Universidade de Lisboa

Faculdade de Ciências

Departamento de Química e Bioquímica

Structural and functional characterization of the bacterial ferrous homeostasis protein FeoA

Vanessa Vieira

Dissertação

Mestrado em Bioquímica

Especialização em Bioquímica

2012

Universidade de Lisboa

Faculdade de Ciências

Departamento de Química e Bioquímica

Structural and functional characterization of the bacterial ferrous homeostasis protein FeoA

Vanessa Vieira

Dissertação

Mestrado em Bioquímica

Especialização em Bioquímica

2012

Orientadores: Dr. Manolis Matzapetakis e Dr. Carlos Farinha

Acknowledgments

I would like to thank all that contributed to the conclusion of my master´s thesis.

First, I have to mention my supervisor in ITQB, Manolis Matzapetakis, which gave me the opportunity to work in this project.

Second, to my lab colleagues Meire Almeida and Ana Pereira that helped me in the laboratory work and in the NMR analysis.

At last, I must also thank my friends and family, especially my parents, for their financial and emotional support.

Contents

Table of Figures

Abstract

The goal of this project was to study the FeoA protein from *E.coli*, in terms of both structure and function as the title suggests.

FeoA is a component of one of the major systems responsible for bacterial ferrous iron uptake, the Feo (**fe**rrous ir**o**n-transport) system. In addition to FeoA it also contains the, FeoB and FeoC proteins.

FeoB acts as a GTP dependent ferrous permease spanning the cytoplasmatic membrane of *E.coli*. Its N-terminal domain is a GTPase which regulates the ferrous influx.

FeoC is believed to be a transcriptional activator or repressor which was recently found in in-vitro experiments to be directly interacting with FeoB potentially modulating its action.

The FeoA function that is an important goal of this project was also further elucidated in June 2012 while this project was ongoing. It was shown in vivo that FeoA also interacts with FeoB protein being involved in its import of Fe (II).

This agrees with our group's hypothesis that the FeoA protein might stimulate the GTPase activity of FeoB.

Our strategy was to probe the mode of interaction of the proteins of the *feo* operon in vitro using NMR spectroscopy. The first step in this task was to characterize each of its components.

In order to determine the FeoA structure through NMR spectroscopy it was necessary to optimize its expression and purification conditions. During the process of purification it was reported for the first time the presence in solution of the FeoA homodimer.

The FeoA monomer was structurally characterize by NMR analysis, where it was verified that the structure is an SH3-domain-like fold in agreement with the crystal structure **(16).** The residues affected by the dimerization were also determined by NMR.

Key words: Feo system; FeoA, NMR, Ferrous Iron Fe^{2+},

Sumário

O objectivo deste trabalho intitulado ""*Structural and functional characterization of the bacterial ferrous homeostasis protein FeoA"* consistiu na determinação da estrutura e função da proteina FeoA da bacteria *E.coli.*

A principal via bacteriana de entrada do ferro ferroso é através do sistema Feo que deriva das palavras inglesas **fe**rrous ir**o**n transport.

O ferro é um elemento essencial para a maioria dos organismos participando em vias metabólicas essenciais. Os sistemas de importação de ferro seja ele ferro ferroso(Fe^{2+}) ou ferro férrico (Fe^{3+}) são importantes alvos terâpeuticos limitando o crescimento celular e são por essas razões alvo de estudos científicos.

O sistema Feo é o principal responsável pela importaçãode ferro ferroso nas bacterias *E.coli* quando estas habitam ambientes anaerobicos e de baixo pH ricos em ferro ferroso.

O operão feo na *E.coli* codifica tres proteinas a FeoA, FeoB e a proteina FeoC.

A proteina FeoB é o principal componente deste sistema actuando como uma permease pela qual o ferro ferroso é transportado para o interior do citoplasma da bacteria.

A sua função de permease está dependente do seu dominio" N-terminal G-protein" que possui actividade enzimática GTPase. A hidrolíse de GTP ocorre a uma taxa extremamente baixa e por isso é especulado que a proteina FeoA seja o factor utilizado para aumentar a sua taxa de hidrolíse.

Recentemente a função da FeoA foi esclarecida num artigo cientifico publicado em Junho de 2012 durante o decurso deste projecto. Revelou-se neste artigo que a proteina FeoA é essencial para a função da FeoB. Sem interacção entre as duas proteinas não ocorre o influxo de Fe^{2+} para citoplasma.

Para determinar a estrutura da proteina FeoA atraves de RMN foi necessario um longo processo de optimização das condições de expressão e purificação.

A analise de proteinas através de RMN pressupõe que estas estejam isotopicamente marcadas seja com ^{15}N ou ^{15}N /^{13}C de forma a diminuir a complexidade dos espectros obtidos. Esta marcação só foi possivel através da expressão de da proteina FeoA em meios marcados com os respectivos elementos químicos. Os meios minímos foram utilizados neste processo visto que possibilitavam a substituição das fontes de carbono e nitrogénio pelos respectivos reagentes equivalentes (^{13}C-Glucose e $^{15}NH_4Cl$)

Durante o processo de optimização da purifição foi detectada a presença do dimero de FeoA (17kDa) em solução. Este resultado foi muito importante pois foi a primeira que foi detectada a presença do dimero de FeoA em solução, apesar de existirem algumas estruturas de FeoA em dimero cuja relevância fisiológica é ainda pouco clara.

A concentração total de proteina afecta a proporção do monomero e do dimero de FeoA. Foi também determinado ao longo deste projecto a impossibilidade de separar ambas as formas visto que se encontram num equilibrio dinâmico.

Nas fases preliminares deste projecto foi produzida uma amostra cuja concentração era muito baixa 10uM que continha maioritariamente o monomero. Esta amostra era inicialmente mais concentrada mas com o tempo a amostra degradou-se e precipitou levando a uma diminuição da concentração total. Esta amostra apenas possibilitou a obtenção de um 1H-15N-HSQC que devido á sua diluição não permitiu a obtenção de espectros essenciais á determinação da estrutura.

O **^{1}H-^{15}N-HSQC** obtido quando comparado com o **^{1}H-^{15}N-HSQC** na qual se encontram ambas as formas possibilitou a distinção dos picos relativos ao monomero e ao dimero.

Para determinar a estrutura da FeoA foi utilizado o **^{1}H-^{15}N-HSQC** como guia para identificar os picos em vários outros espectros. Os espectros **HNCO**,**HN(CA)CO**, **HNCACB** e **HN(CO)CACB** foram utilizados para a identificação dos atomos do backbone da proteina ao mesmo tempo que nos permitiram ligar os varios picos de acordo com a sequência da proteina FeoA.

Por outro lado os espectros **^{15}N-TOCSY**, **hCCH-TOCSY**, **^{1}H-^{13}C-HSQC** permitiram a identificação dos atomos da cadeia lateral dos aminoacidos.

No fim este conjunto de picos identificadfos foram utilzados nos espectros NOESY que nos permitem identificar interaçãoes até um maximo de 5 A entre H-H. Esta informação permitiu-nos produzir a estruttura da FeoA ao fim de um longo processo de analise por parte de vários softwares e fases finais de refinamento.

A estrutura final da FeoA da *E.coli* possui uma estrutura SH3 domain-like fold semelhante á estrutura determinada através de cristalografia do organismo *Stenotrophomonas maltophilia*.

Palavras chave: Sistema Feo, FeoA, RMN, Ferro ferroso, Fe^{2+}

1. Introduction

The purpose of this project entitled *"Structural and functional characterization of the bacterial ferrous homeostasis protein FeoA"* is the study of FeoA protein from *E.coli.* This protein is one of three comprising the *feo* operon, which is believed to be one of the main paths of ferrous iron influx in bacteria. It is a small, 8.5 kDa protein and the only structure of this protein described in the literature so far, is of a possibly non physiological dimeric form, obtained by X-Ray Crystallography. the structure is reminiscent of SH3 domains, which suggests that it might be involved in protein-protein interactions.

In this work it was possible to determine the FeoA structure in solution with the use of NMR spectroscopy. The project included the sample preparation, establishing and optimizing the expression and purification protocols, the NMR data collection and structure determination.

1.1 Role and properties of Iron

Iron is an essential element for most forms of life, although there are some organisms like the prokaryotes *Lactobacillus plantarum* and *Borrelia burgdorferii* that can grow in the total absence of iron. In these bacteria the Mn^{2+} substitutes for iron as the metal component of enzymes that normally contain Fe^{2+}. **(1; 2)**

For the organisms that depend on iron, this element is responsible for the biological activity of many proteins and enzymes participating in essential metabolic pathways such as amino acid synthesis, respiration and DNA biosynthesis.

Despite being one of the most abundant metals on earth, there is practically no free iron available for bacteria, in most of the environment they colonize. In order to meet their iron needs bacteria evolved multiple iron acquisition systems, reflecting the diversity of their potential iron sources. **(3)**

Thus, in order to achieve effective iron homeostasis, organisms must balance their need to efficiently scavenge iron from their surroundings to ensure that adequate supplies are maintained, with the careful management of cellular free iron levels to guard against iron induced toxicity.**(3)**

Due to its importance iron depletion inhibits the growth of many microorganisms, including that of pathogenic bacteria. In addition to the low availability of iron in most environments, the pathogenic bacteria have also to deal with the host iron-withdrawal

responses. In the case of *E.coli* most strains are not pathogenic, however, there are still 200 *E.coli* strains that can cause life-threatening diseases such as diarrhea and urinary infections. For these pathogenic strains as well as for other pathogenic species, their iron import systems in all their variety are potential therapeutic targets to limit cell growth. **(4)** Hence, the research within this project may have direct impact in the design of novel therapeutic strategies.

1.2-Iron acquisition

1.2.1 Iron acquisition: The import of ferric Iron (Fe^{3+})

In bacteria the transport of iron into the cytoplasm is accomplished via multiple routes, depending on the oxidation state of iron.

To better understand the role of iron on the multiple import systems of bacteria, we need to discriminate between its two readily inter-convertible redox states: the Fe^{3+} ferric iron and the Fe^{2+} ferrous iron. These two distinct oxidation states exist under different conditions; with the ferric form Fe(III) dominating at what we perceive as ambient conditions (neutral pH and oxygen rich atmosphere), while the ferrous form Fe(II) is predominant in acidic and/or anaerobic environments.**(3)**

Although the ferric form is the most abundant due to its prevalence in aerobic and neutral pH environments, it is characterized by its poor solubility (10^{-18} M at pH 7.0 for ferric state versus 0.1 M at pH 7.0 for ferrous) making the bioavailability only possible in the form of various insoluble minerals.

This low solubility of ferric iron has driven the majority of bacteria to produce or secrete iron-binding agents called siderophores (molecules that chelate ferric iron with high affinity) that bind iron and transport it into the cell in the form of ferric complexes. In Gram-negative bacteria, such as *E. coli*, siderophore-bound Fe(III) is recognized by outer membrane receptors and is then transported into cells in a process that requires the TonB–ExbB–ExbD protein complexes.**(5)** (Figure 1)

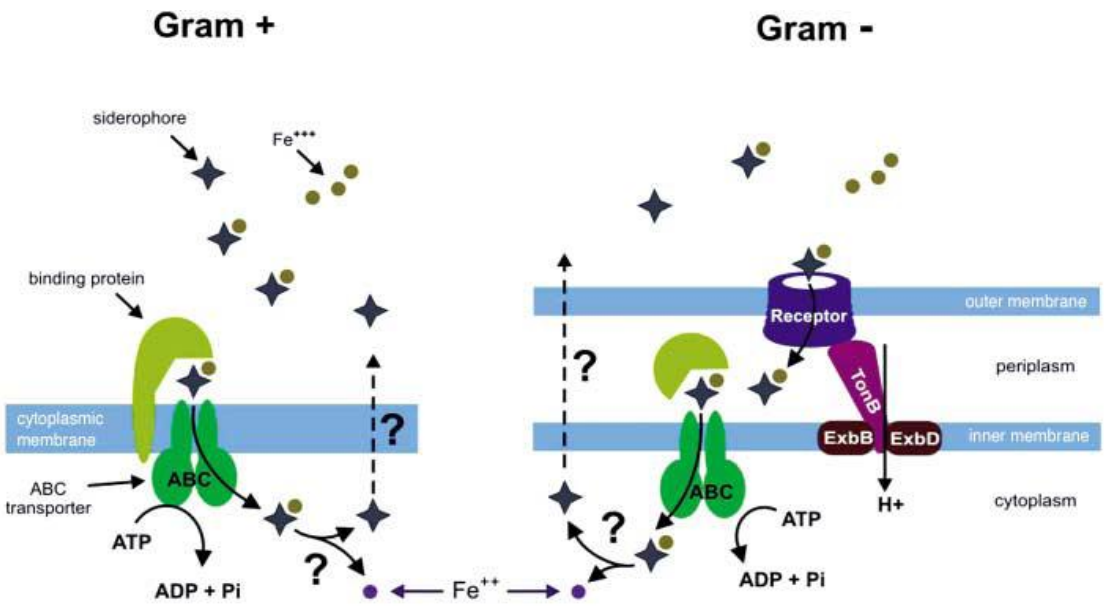

Figure 1: Organization of siderophore excretion and internalization in gram-positive and gram-negative bacteria. (5)

Iron acquisition: The import of ferrous Iron (Fe^{2+})

Under anaerobic conditions and acidic pH, Fe (II) is the major form of bioavailable iron. This highly soluble ion is the preferred form of iron utilized by bacteria under these conditions since, unlike the ferric form, it can be transported as a free ion. In the case of Gram-negative bacteria the ferrous iron diffuses freely through the outer membrane porins and is then transported through the cytoplasmic membrane by an ABC ferrous ion transporter conserved in many species. (6)

The transport through the cytoplasmatic membrane is accomplished by a transport system like the metal-ABC permeases and Nramp-like transporters such as SitABCD and MntH respectively, that display specificity for divalent metals, such as Fe2+ and Mn2+ but are primarily involved in the uptake of manganese rather than iron. **(4)** The major route for bacterial-ferrous-iron uptake would appear to be, in many cases, via Feo (**Fe**rrous ir**o**n transport) system. The Feo system appears to be unlike any other bacterial ion transporter. However, despite its importance and novelty, surprisingly little research has been performed on the Feo pathway.

While the Feo pathway is not the prevalent mode of iron uptake for most bacteria, it is very important in bacteria that inhabit anaerobic environments like the human intestine, where the ability to import ferrous iron is crucial. **(7)**

1.3- The Feo operon in *E. coli*

The analysis of the *feo* locus in *E.coli* proved the existence of three closely associated genes likely to form an operon, the *feoABC* . **(8;9)**

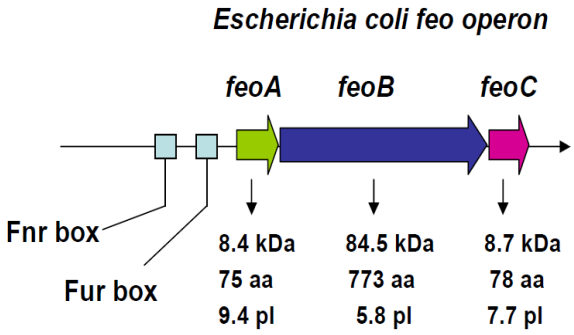

Figure 2: Gene organization of the Feo system in *E.coli*. The size, molecular weight and pI of each of the proteins encoded is also shown.

The operon encodes three predicted proteins: FeoA, a small 75-residue hydrophilic protein; FeoB, a large 773-residue protein with an integral membrane domain (_270–773) likely to act as the ferrous permease; and FeoC a small 78-residue hydrophilic protein found so far only associated with c-proteobacterial Feo systems.

The role of FeoB as a GTP gated transmembrane permease is starting to be understood **(10)**, and was shown recently to be essential for transport of periplasmic Fe(II) to the cytoplasm. However, little is known about the function of the two small proteins FeoA and FeoC **(6)**. FeoA is proposed to function in a concerted mechanism with FeoB for transport of ferrous iron and FeoC only recently (September 2012) was shown to interact with FeoB modulating its activity **(28).**

1.4- Control of the feo operon in E.coli.

Most of the bacterial genes involved in iron assimilation are expressed only under conditions of iron deficiency. When the intracellular iron concentration rises, these genes are repressed. It seems that the iron acquisition systems are not simultaneously induced in response to iron limitation irrespective of the nature of the iron source.

Several iron acquisition systems are positively regulated by their respective iron sources.

The expression of the feo operon is regulated by Fur and Fnr transcriptional regulators that activate the expression under anaerobic conditions (Fnr) and repress it at high-iron concentrations (Fur).

The upstream region of *feo* operon contains the binding site for the regulatory protein Fur, which acts with iron (II) as a corepressor in all known iron transport systems of *E.coli,* switching the operon off when the intracellular levels of Iron (II) are high.

In addition, an Fnr binding site was identified in the promoter region. Fnr is the transcriptional activator of anaerobic respiratory genes. *feoABC* is anaerobically induced by Fnr, consistent with the induction of Feo activity under anaerobic conditions. Fe(II) has also been shown to influence the activity of Fnr, since iron limitation led to a reduced activity of Fnr.**(3)**

1.5-The FeoB protein

FeoB is the major component of the Feo system. It acts as the permease through which ferrous iron is transported into the cell. In *E. coli* the FeoB protein is composed of a hydrophilic N-terminal domain (FeoB-N; residues 1 to 270) and a C-terminal integral-membrane domain (FeoB-C; residues _271–773) that is predicted to consist of eight transmembrane α-helices.**(6)**

FeoB transport activity is strongly dependent on its N-terminal G-protein domain. The N-terminal 'G-protein region' has GTPase activity, as observed for the homologous small regulatory G-proteins of eukaryotes.

This hydrophilic domain is starting to be understood and was shown to be essential for transport of periplasmic Fe(II) to the cytoplasm.**(11)**

The N-terminal GTPase is composed of two regions, the catalytic GTPase (FeoB_G, residue 1 to 160, ~18kDa) and the GDP Dissociation Inhibitor (FeoB_GDI, residue 160 to 270, ~12kDa) domains.**(12)**

This domain is able to hydrolyze GTP to GDP but at an extremely low rate because of its very high affinity for the reaction product, GDP, which results in inactivation of the enzyme after one turnover cycle. Because of these properties FeoB-N terminal is believed to require some factor or event to increase its hydrolysis rate.

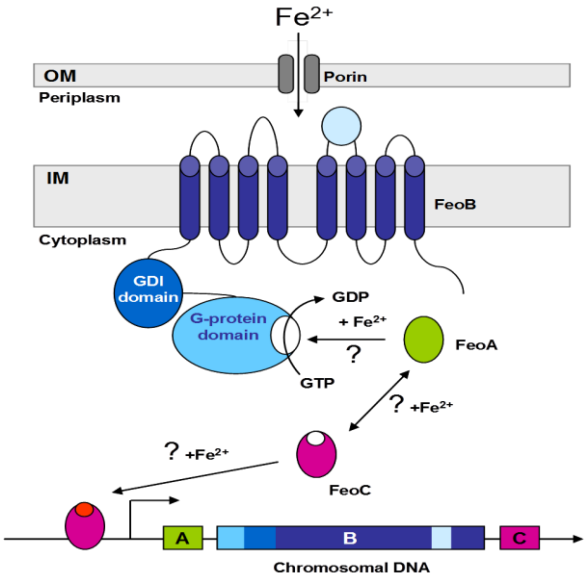

Figure 3: Graphical depiction of the proposed function of the Feo system. The question marks denote unconfirmed interactions.

The eukaryotic G-proteins have a similar mechanism of function as the FeoB, they also possess low rate in hydrolyzing GTP. In the eukaryotic G-proteins this is overcome by the interaction with GAPs proteins through their SH3 domain that stimulates GTPase activity.

It is very interesting that in the crystal structure of FeoA an SH3-like-fold- domain was identified that could point to a modulation of FeoB activity by FeoA protein.**(15)**

1.6- The FeoC protein

In *E.coli* the feoB gene precedes a gene called feoC which is also found similarly located within the feo locus of other (gamma)-proteobacteria.

Multiple-alignment of the FeoC proteins (15 examples in the NCBI database, Sept 2005) shows that they possess four conserved Cys residues (CxxGxCKxCPx4-7C) likely to provide a binding site for an [Fe–S]-cluster.**(6)**

Furthermore, the solution structure of FeoC has recently been deposited in the Protein Data Bank revealing that the protein is a monomer with a winged-helix fold **(14)** within its N-terminal region. This fold belongs to the family of helix-turn-helix folds and is normally DNA (sometimes RNA) binding in function supporting a likely role as a transcriptional regulator directly controlling feoABC expression.

Also in a recently new paper it was shown *in-vitro* that FeoC binds to FeoB disrupting the formation of a potential pore by interfering with FeoB trimerization. These results provide evidences suggesting that FeoC may play a role in regulating Fe^{2+} transport, in addition to or in conjunction with its presumed gene regulator role. **(28)**

1.7- The FeoA protein

The function of FeoA is starting to be understood only in recent times (June 2012) and only a few structures of the protein are available, in data banks such as the crystallographic structure of FeoA from *Stenotrophomonas maltophilia* **(15).**

This FeoA structure was found to adopt an SH3-domain-like fold similar to that adopted by SH3-domains of eukaryotic proteins, although no sequence identity exists among these proteins.

1.7.1- FeoA crystal structure

The only tertiary structure published of FeoA is the crystal structure from *S. maltophilia (Sm)* that revealed a unique prokaryote SH3-domain-like- fold that is rarely observed in the bacterial kingdom.

Interestingly, in the crystal structure the *Sm* FeoA is found as a homodimer that is cross-linked by two zinc ions and six chloride ions**.(15)** It is likely that this mode of dimerization described in this paper is not physiologically significant since very little or no dimer was observed in solution studies. In addition Chloride is very rarely seen as a ligand in water exposed metal complexes in solution.

The eukaryotic SH-3 domain-like-fold adopted by FeoA bears low sequence identity to eukaryotic proteins that have a similar SH-3 domain. The homology exists at the structural level only. **(16)**

The SH3 domains are found in a number of eukaryotic proteins involved in signalling (e.g. Ras-GTPase activating protein or GAP) and in cytoskeletal dynamics (e.g. myosin 1B).**(16)**

These SH3 domains are composed of 50–70 residues that form a barrel with five anti-parallel β-strands and are involved in mediating protein-protein interactions with target proteins that contain 10-residue Proline rich motifs. **(17)**

One of the other prokaryotic proteins that have an SH-3 domain-like-fold in the presence of manganese and cobalt ions is diphtheria toxin regulator, DtxR **(18).**The SH-3 like domain of DxtR also exhibits structural homology to FeoA.

The diphtheria toxin regulator protein functions as the major iron-responsive regulator in Corynebacterium diphtheriae.

Both FeoA and the SH3-like-domain-fold of DtxR proteins have no clear sequence similarity with eukaryotic SH3- domain proteins, but their common fold does suggest a common role in mediating protein–protein interactions. This raises the possibility that FeoA interacts with its partner protein, FeoB, thereby mediating FeoB-dependent Fe (II)-uptake activity through this SH-3 domain.

1.7.2-FeoA function

The function of FeoA protein recently became clearer with the June 2012 paper "*The FeoA protein is necessary for the FeoB transporter to import ferrous iron". (***19)**

With the use of a Bacterial two-hybrid assay it was determined that the FeoA protein directly and specifically binds to the FeoB transporter in vivo. This FeoA-FeoB interaction appeared necessary for FeoB mediated Fe(II) uptake because *Salmonella* expressing the mutant FeoA that could not interact with FeoB failed to uptake Fe(II) via the FeoB transporter.(19) One of the residues found to be important for FeoA interaction with FeoB is Leucine26. This residue is highly conserved in the FeoA proteins from various bacterial origins as well as the structural homologous DxtR protein.

The absence of a functional FeoA protein resulted in impaired import of Fe(II) by FeoB.

As mentioned before the FeoB N-terminus has been demonstrated to possess GTPase activity.

Similar to eukaryotic G-proteins, the FeoB G-protein possesses weak GTPase activity and displays very slow GTP hydrolysis. In the eukaryotic systems the presence of GTPase activating proteins (GAPs) that directly interact with the G- proteins stimulates their GTPase activity. **(6)**

The results of this recent paper raise the possibility that the FeoA protein might stimulate the GTPase activity of FeoB protein in the same way as the eukaryotic GAPs.

Interestingly the presence of the SH3 domain-like fold in FeoA evidenced by the tertiary structure of FeoA from *Stenotrophomonas maltophilia* gives credit to this

hypothesis since in eukaryotes the SH3 domains are present in many signalling proteins including GAPs known to mediate protein–protein interactions. **(15)**

Therefore it could be suggested that the FeoA protein resembles eukaryotic GAPs, which further supports the hypothesis that the FeoA protein acts as a modulator for the FeoB's G-protein function.

1.8-Protein structure determination from NMR data

Nuclear magnetic resonance (NMR) is a very versatile technique 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 organic molecules and biomolecules.

The available nuclei suitable for nuclear magnetic resonance are those with non-zero spin quantum number. Thus, only unpaired nuclear spins are detected in NMR spectroscopy. The most widely used nuclei in biomolecular NMR are the ones with I1/2 (like the ^{1}H, ^{15}N and ^{13}C).

Although NMR was discovered in 1946, its application to biological systems only started in the late 1970s and early 1980s. Major breakthroughs happened in this period that revolutionized the field, allowing the use of NMR in determining the three-dimensional structures of macromolecules. **(20)**

Due to differences / advantages with respect to X-Ray crystallography, the interest in protein structure determination by NMR is immense. Namely, NMR can provide complementary structural information in a more physiologically relevant solution environments. Moreover, since some biomolecules are difficult to crystallize, NMR can be used as an alternative method for obtaining three-dimensional structures.

More recently, in the late 1980s and early 1990s, the NMR limit of analysis was shifted from molecules of size less than 10kDa to a upper limit of 35 kDa due to the development of multidimensional heteronuclear methods as well as advances of molecular biology that allow to overexpress isotope labelled proteins such as $^{15}N/^{13}C$ double labelled proteins.

Today one fifth of the macromolecular structures deposited in the PDB (Protein DataBank) are derived from NMR spectroscopy. Despite its size limitation for macromolecular structure determination, NMR has the following unique features:

(a) It allows structural studies in a physiological relevant solution environment, which avoids experimental artefacts such as crystal packing seen in crystal structures.

(b) It allows structural studies of some molecules that are difficult to crystallize such as flexible protein domains, weakly bound protein complexes.

(c) It can provide information about protein dynamics, flexibility, folding/unfolding transitions.

The structure determination of proteins by NMR spectroscopy includes normally three stages:

(a) Sample preparation, NMR experiments for data collection and data processing

(b) Sequence specific assignment, NOESY assignment, assignments of other conformational restraints such as J coupling and hydrogen bonding

(c) Structure calculation and structure refinement

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 angle 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.

2. Plan and Methods

The main goal of this project is to determine the structure of FeoA protein using NMR spectroscopy. The protein samples produced for NMR have to be isotopically labelled in order to decrease the complexity of spectra making the analysis of data much faster and easier. The expression and purification protocols have to be optimized in ways that are only relevant for this technique. The use of expensive isotopically labelled reagents such as ^{13}C-Glucose and $^{15}NH_4Cl$ dictates that the expression of the protein has to be optimized to give the highest possible yield while using smaller amounts of these reagents.

Besides being isotpically labelled the protein sample used for NMR analysis, also has to be pure and stable. This is achieved with the next stage of the protocol, the optimization of the best purification scheme.

The FeoA project had already some preliminary results done by Meire Almeida a lab member. This preliminary results that included some NMR spectra, helped simplifying the design of the various protocols.

2.1. Expression

2.1.1. Expression system

The gene coding for FeoA protein was amplified from genomic DNA from DH5α *E.coli* (NZY5α by NZYtech genes & enzymes) and inserted into the vector pET-21c (+) and was transformed into BL21(DE3) strain cells of *E. coli* (from NZYtech genes & enzymes). A glycerol stock (80% glycerol) of BL21 cells was prepared and stored at -80^{o}C for further expression and purification trials. (Meire Almeida).

The pET-21c (+) vector used to overexpress FeoA protein has a selective marker for ampicillin (amp^{R}) and a LacI gene that induces the expression of FeoA with the addition of IPTG. **(21)** This plasmids were available prior to the start of this work. (Figure 4)

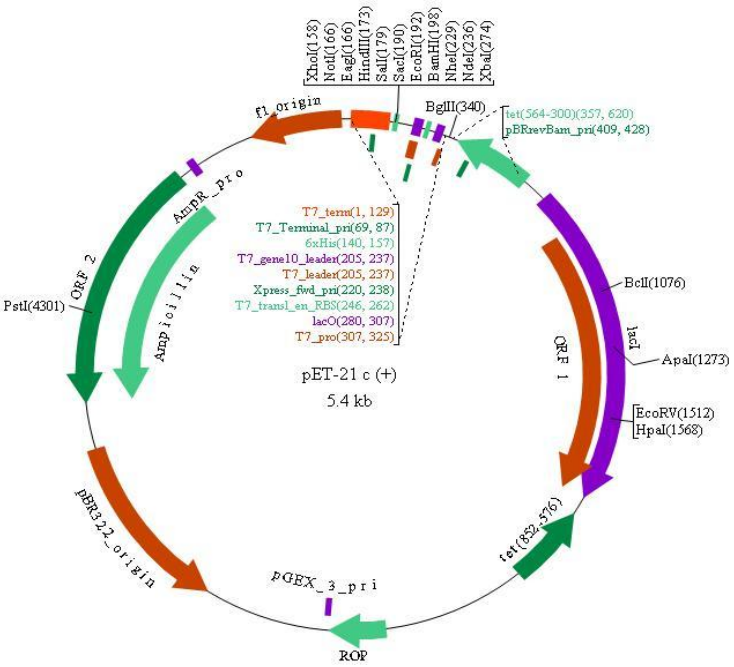

Figure 4: pET-21c (+) vector from Novagen used to clone the FeoA gene

2.1.2. Optimization of the expression conditions

To maximize protein expression and to increase the soluble fraction in bacterial lysates for posterior purification, a series of expression conditions was varied:

1- The cell density (OD_{600nm}) was measured as a function of time in hours and the results were used to make a growth curve. The growth of the BL21 strain transformed with the pET21-c (+) was monitored in LB medium with 100 mg/ml of ampicillin at 37^{o}C.

2- The induction of protein expression was performed in the middle of the log-phase (determined in step 1 OD_{600nm}) with a range of different IPTG concentrations from 0.1mM to 2mM.

3- The length of induction of protein expression was varied from 1 hour to overnight as well as the temperature from 20^{o} C to 37^{o}C.

2.1.3. Overexpression of isotope-labelled proteins

After the optimization of the best conditions to express FeoA protein in order to have the biggest amount possible while minimizing the use of expensive labelled reagents, it was time to overexpress FeoA both single ^{15}N labelled and double labelled $^{13}C/^{15}N$ for NMR analysis.

The goal of this stage was to overexpress in large scale (mg) isotope-labelled FeoA for NMR studies, for structure determination requiring that our target protein be uniformly single ^{15}N labelled and double labelled $^{13}C/^{15}N$. This was accomplished by growing cell cultures in minimal media in which $^{15}NH_4Cl$ and ^{13}C- Glucose were the only sources for nitrogen and carbon atoms.**(22)**

Minimal media is the most commonly used and cheap medium to overexpress isotope labelled proteins for NMR sample preparation. The minimal media is a mixture of M9 salts, glucose (as the sole carbon source) and ammonium chloride or ammonium sulphate (as the sole nitrogen source) that is formulated in either H_2O or D_2O depending on the size of the protein. Deuteration is necessary for proteins greater than 25kDa in order to simplify the spectra. Since FeoA is 8,4 kDa there is no need to use D_2O in the growth media. Often, depending on the *E.coli* strain used, additional reagents, such as cofactors, are added to the media to enhance growth. In the case of BL21 (DE3) *E.coli* strain used the addition of vitamin mix and some metals were important for a better yield.

10X M9 Minimal Media(1L)		Media Composition (1Liter)	
$Na_2HPO_4.7H_2O$	**64g**	Sterile MilliQ H_2O	880ml
KH_2PO_4	**25g**	10x M9 media	100ml
NaCl	**2.5g**	1M $MgSO_4$	2ml
		1M $CaCl_2$	20 µl
		Metals Stock	20 µl
		Vitamin Mix stock	10ml
		Antibiotic 100 mg/ml	1ml
		^{13}C-Glucose	3.2g
		$^{15}NH_4Cl$	1g

Table 1: Composition of the optimized M9 minimal media to express labelled FeoA protein.

Because of the lower nutritional content of minimal media, bacterial growth ~~in it~~ is often slow and sometimes irreproducible. In order to achieve higher cell densities and have faster and more reproducible growths, cells were first grown in rich media such as LB and then transferred to minimal media prior to induction.

2.1.4. Expression protocol

1-A pre-culture was prepared the night before by using 1ml of the glycerol stock of BL21 cells transformed with the FeoA containing plasmid, to inoculate 50 ml of LB medium containing 100µg/ml ampicillin. The culture was grown overnight at 37^{o}C at 150 rpm.

2-The pre-culture was then used the next morning (after 16 hours) to inoculate a fresh LB medium (100µg/ml ampicillin). The cultures were then placed in a shaker at 37^{o}C at 150 rpm. The cells were grown until an OD $_{600nm}$ 0.6-0.8 and were then collected by centrifugation at 8000 rpm at 25^{o}C.

3-The cell pellet obtained was resuspended in M9 minimal medium of ½ LB volume and after 20 minutes at 25^{o}C 90 rpm the induction started with the addition of 1 mM of IPTG and 150 µl of Iron(II)(150µM) . The OD_{600nm} was used to monitor cell density and samples were taken each hour during the 4 hour induction period for analysis by SDS-PAGE.

4- After the 4 hours of induction the cell culture was subjected to another centrifugation at 8000 rpm at 4^{o}C. The cell pellet was then stored at -80^{o}C. to be used when there was a need to purify FeoA for NMR studies.

2.2. Purification

FeoA is a small protein of 75 amino acids with a molecular weigth of 8370.8 Da and the presence of only one cysteine residue at the C-terminal. It has a high isoelectric point (P.I.) of 9.4 making a purification step based on charge such as the cation exchange chromatography ideal to purify our target protein. Knowing that the majority of proteins have isoelectric points within the pH range of 5.5 to 7.5, at pH 7, the majority of proteins will have negative or even neutral charge, allowing for almost selective binding of FeoA to the cationic column. This was our strategy to capture the FeoA protein and separate it from the majority of contaminants. **(23)** A final step of size exclusion chromatography would then separate FeoA from the remaining contaminants.

2.2.1. Optimization of purification protocol

Cell lysis: Cells pellets were removed from -80^{o}C and re-suspended in lysis buffer (50mM Tris pH 7.0, 150mM NaCl, 1mM PMSF plus protease inhibitor cocktail without EDTA - Roche). The use of reducing agents such as DTT (2mM) or B-Mercaptoetanol (2mM) in the lysis buffer was only present when reducing conditions were been explored while purifying FeoA protein.

Cell disruption was achieved by passing the solution three times in a large French Press at 1000 psi pressure at room temperature. The cell lysate was subjected to ultra-centrifugation at 50 000 rpm during 1 hour at 4^{o}C on a Beckman TL100 ultracentrifuge.

Dialysis: After ultra-centrifugation the supernatant was collected and diluted with dialysis buffer (20mM Tris pH 7.0) in a 2X dilution and then transferred to a dialysis membrane (MWCO 3,5 kDa) and was kept overnight in 20mM Tris pH 7.0. The 1 liter buffer solution was changed 2 to 3 times in a period of 16 to 24 hours. The dialysis

solution also included reducing agents such as DTT (2mM) or B-Mercaptoetanol (2mM) when reducing conditions were necessary for the study of the protein.

Cationic exchange Chromatography: The target protein was purified by cation exchange chromatography on a 5 ml Hi-Trap SP FF column that was equilibrated with buffer A (20mM Tris pH 7.0) and eluted with a linear gradient (0- 25%) of buffer B (20Mm Tris pH 7.0 1M NaCl). The buffers also included reducing agents such as DTT (2mM) or B-Mercaptoetanol (2mM) when the elution behaviour of the protein was studied in non oxidative conditions.

The fractions containing the FeoA protein were always verified by SDS-PAGE.

Size Exclusion Chromatography: A second step of purification was also optimized. The FeoA fractions pooled together from the cation exchange chromatography were injected in the size exclusion column (Hi Load 16/60 Superdex 75 prep grade or Superdex 75) that was equilibrated with buffer A (20 mM Tris pH 7.0 2 mM DTT) with different salt concentrations (0-150 mM NaCl) to assess the best salt concentration in terms of protein stability.

Protein concentration: The final sample collected from the size exclusion chromatography was concentrated to produce a suitable NMR sample (0.6-1mM). The sample was concentrated using a 3kDa MWCO Vivaspin 15R device.

The concentration conditions were also optimized to reduce aggregation and degradation of the sample to achieve the highest possible protein concentration.

Protein concentration determination: The protein concentration was determined by absorbance at 280nm in the NaNo Drop 2000. The absorbance was measured after the auto-zero with the corresponding buffer solution (FeoA Extinction coefficient: 8 400).

2.3.NMR sample preparation and data collection

The final labelled protein samples had an average concentration in the range of 0.6-0.8 mM. This concentration was suitable for NMR analysis although higher concentrations (1mM) are always desired for structural studies.

The buffer used for NMR analysis as a composition of 20 mM Tris pH 7.0, 150 mM NaCl, 2mM DTT and 0.5 mM EDTA. In the final sample it was also added 10% D_2O needed for signal lock.

Samples were loaded into a 5 mm NMR tube (NEWERA UL-5) and all the NMR data were recorded on an 800 MHz Bruker Avance III NMR spectrometer at 298K using a TXI Z axis gradient room temperature probe.

^{1}H 1D and ^{1}H-^{15}N-HSQC spectra were recorded with the single ^{15}N labelled FeoA sample. The ^{13}C/^{15}N double labelled sample allowed us to record the HNCO, HN(CA)CO, HNCACB, HNcoCACB, ^{15}N-TOCSY, hCCH-TOCSY, ^{1}H-^{13}C-HSQC, ^{15}N-NOESY and ^{13}C-NOESY spectra.

2.4 Data Processing and Analysis

Data collection and processing was done with Bruker´s **TopSpin 2.1** software (Bruker Biospin). Both CARA 1.8.4.2 and CCPN 2.2.2 software were used to analyze the collected NMR spectra and to perform sequential assignment.

For structure determination it was used the CYANA, UNIO (ATNOS CANDID), RECOORD, DANGLE (h-bonds) and ECI (data submission) software.

3. Results and Discussion

3.1. Expression of FeoA protein

3.1.1-Growth Curve

The growth curve of BL21 strain transformed with pET 21c (+) was used to decide the best induction point.

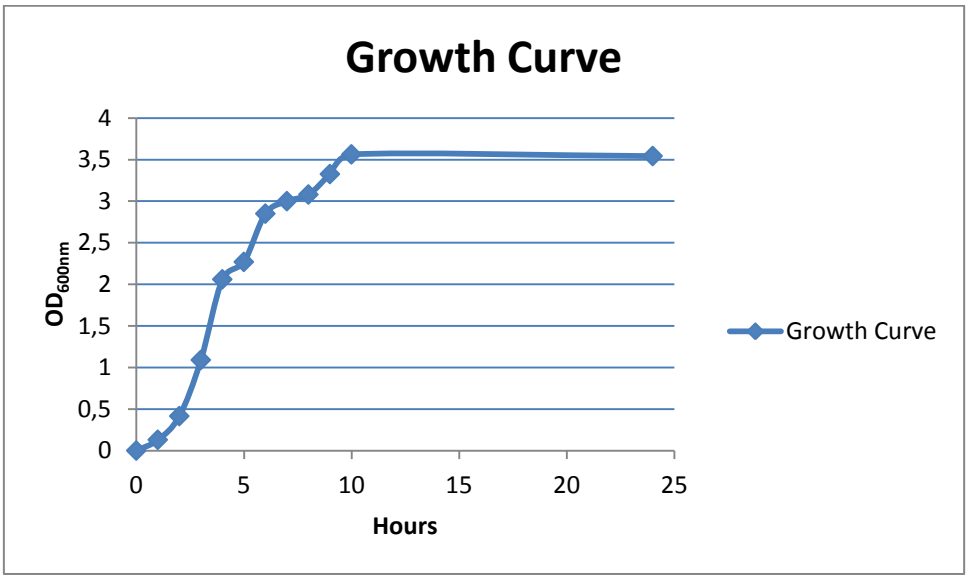

Figure 5: Growth curve of BL21 transformed with pET 21c(+).

The Growth Curve enabled us to determine the beginning of the exponential phase, when the cells are usually in their healthiest state. This stage corresponds in our BL21 strain to an OD_{600nm} of around 0.6-0.8. This OD_{600nm} was then used as the point to start the induction of FeoA expression in minimal media.

3.1.2- Best IPTG concentration

It is important to determine the best IPTG concentration in order to have an optimized expression of FeoA. The FeoA gene was inserted into the pET 21c(+) vector and transformed into a BL21(DE3) strain of cells. The pET manual states that for a pET construction carrying the T7/lac promoter such as the pET21c(+) needs a final concentration of 1mM of IPTG for full induction. This is only recommended if the plasmid was established in a DE3 strain such as our case.

3.1.3 –Time and temperature of induction

In Figure 6 it is shown the induction of FeoA expression in minimal media at different temperatures at 25°C and 37°C in a 4 hour induction period.

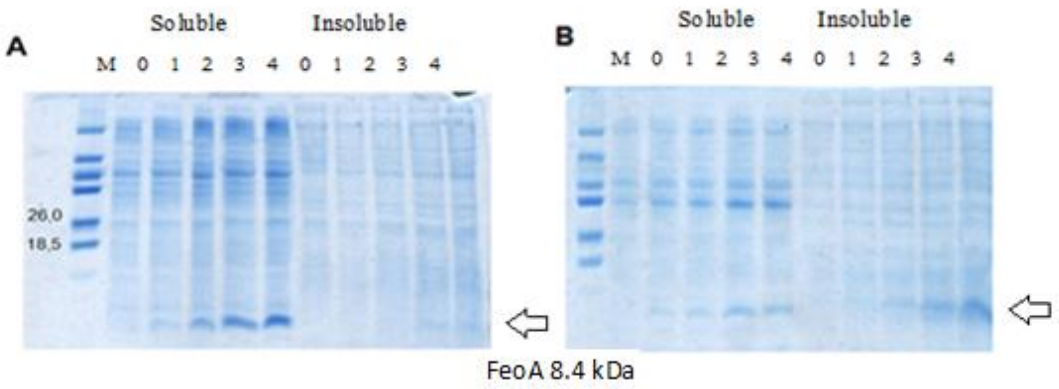

Figure 6: SDS-PAGE (Tris-Glycine 12.5%), A **soluble/insoluble test, *E. coli* induced at 25°C. Lane 1: Marker. Lane 2-6: soluble fractions from 0 to 4h of induction. Lane 7-11: insoluble fractions from 0h to 4h of induction.** B: **Soluble/insoluble test. *E. coli*, induced at 37°C. Lane 1: Marker.Lane2-6 Soluble fractions from 0 to 4h of induction. Lane 7-11: Insoluble fraction from 0 to 4h of induction.**

The induction of protein expression at 37°C decreases the solubility of FeoA protein since a significant part of the protein is localized in the insoluble fraction. The expression at low temperatures such as 25°C is needed in order for the majority of the protein to be soluble.

In Figure 7 it is shown the induction of FeoA protein at 25°C during an overnight period that lasted 16 hours.

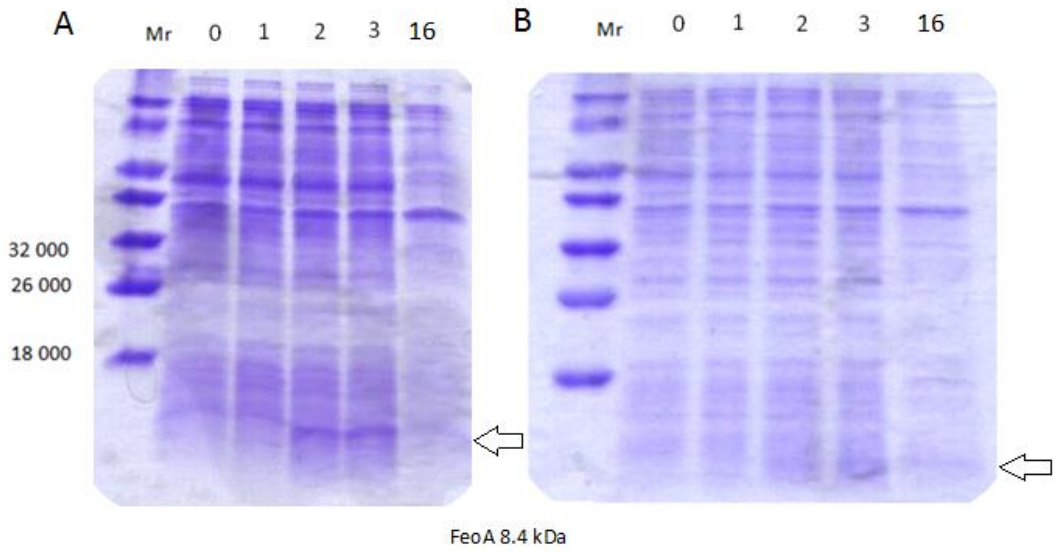

Figure 7: SDS-PAGE (Tris-Tryc 12.5%), A **soluble test, *E. coli* induced at 25°C for a16 hour period. Lane 1: Marker. Lane 2-6: soluble fractions from 0 to 16h of induction.** B: **Insoluble test. *E. coli*, induced at 25°C for a 16 hour period. Lane 1: Marker. Lane 2-6 Insoluble fractions from 0h to 16 of induction.**

The 16 hour period of induction of the FeoA expression leads to a degradation in the overall protein content. The FeoA protein is not present in the soluble or insoluble fraction.

The induction of FeoA expression during an overnight period posses not advantage in terms of amount of protein since the protein degrades.

The induction of FeoA expression has to be performed at 25°C and for a period of 4 hours.

3.2. Purification of FeoA protein

The purification results are divided in two parts: the first is related to the optimization of the best, most efficient purification scheme and the second part is related with the actual results of the purification protocol when applied to the various labelled forms of FeoA, such as single labelled ^{15}N and double labelled $^{13}C/^{15}N$.

3.2.1.Optimization of the purification scheme

The FeoA protein has a relatively high isoelectric point (pI) of 9.4 when compared to other proteins, making a purification step based on charge such as the ion exchange chromatography a perfect strategy to isolate our target protein.

For the ion exchange chromatography the choice of the pH value was very important not just to allow the most efficient purification but also to maintain the stability in solution of the protein.

At a pH below their isoelectric point such as pH 7.0, used in all the solutions, the FeoA protein will have a positive charge binding to a negatively charged medium or *cation exchange* and this is our purification strategy for the first step.

The protein is then eluted with increasing ionic strength (NaCl in this case). Knowing that the majority of proteins have isoelectric points within the range of 5.5 to 7.5 this means that at pH 7 these proteins will not be able to bind the column allowing the separation of our target protein from critical contaminates such as proteases from the crude sample.(23)

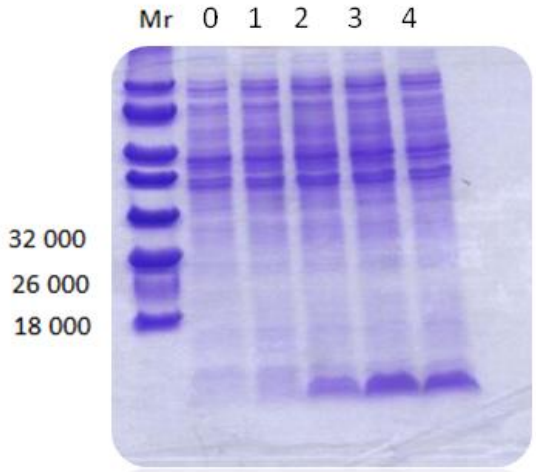

Figure 8: SDS-PAGE (Tris-Glyc 15%) of FeoA expression at 25°C. Soluble fractions from hour 0 to hour 4.

The FeoA protein exists mainly in the soluble fraction of the cell extract as can be seen on Figure 8. The supernatant solution contains our protein which is directly loaded into the column**.** To allow maximal binding of our target protein to the column

the supernatant solution is subjected to a dialysis procedure to lower the ionic strength and also to adjust the pH to similar values to that of the starting buffer.

The elution of FeoA protein shows an interesting profile of three peaks (Figure 9), all proven to be our studied protein as verified by an SDS-PAGE analysis of the different peaks.

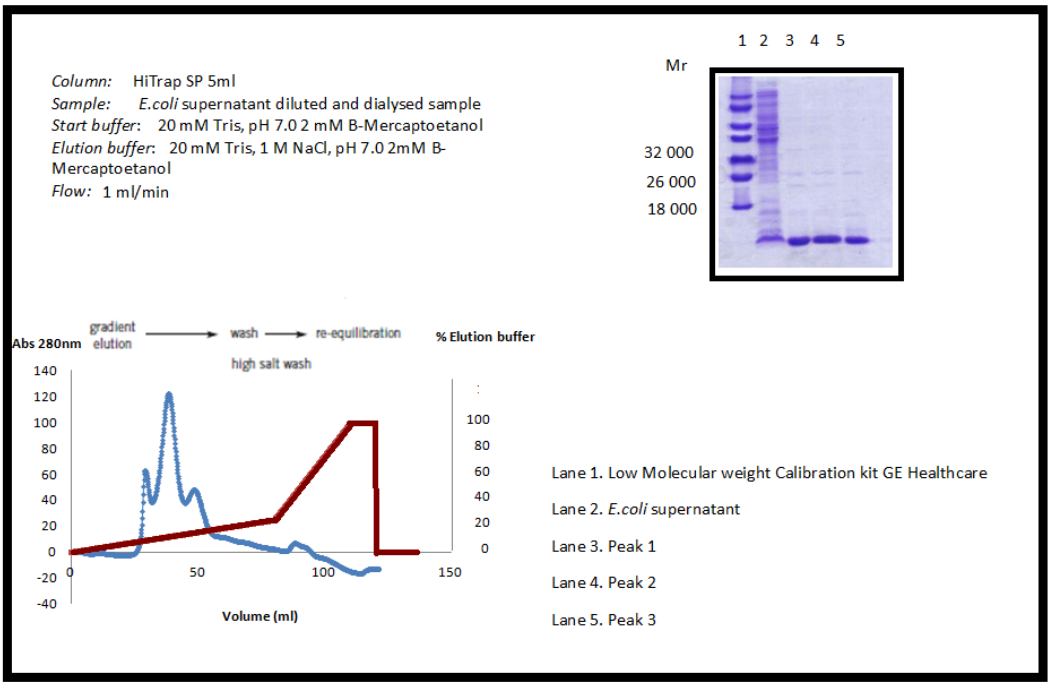

Figure 9: Cation exchange chromatography using a linear gradient of NaCl to elute the FeoA protein. The UV (protein/blue) and conductivity (salt/red) traces show the elution of protein peaks and the changes in salt concentration during elution. An SDS-PAGE (Tris-Gly 15%) shows the presence of our target protein in the three peaks.

The elution profile of FeoA protein shows the existence of three peaks. These three peaks belong to the FeoA protein as was demonstrated by the SDS-PAGE analysis. These different forms of FeoA differ by charge since they elute at different salt concentration.

The columns used in this cation exchange chromatography (HiTrap SP 5ml) have some degree of resolution that allows them to separate proteins that differ by small charge differences.(23) Even if the differences shown by FeoA protein are minimal they will be detected by this columns.

These charges differences cannot be explained by the presence of one cystiene residue in the C-terminal of FeoA protein (Figure 10). The uses of high levels of B-Mercaptoetanol maintain this residue in their reduced form.

```
10         20         30         40         50         60
MQYTPDTAWK ITGFSREISP AYRQKLLSLG MLPGSSFNVV RVAPLGDPIH IETRRVSLVL

        70
RKKDLALLEV EAVSC
```

Figure 10: The FeoA sequence from *E.coli* K12 strain.

In order to understand the nature of the different forms of FeoA protein a size exclusion chromatography was performed. The three peaks were injected into the size exclusion column, to observe if the difference could be related to different oligomeric states (Figure 11). The buffer used in the size exclusion is similar to the buffer used to elute the target protein, with the appropriate salt concentration (150mM NaCl).

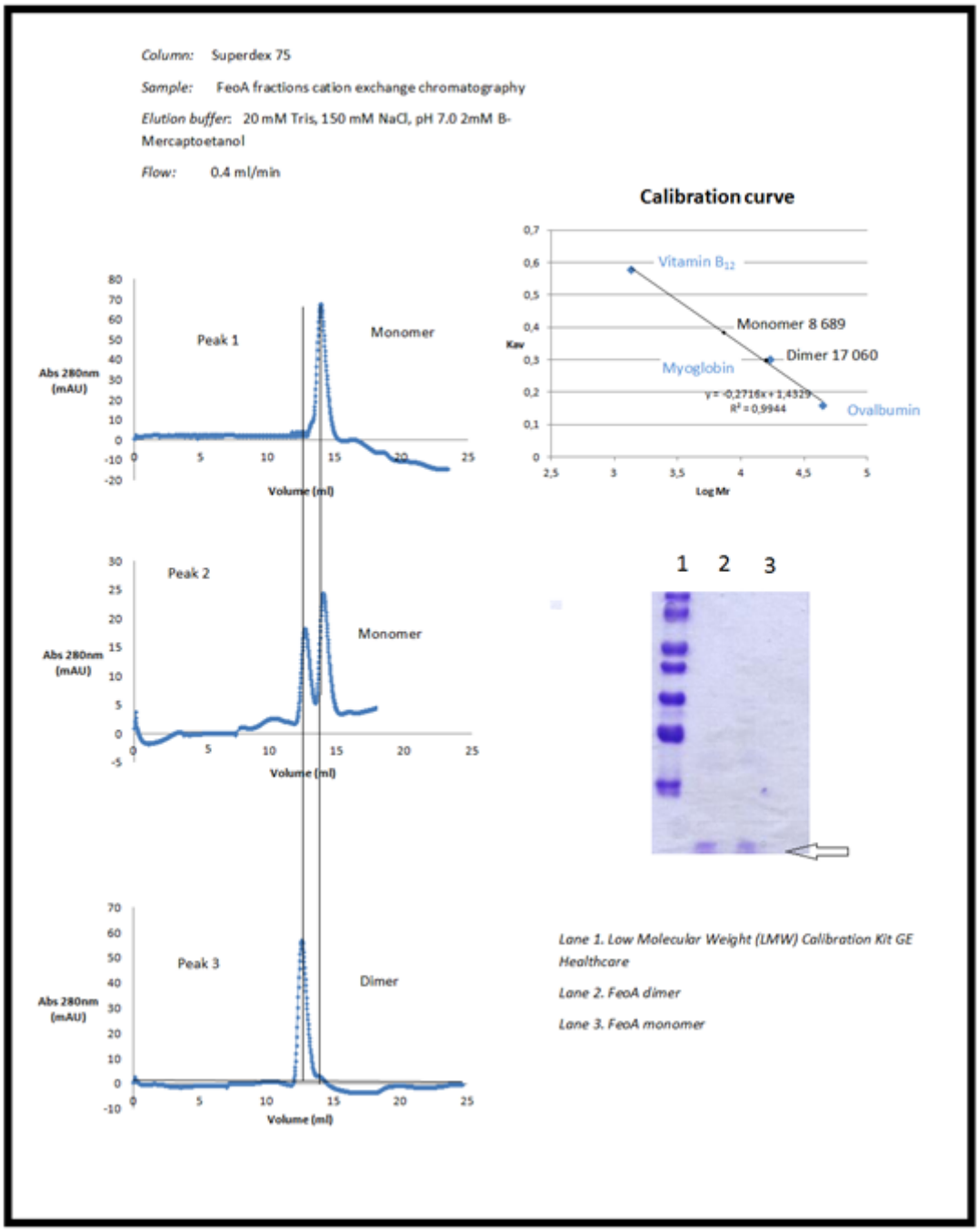

Figure 11: Size exclusion chromatography of the three peaks collected in the previous ion exchange chromatography, the corresponding SDS-PAGE (15%Tris-Gly) of the dimer and monomer fractions and the calibration curve of the Superdex 75 column. The UV (protein/blue) trace show the elution of protein peaks during the chromatography.

With the use of the size exclusion technique we could see the existence of two oligomeric states of our target protein, a dimeric (17 060Da) and a monomeric state (8 689Da). The peak 1 seen in the chromatogram of cation exchange chromatography includes the monomeric state, the peak 3 the dimeric state and the peak 2 the simultaneous presence of dimer and monomer of FeoA protein**.** Even the relative

absorbance at 280 nm is higher in the peaks that contain the dimeric state such as the peak 2 and 3, in agreement with a larger concentration of protein on those fractions.

The nature of the dimeric state is proven not to be related to the disulfide bonds mediated by the cysteine residues since the presence of high amounts of reducing agents excludes that hypothesis.

The nature of interaction of the FeoA homodimer could be related to electrostatic interactions, hydrophobic interactions or the simultaneous co-elution of a cofactor essential for the dimerization. It can be seen by size exclusion chromatography that the dimer and monomer forms are influenced by the overall protein concentration (Figure 12). The increase in total protein concentration favors the existence of the dimer form and the dilution favors the monomer state, although at not the same intensity.

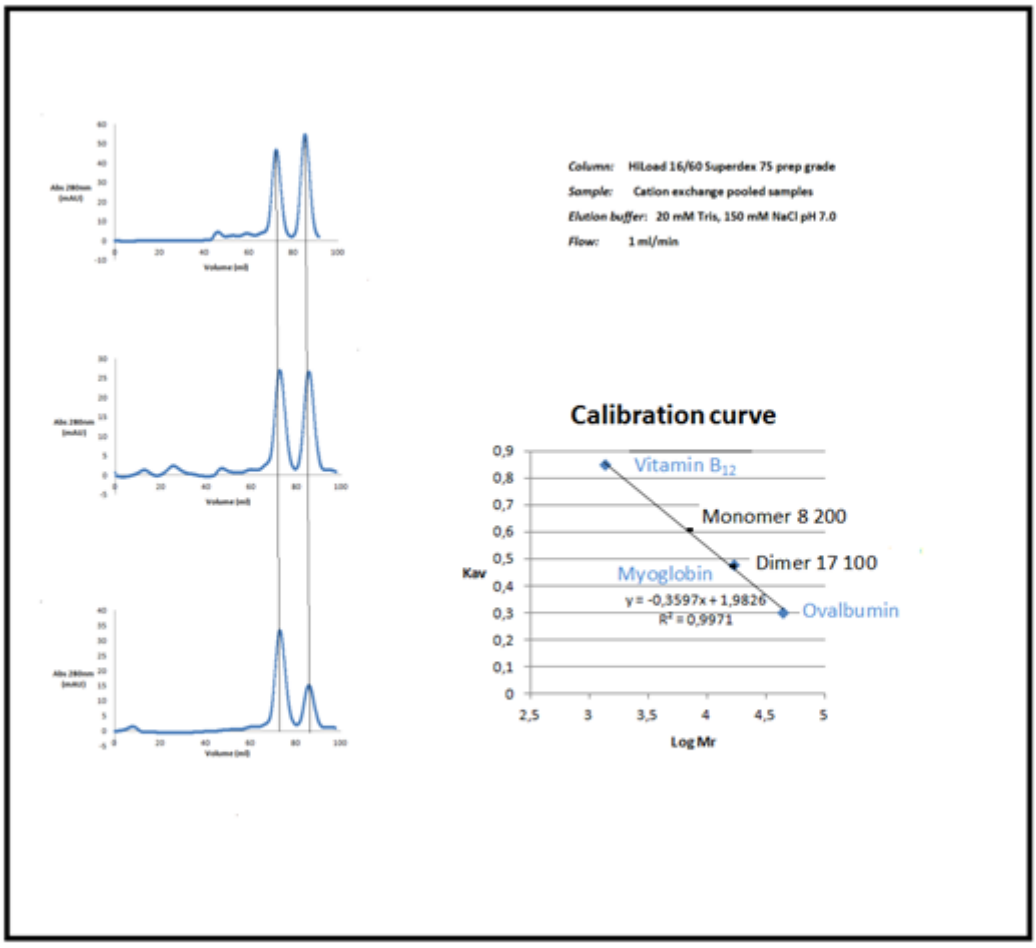

Figure 12: Size exclusion chromatography of different solutions with different concentration and the calibration curve of the HiLoad 16/60 Superdex 75 prep column. The UV (protein/blue) trace shows the elution of protein peaks during the chromatography.

Proeminent form	Concentration	Area (Dimer) mAU*ml	Area (Monomer) mAU*ml
Dimer	3,2 mg/ml	144	38
Dimer/Monomer	0.8 mg/ml	48	39
Monomer	0.2 mg/ml	13	33

From this experiment it was verified that the concentration affects the overall proportion among the dimer and monomer.

Although the nature of the dimer and monomer is not understood it is clear from this experiment that the concentration plays a key role. The question then arose on why was it possible to isolate the different forms after the ion exchange in order to inject them in the gel-filtration (Figure 11).

The different forms were eluted at different protein and salt concentration in the ion exchange chromatography and those specific conditions allowed the isolation and stability of the different forms.

It is known that high salt concentrations favour hydrophobic interactions, which could be responsible for the dimer formation. By analyzing one more time the ion exchange chromatography (Figure 9) we verify that the dimer elutes at a greater concentration of NaCl (150mM) than the monomer (90mM). It was also performed a size exclusion chromatography with no NaCl present and there was no presence of dimer or monomer just aggregation and degradation of the sample. (Figure 13)

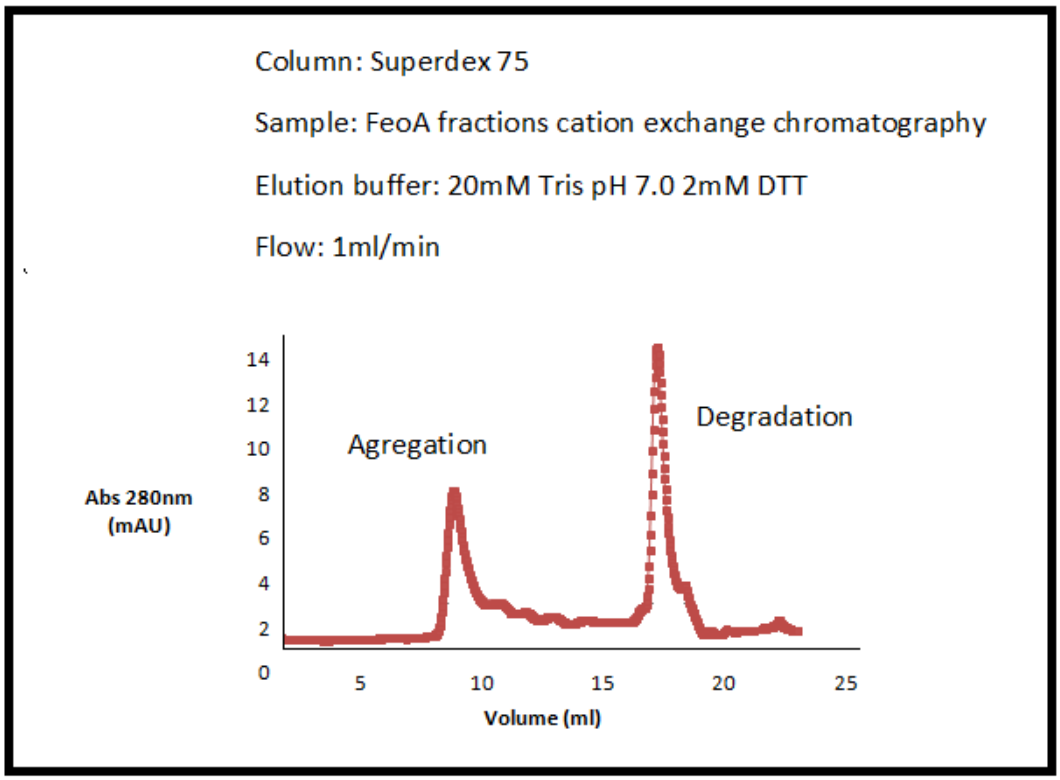

Figure 13: Size exclusion chromatography performed without the presence of salt. The UV (protein/blue) trace shows the elution of protein peaks during the chromatography.

Also the concentration of the FeoA protein in a buffer with no salt present showed aggregation and was not suitable for NMR analysis. All of this results determined that the solution should always include 150 mM of NaCl. The protein requires a certain salt concentration in order to be stable.

The isolation of the FeoA monomer was never possible since the process of protein concentration always produces a final sample were the dimer is the predominant form. The only sample produced for NMR with the monomer form was very diluted (10µM) due to degradation. (Preliminary results done by Meire Almeida)

The dissociation constant of the dimer was studied by gel-filtration, but it was never determined due to difficulties in analyzing the results at low protein concentration due to baseline problems.

By comparing the SDS-PAGE of FeoA protein after the cation exchange chromatography and the chromatogram of the size exclusion it is clear that the second purification step is crucial to eliminate remaining impurities (Figure 14).

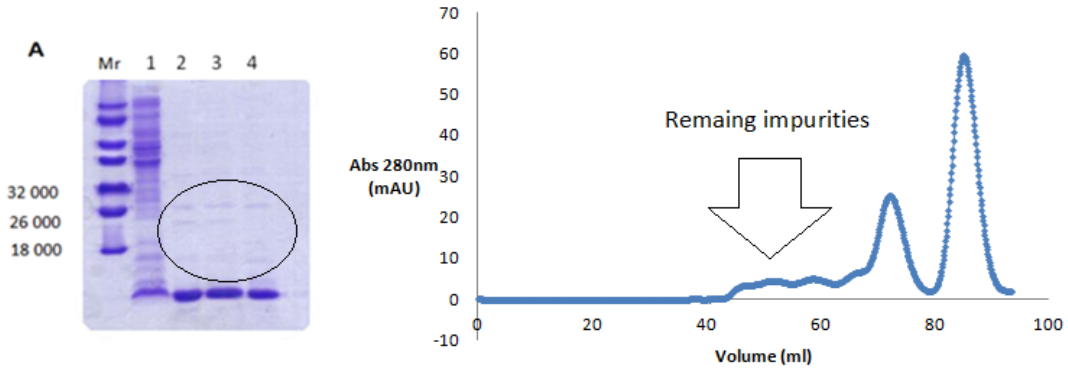

Figure 14: Comparison between the SDS-PAGE (Tris/Gly 15%) after the cation exchange (A) and size exclusion (B) chromatography in terms of presence of impurities.

The SDS-PAGE of the FeoA samples after the cation exchange chromatography shows the co-elution of impurities together with our target protein and the separation of some of those impurities in the size exclusion chromatography.

Although samples of sufficiency purity were produced after the cation exchange chromatography, the gel filtration removed some of the impurities that lead to the protein degradation seen in preliminary NMR spectra (previous work done by Meire Almeida).

This SDS-PAGE and the preliminary NMR results proved the need for an extra purification step.

The NMR protein sample has to have a high level of purity, not just to allow the proper analysis of the NMR spectra but also because it increases the stability of the target protein for longer periods since it eliminates impurities that degrade or interact with our sample affecting the NMR analysis.

The previously used size exclusion chromatography can act as a polishing step to achieve final purity by removing the remaining traces of impurities and closely related proteins that affected the NMR analysis **(30).**

The use of this technique demandes the concentration of the sample prior to gel filtration in order to minimize sample volume and facilitate a rapid high resolution size separation.

Because of that, efforts were made to obtain a single peak in the cation exchange chromatography, that would translate in a more concentrated solution of FeoA and a lower volume to inject in the next purification step.

This was accomplished with one single change in the cation exchange protocol, the absence of reducing agent (Figure 15). This oxidative conditions created a single peak in opposition to the three peaks present at reducing conditions.

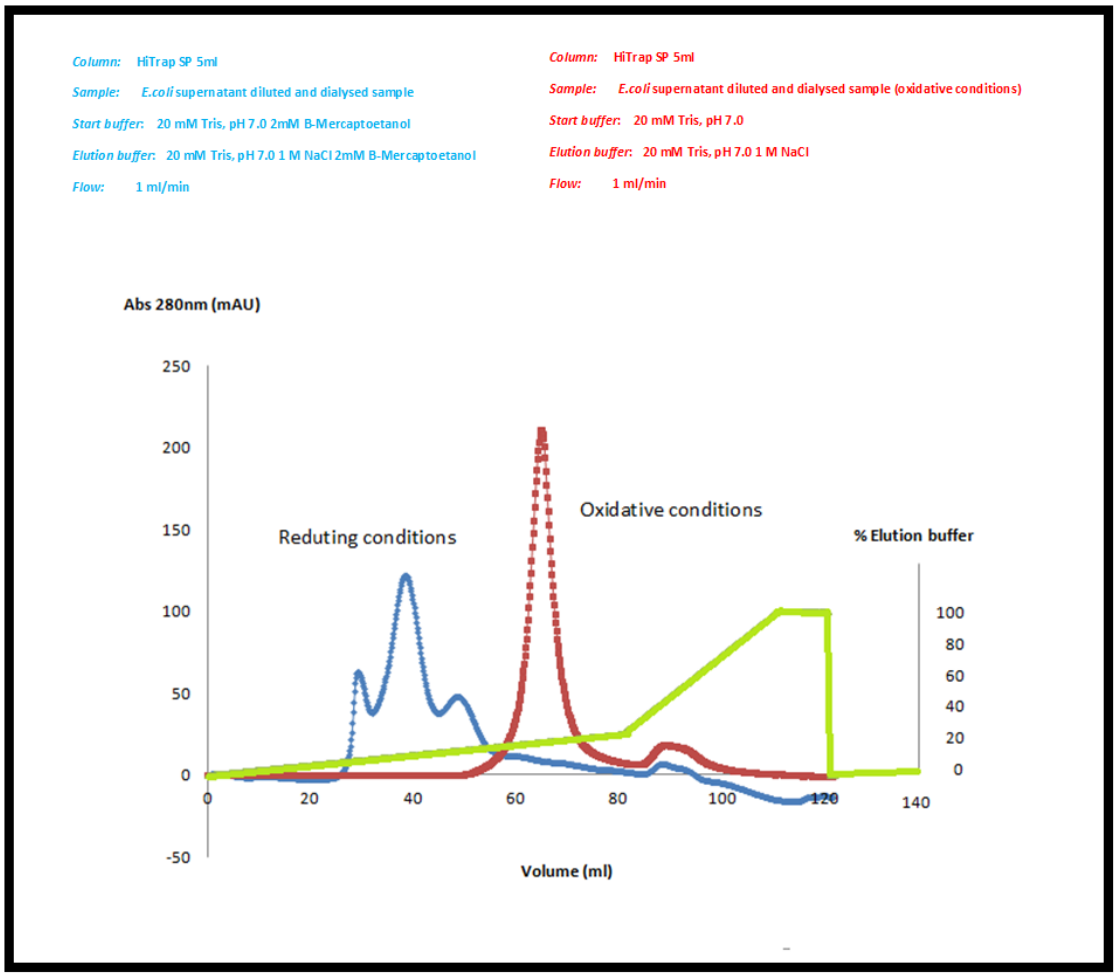

Figure 15: Comparison between the cation exchange chromatography performed in reducing conditions (2mM B-Mercaptoetanol) and at oxidative conditions (no B-Mercaptoetanol).

The optimization of the purification scheme allowed the production of a stable and pure FeoA sample. The cation exchange chromatography acted as a first capture step where the majority of impurities were eliminated. A second polishing step created a final sample with a high purity level that was the condition for the stability at long term of the protein sample.

With the optimization of the purification scheme and the evaluation of the best conditions to stabilize the purified protein the scheme was tested in a non labelled sample, to assess the overall reproducibility and protein yield.

In the Figure 16 it is shown the complete purification scheme for a non labelled sample and the final yield.

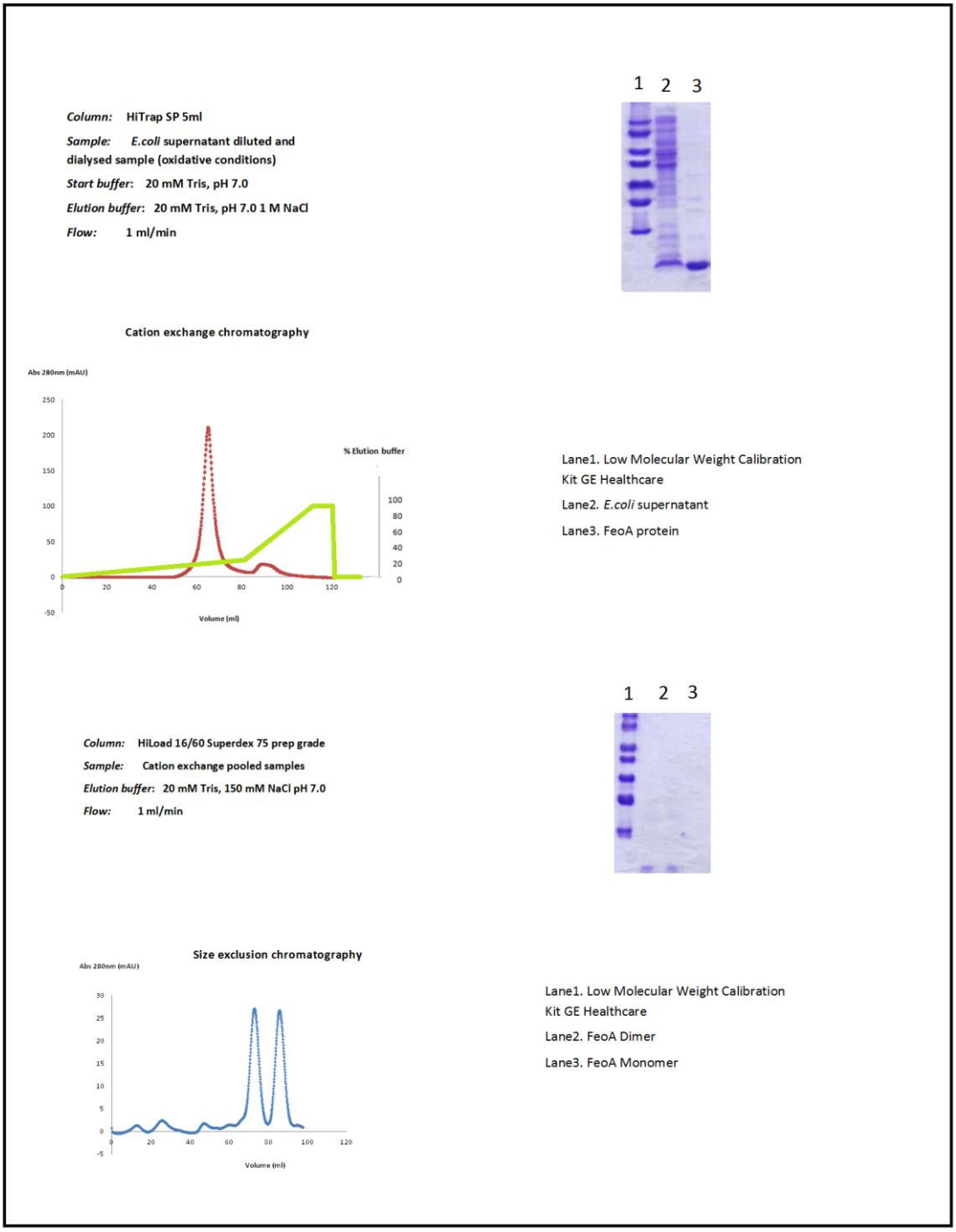

Figure 16: Final optimized purification scheme for the unlabelled FeoA protein.

Unlabelled FeoA	Concentration (mg/ml)	Volume (ml)	Concentration (mM)	Protein yield (mg protein/LCulture)
After Concentration	6,2	0.5	0.74	3,1

Figure 17: Final yield for a two liter culture necessary to achieve the necessary protein concentration for NMR. The final volume of 500 µl is the final volume used in the NMR tubes.

3.2.2. Purification of the single labelled ^{15}N and double $^{13}C/^{15}N$ labelled forms of FeoA

With the optimization of the purification scheme complete, the production of isotopically labelled samples was achieved for further NMR analysis.

In figure 3.12 and 3.13 it is shown the purification results of the single labelled ^{15}N and double labelled $^{13}C/^{15}N$ sample and the respective protein yield.

The purification results are shown not because the labelled proteins behave different from the non labelled, but to show the reproducibility of the purification scheme and the final protein yield.

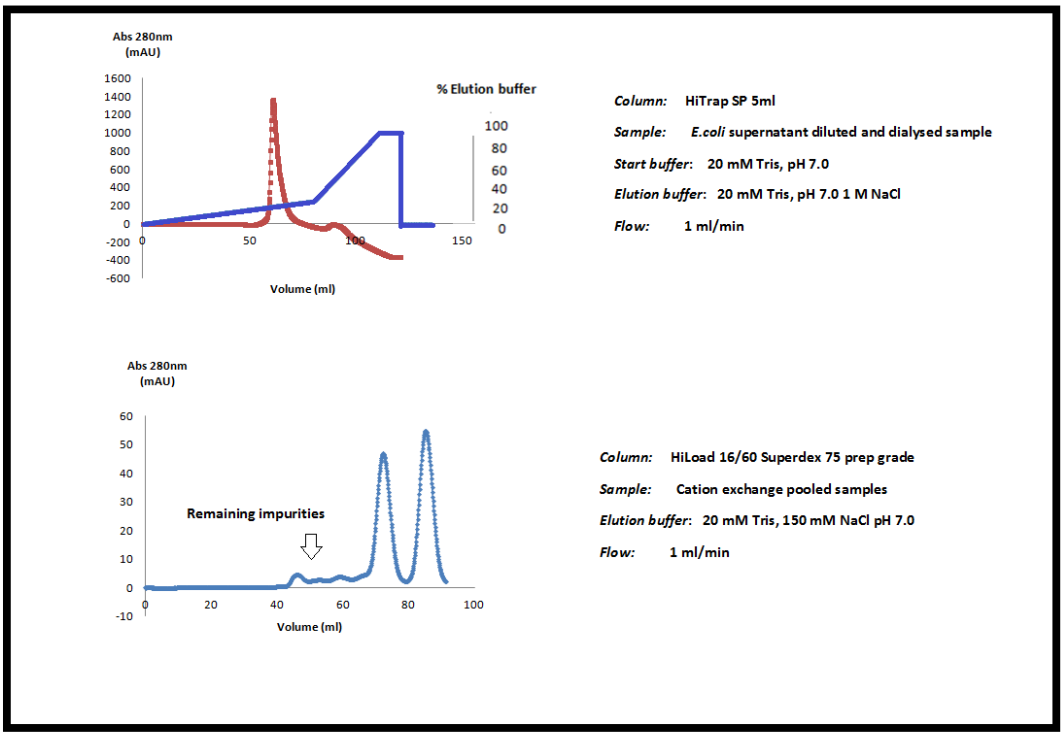

Figure 18: Purification scheme of the single labelled ^{15}NFeoA.

^{15}N FeoA	Concentration (mg/ml)	Volume (ml)	Concentration (mM)	Protein yield (mg protein/L Culture)
After Concentration	6.0	0.5	0.711	3.0

Figure 19: Yield of the single labelled ^{15}N FeoA.

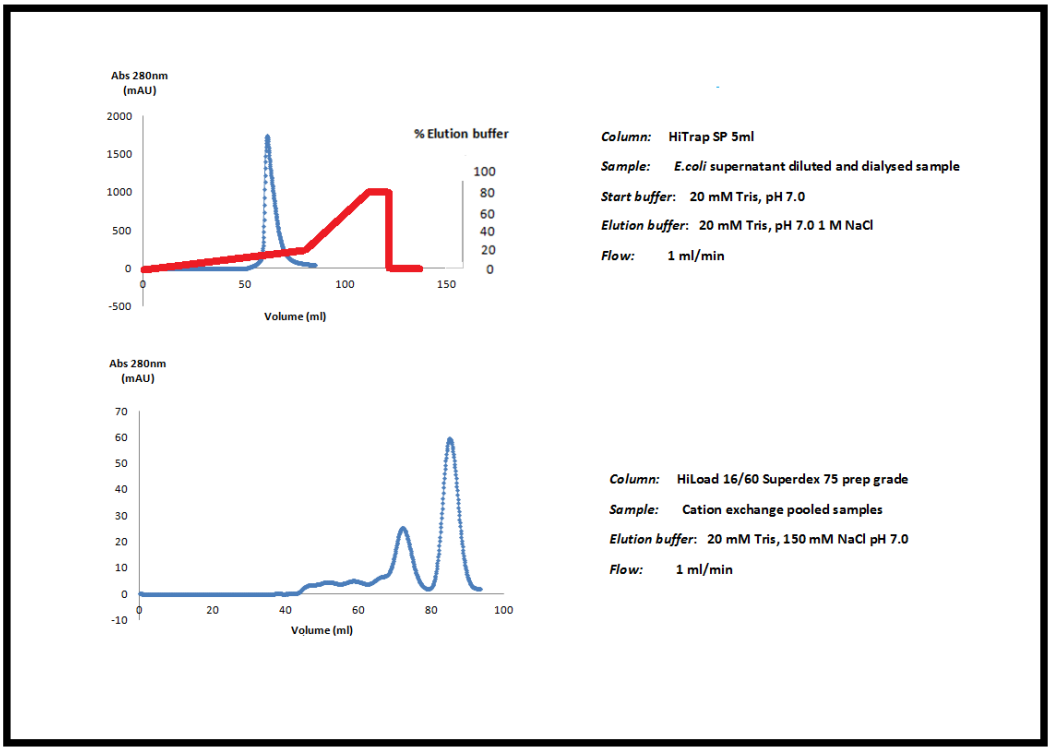

Figure 20: Purification of the double labelled $^{15}N/^{13}C$ FeoA.

$^{15}N/^{13}C$ FeoA	Concentration (mg/ml)	Volume (ml)	Concentration (mM)	Protein yield (mg protein/L Culture)
After Concentration	5.4	0.5	0.64	2.7

Figure 21: Yield of the double labelled $^{15}N/^{13}C$ FeoA.

3.3.- NMR data analysis

With the different protein samples produced for NMR we were able to learn more about our samples, their stability and the structure of the protein. Initial investigation of the samples 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.

First, we studied the sample stability using the non labelled sample for the initial steps of the optimization process due to its lower cost.

Figure 22 is showing the overlay of the **^{1}H 1D** -spectrum at day one and day seven of a sample that remained at 25^{o}C. The peaks overlay very well, proving that the conditions of the sample (buffer composition, salt concentration, pH value) are suitable for a long term NMR analysis such as a week period. Closer analysis of the amidic (H^{N}) region (protons attached to the nitrogen atoms) shows that they are very well dispersed. This is a strong indication that the protein is folded and not denaturated. In addition the presence of peaks in the region between 4.7 and 6.5 ppm that are attributed to Hα protons was strongly indicative of the presence of beta sheet secondary structure in the protein, an observation that is consistent with models of feoA.

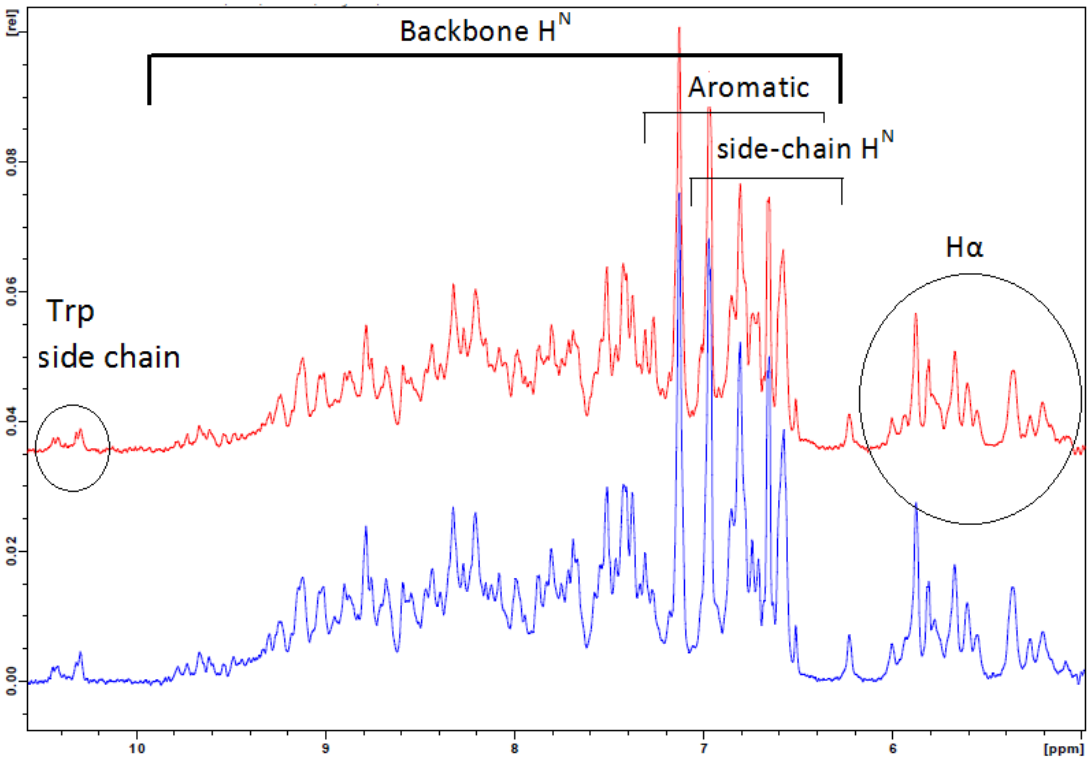

Figure 22 Display of ^{1}H 1D spectra of FeoA protein at day one (blue) and at day seven (red). The typical Backbone H^{N} region is shown as well as the Hα from the β-sheets.

Further investigation of the sample was possible with the use of the **^{1}H-^{15}N-HSQC** spectra. These spectra contain one cross-peak (chemical shift) for each ^{1}H-^{15}N pair connected by a single covalent bond. This peak results from the magnetization transfer from the hydrogen to the attached ^{15}N nuclei via the J-coupling (Figure 23). The spectrum is expected to show all the backbone amides (except prolines that are not protonated) as well as the side chain amides of Tryptophan, Asparagines and Glutamines. The Arginine side chain amides and Histidine are also sometimes visible on the **^{1}H-^{15}N-HSQC** spectrum.

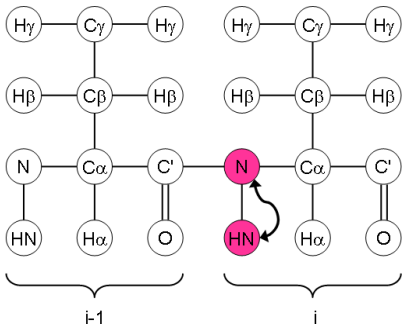

Figure 23: Scheme of the ^{1}H-^{15}N-HSQC 2D experiment magnetization pathway.

In Figure 24 we see the **^{1}H-^{15}N-HSQC** of a 0.64mM sample of FeoA at pH 7. Initial inspection of the spectrum indicates that FeoA is folded because the H-N peaks are spread throughout the spectra. An unfolded protein will have their peaks concentrated in the center of the **HSQC** spectrum.

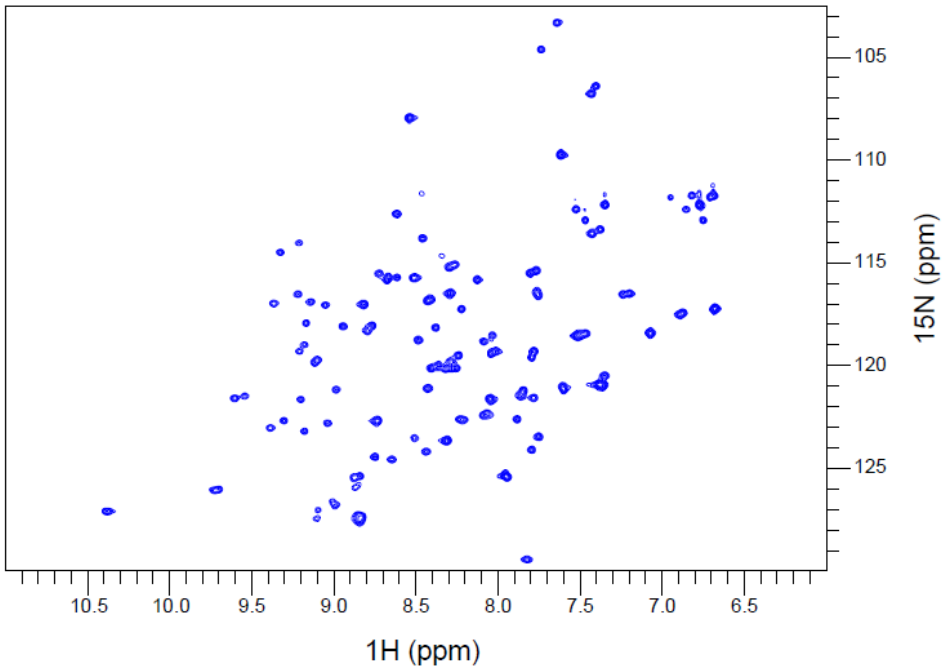

Figure 24:^{1}H-^{15}N-HSQC from a FeoA protein sample with a final concentration of 0.64 mM.

However it also revealed the presence of more peaks than the ones predicted for the 75 aminoacid protein. The number of expected peaks in a **^{1}H-^{15}N-HSQC** is given by the total number of the aminoacids minus the prolines that do not contain a protonated amide. It also contains a pair of peaks originating from the side-chains of the glutamines, asparagines and one for each side-chain of a tryptophan. In the case of the FeoA protein it was expected a total of 77 peaks but the final spectrum contained about 120 peaks. This increased number of peaks relative to the expected is not common for a sample that contains only one form of the target protein.

Given our size exclusion results, where we observed the simultaneous presence of the monomeric and the dimeric forms of feoA, we interpreted our NMR findings as those belonging to a slow exchanging mixture of monomer and dimer.

To verify this observation, we collected another **^{1}H-^{15}N-HSQC** spectrum on a 10µM sample. The resulting spectrum contained a significantly smaller number of peaks,

consistent with a monomer. Figure 25 shows the overlay of the concentrated and the diluted samples.

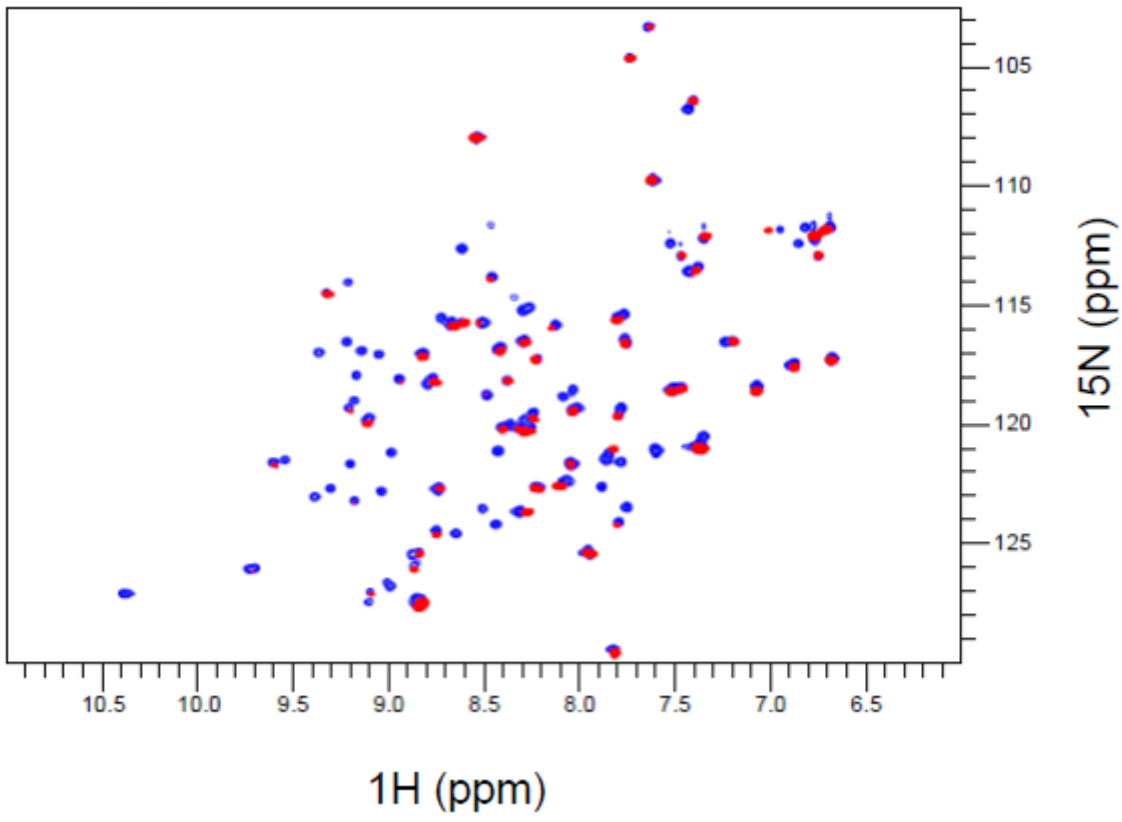

Figure 25: Overlay of ^{1}H-^{15}N-HSQC with the prevalence of the monomer form 10µM (red) and the prevalence of both forms 0.64mM (blue).

However, while we were able to collect **^{1}H-^{15}N-HSQC** data at that low concentration, it was not possible to use these conditions for any more complicated spectra. As we can see in Figure 25, although we have multiple forms in solution, the resulting peaks are still very well dispersed with relatively small overlap. For this reason we decided to use the more concentrated sample for data collection and then to use the monomer spectrum to help us distinguish between the monomer and the dimer peaks.

With the two different **^{1}H-^{15}N-HSQC** it was possible to distinguish between the monomer and dimer peaks. Some chemical shifts did overlay presenting no changes but a significant number of resonances did change. The correlation of the monomeric and the dimeric peaks was possible with the use of the **HNCO** and the **^{15}N-NOESY** spectra. While the CO resonances of each aminoacid were not identical in the monomeric and the dimeric forms, the pattern was very similar.

This part of the work enabled us to get a list of monomer-dimer pairs and to use only the monomer peaks in the next phase of the analysis. The monomer peaks were the

only ones used in the assignment since our goal was to determine the structure of FeoA as a monomer and only in the end to analyze the dimer structure.

3.3.1-Backbone assignment

Having decided a strategy for the interpretation of the spectra, we started the more detailed analysis of our NMR data.

The first step to determine the FeoA structure was to identify the observed signals in the different NMR spectra and associate them with the atoms that they correspond to. This is called the assignment process. Once the process of assignment is complete the information will be used as a starting point for the structure calculation of FeoA protein.

At this stage the double labelled sample allowed the recording of many 3D spectra that simultaneously used the information of ^{1}H, ^{15}N and ^{13}C dimensions for the backbone and side chain assignment.

The most common strategies for resonance assignment use the **^{1}H- ^{15}N-HSQC** as an initial point. Starting from the H-N group it is possible, using different experimental schemes, to make correlations between neighbouring aminoacids using the information from either the directly attached carbonyl CO or the Cα. The experiments involving the CO allowed us to obtain information about the previous aminoacid (i-1), while those involving Cα, allowed us to correlate resonances within the same aminoacid (i) as well as those of the previous aminoacid (i-1).

In summary the first stage of the assignment process which is called sequential assignment uses these different correlations of the previous aminoacid and current aminoacid to identify pairs of peaks that correspond to sequential aminoacids (Figure 26). Repetition of this process results in the linking of the **^{1}H-^{15}N-HSQC** peaks in chains corresponding to the primary sequence of the protein.

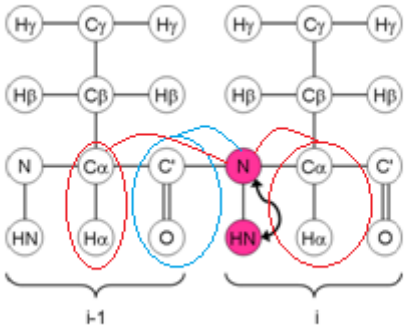

Figure 26: Scheme showing the correlations of the previous aminoacid CO (blue) and current aminoacid Cα as well as the previous aminoacid (red).

The detailed procedure that we followed to achieve this is the following.

The backbone assignment used the **^{1}H-^{15}N-HSQC** H-N correlations as a guide to extract the information from triple resonances experiments such as the **HNCO**, **HN(CA)CO**, **HN(CO)CACB** and **HNCACB** spectra. Figure 27 shows the graphical interface of the program CARA that allowed as to link these resonances. As can be seen, the **^{1}H-^{15}N-HSQC** acts as a guide for the analysis of the 3D data that are displayed in the column to the left. These columns are located on the Z axis centered above each amide peak.

The first stage of the analysis utilized the spectra **HNCO** and **HN(CA)CO.**

The **HNCO** (Figure 27) allowed the assignment of the CO_{i-1} and the **HN(CA)CO (**Figure 28**)** the assignment of CO_{i-1} and CO_i.

The **HNCO** (Figure 27) was used to assign each NH_i with the previous CO_{i-1}, in backbone correlations. Magnetisation is passed from ^{1}H to ^{15}N and then selectively to the carbonyl ^{13}C via the $^{15}N^{H}$-^{13}CO J-coupling. Magnetisation is then passed back via ^{15}N to ^{1}H for detection. The chemical shift is evolved on all three nuclei resulting in a three-dimensional spectrum.

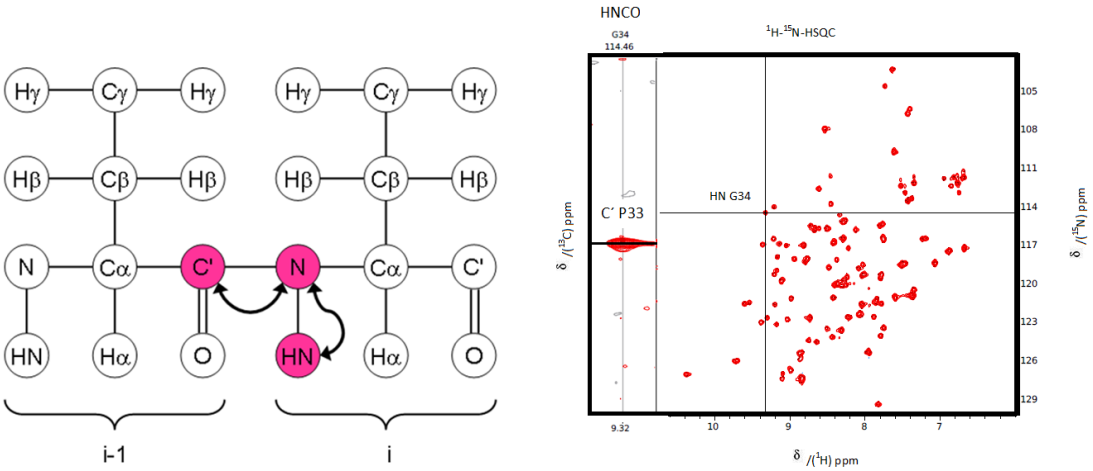

Figure 27: Scheme of the HNCO 3D magnetization pathway and the corresponding FeoA HNCO.

In a similar way the **HN(CA)CO** 3D spectrum (Figure 28) also provided NH-CO correlations, but in this experiment the magnetisation is transferred from ^{1}H to ^{15}N and then via the N-Cα J-coupling to the ^{13}Cα. From there it is transferred to the ^{13}CO via the ^{13}Cα-^{13}CO J-coupling. For detection the magnetisation is transferred back the same way: from ^{13}CO to ^{13}Cα, ^{15}N and finally ^{1}H. The chemical shift is only evolved on ^{1}H, ^{15}N and ^{13}CO and not on the ^{13}Cα. The result is a three-dimensional spectrum.

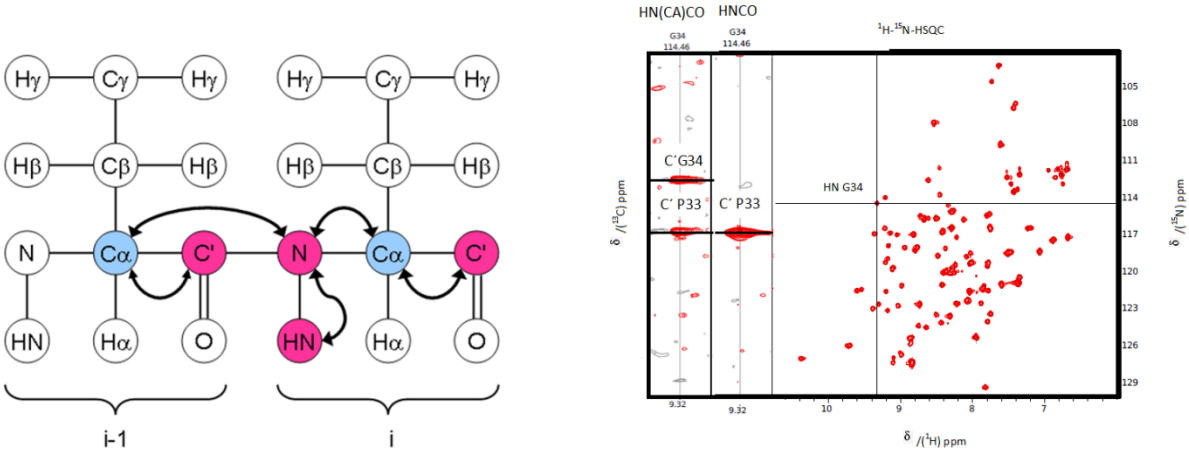

Figure 28: Scheme of the HN(CA)CO 3D magnetization pathway and the corresponding FeoA HN(CA)CO.

As can be seen in the FeoA **HN(CA)CO** spectrum (Figure 28) each H-N group is associated with two carbonyl groups. This is because the amide nitrogen is coupled both to the Cα of its own residue(i) and that of the preceding residue(i-1).

But because the coupling between N_i and $C\alpha_i$ is stronger than that between N_i and $C\alpha_{i-1}$, the H_i-N_i-CO_i peak generally ends up being more intense than the H_i-N_i-CO_{i-1} peak.

With this two sets of spectra it was possible to identify the CO_i and the CO_{i-1}. It was then possible to lay out all the 3 dimensional strips originating from each H-N in such a way that the CO_i -resonances of one H-N were matched with the CO_{i-1} of another. By repeating this process we could sequentially link the different H-Ns as seen in Figure 29.

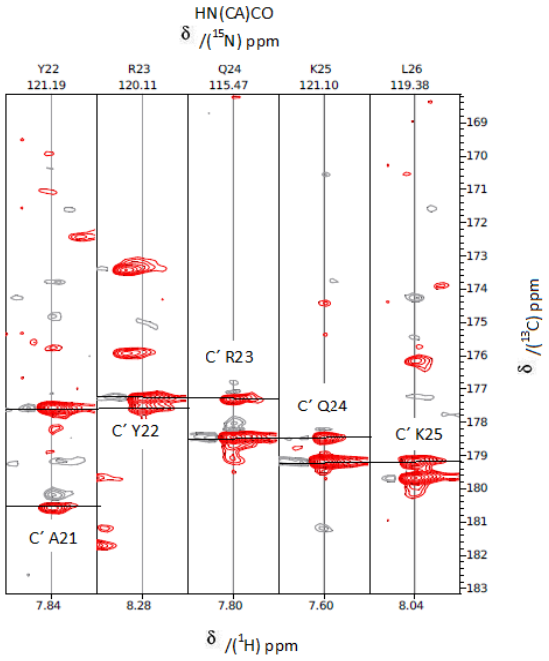

Figure 29: Procedure for the sequential alignment of HN(CA)CO strips.

The **HNCO/HN(CA)CO** set of experiments is often not enough for the identification of all the sequential resonances because carbonyls cannot be used to identify the chemical nature of the different amino acids since there is no good correlation between carbonyl chemical shifts and amino acid type. As can be seen in Figure 30 when taking into account the error bars the variation in the chemical shift of the CO groups in the different aminoacids falls in the same range, making it impossible to be used to distinguish between different aminoacids. To be able to match the fragments we have obtained so far to specific amino acid fragments of the primary sequence,

another set of experiments was needed. These were the **HNCACB** and **HN(CO)CACB** that allowed us do assign the Cα and Cβ atoms . Contrary to the CO, Cα and Cβ atoms can be used in identifying the different aminoacids. As can be seen on the Figure 30 the Cα and Cβ are specific of each aminoacid.

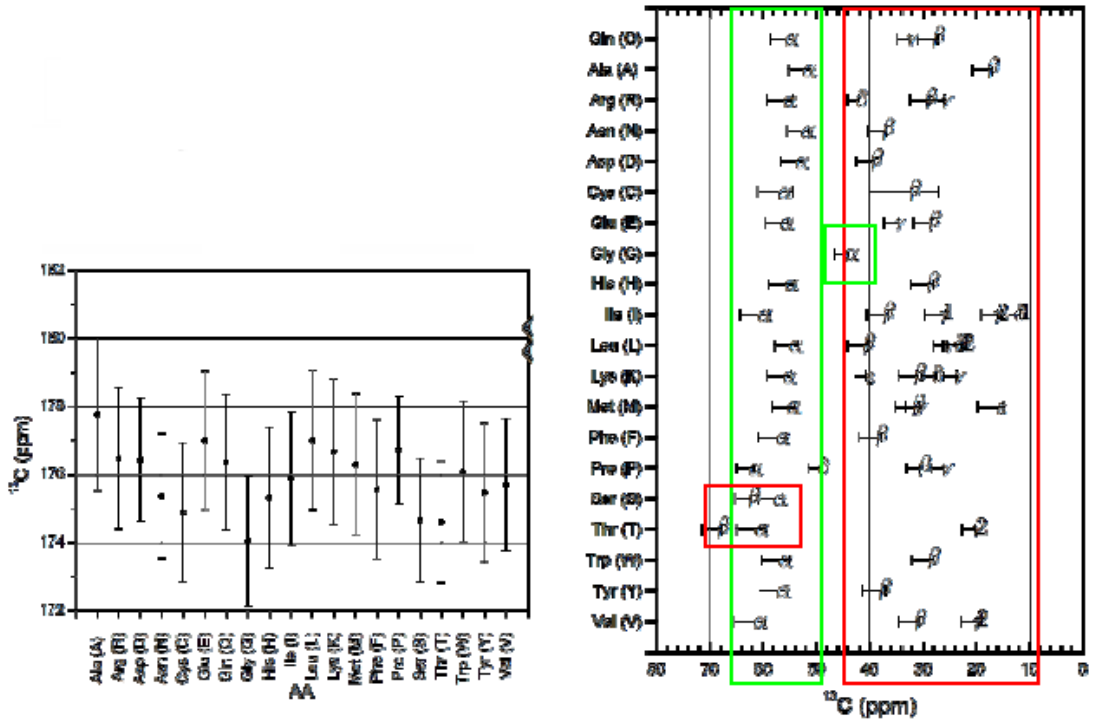

Figure 30: Chemical shift variation for the CO, Cα and Cβ in the different aminoacids.

In the **HN(CO)CACB** 3D spectrum (Figure 31) the magnetisation is transferred from ^{1}Hα and ^{1}Hβ to ^{13}Cα and ^{13}Cβ, respectively, and then from ^{13}Cβ to ^{13}Cα. From here it is transferred first to^{13}CO, then to ^{15}N^{H} and then to ^{1}H^{N} for detection. The chemical shift is evolved simultaneously on ^{13}Cα and ^{13}Cβ, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are ^{15}N^{H} and ^{1}H^{N}. The chemical shift is not evolved on ^{13}CO.

As can be seen in the FeoA **HN(CO)CACB** each NH peak in the **^{1}H-^{15}N-HSQC** is associated with the chemical shift information of the ^{13}Cα and ^{13}Cβ carbons of the preceding residue (residue i-1). Each strip of this spectrum will contain two peaks, the^{13}Cα and the^{13}Cβ of the preceding NH group.

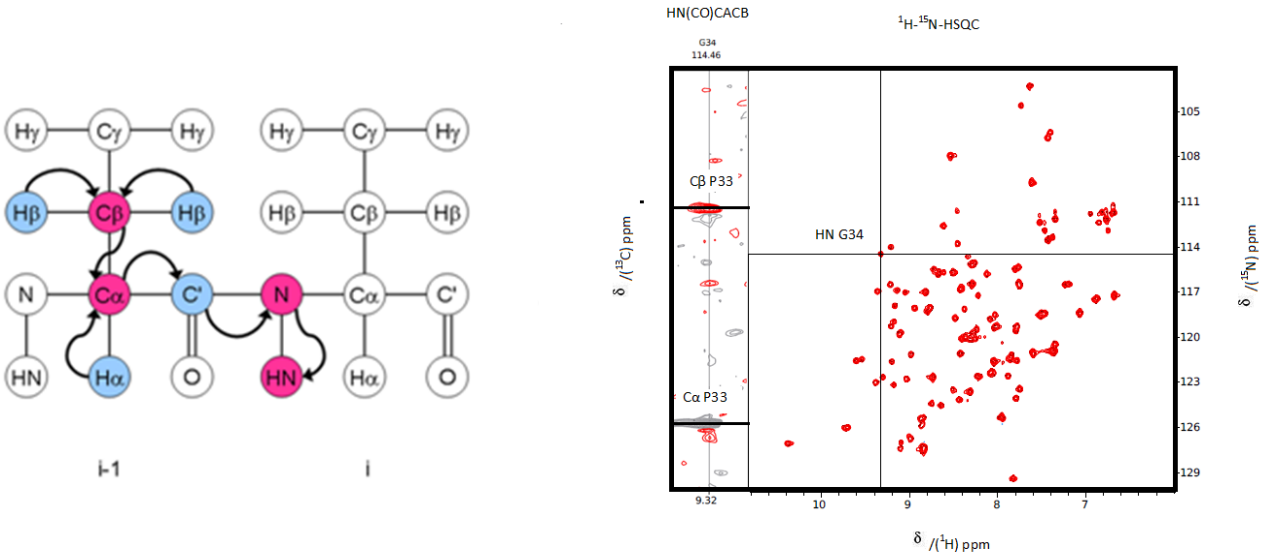

Figure 31: Scheme of the HN(CO)CACB 3D magnetization pathway and the corresponding FeoA HN(CO)CACB.

In the **HNCACB** 3D spectrum (Figure 32) the magnetisation is transferred from ^{1}Hα and ^{1}Hβ to ^{13}Cα and ^{13}Cβ, respectively, and then from ^{13}Cβ to ^{13}Cα. From here it is transferred first to $^{15}N^{H}$ and then to $^{1}H^{N}$ for detection. Transfer from $C\alpha_{i-1}$ can occur both to $^{15}N_{i-1}$ and $^{15}N_{i}$, or viewed the other way, magnetisation is transferred to $^{15}N_{i}$ from both $^{13}C\alpha_{i}$ and $^{13}C\alpha_{i-1}$. Thus for each H-N group there are two Cα and Cβ peaks visible. The chemical shift is evolved simultaneously on ^{13}Cα and ^{13}Cβ, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are $^{15}N^{H}$ and $^{1}H^{N}$.

In the **HNCACB** spectrum (Figure 32) each strip contain ^{13}Cα and ^{13}Cβ from the preceding residue ($^{13}C\alpha_{i-1}$ and $^{13}C\beta_{i-1}$). Since the coupling between N_i and $C\alpha_i$ is stronger that of the N_i and $C\alpha_{i-1}$, the H_i-N_i-$C\alpha_i$ /H_i—N_i-$C\beta_i$ peaks will be stronger than the H_i-N_i-$C\alpha_{i-1}$/H_i-N_i-$C\beta_{i-1}$ peaks. Glycines residues are exception because they don't have Cβ. The Cα and Cβ are very easy to distinguish because they have opposite signs.

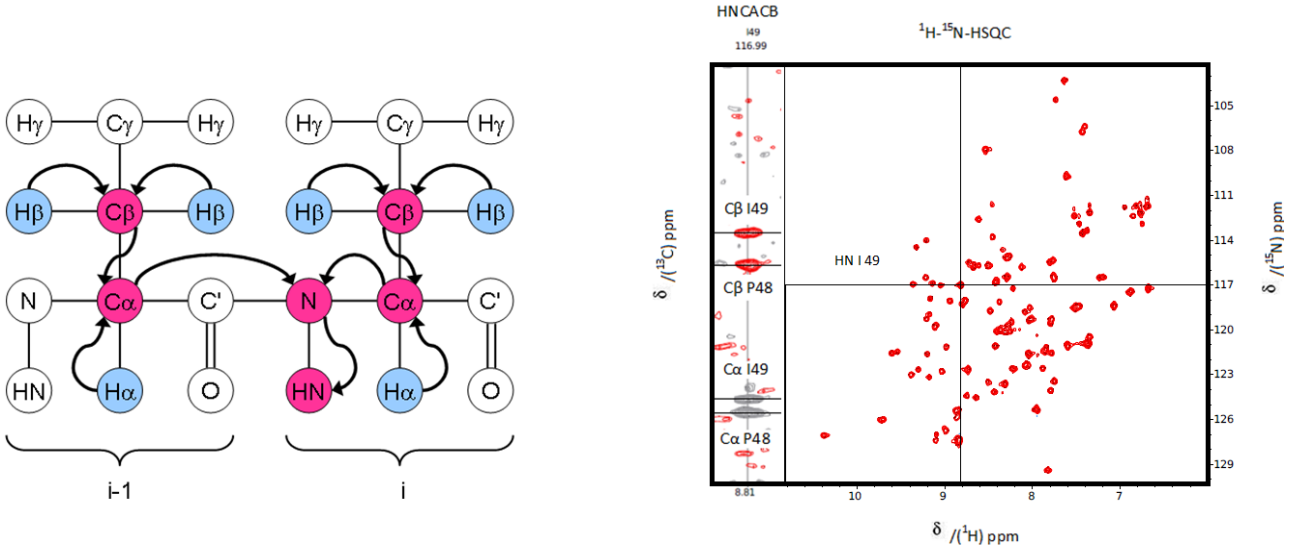

Figure 32: Scheme of the HNCACB 3D magnetization pathway and the corresponding FeoA HNCACB.

The analysis and the strip matching of the **HNCACB/HN(CO)CACB** spectra is done in the same way as with the **HNCO/HN(CA)CO** described earlier.

At the end of the analysis we are now able to link our fragments to specific aminoacids in the sequence. This is done in a semi-automated manner by the CARA software that was used in sequential backbone assignment. The identity of the resonances that composed this fragments was calculated by CARA through the information of Cα and Cβ that are characteristic of the aminoacid type. This information was extracted from the **HNCACB** and **HN(CO)CACB.**

At this stage we were able to obtain an **^{1}H-^{15}N-HSQC** spectrum where all the previously unknown peaks could be associated with specific aminoacids (Figure 33).

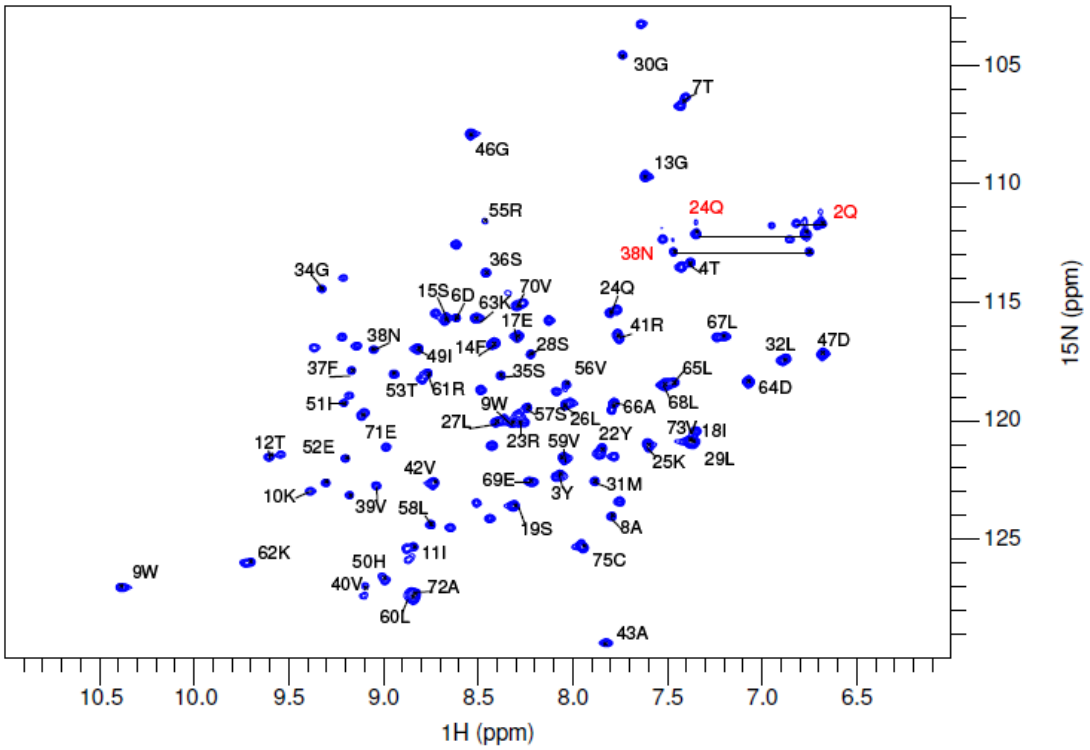

Figure 33: ^{1}H-^{15}N-HSQC where all the H-N peaks were assigned to a FeoA aminoacid. The red H-N group correspond to the side chain H-N groups of Glutamine and Asparagine.

3.3.2-Side chain assignment

When sequential backbone assignment was complete, the next step was to assign the remaining side chain atoms. For this five spectra were used: **HN(CO)HAHB**, **hCCH-TOCSY** and **^{13}C-HSQC**, as well as the **^{15}N-HSQC-NOESY** and the **^{13}C-HSQC-NOESY** for both aromatic and aliphatic side chain atoms.

In the FeoA **HN(CO)HAHB** **(**Figure 34**)** the magnetisation is transferred from ^{1}Hα and ^{1}Hβ to ^{13}Cα and ^{13}Cβ, respectively, and then from ^{13}Cβ to ^{13}Cα. From here it is transferred first to^{13}CO, then to $^{15}N^{H}$ and then to $^{1}H^{N}$ for detection. The chemical shift it not evolved on any of the carbon atoms. Instead, it is evolved on the ^{1}Hα and ^{1}Hβ, the $^{15}N^{H}$ and $^{1}H^{N}$. This results in a three-dimensional spectrum with one nitrogen and two hydrogen dimensions.

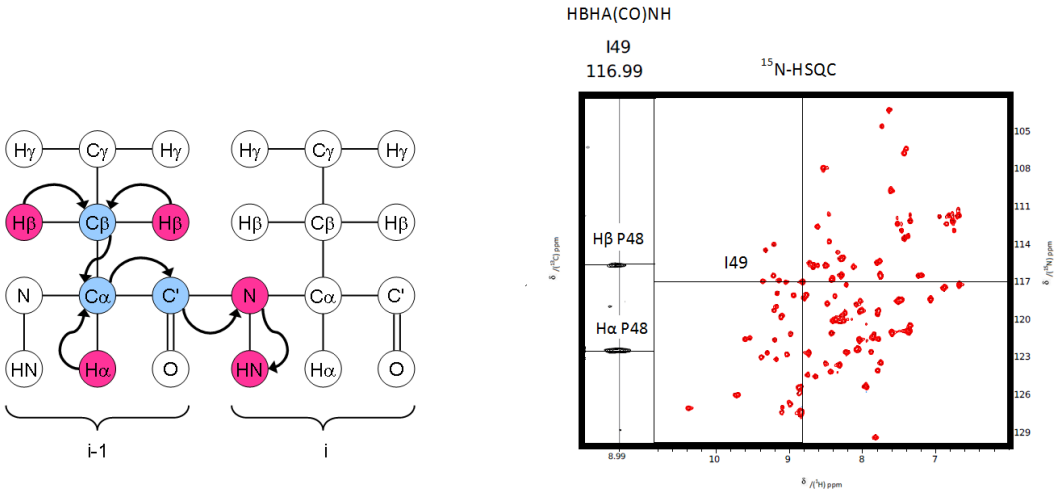

Figure 34: Scheme of the HBHA(CO)NH 3D experiment magnetization pathway and the respective FeoA HBHA(CO)NH.

With this spectrum each H-N strip contained links to the Hα and Hβ of the previous residue, allowing for the assignment of this side-chain atoms.

The Hβ and Hα chemical shifts obtained from the **HN(CO)HAHB** were then linked to their corresponding aminoacid based on the sequential assignment information that we already had. The remaining atoms of the side chains were obtained from the analysis of another spectrum the **hCCH-TOCSY**.

The **hCCH-TOCSY** experiment correlates all aliphatic ^{1}H and ^{13}C spins within residues, and is used to assign ^{1}H and ^{13}C resonances and connect the side-chain chemical shifts with the backbone assignments.

In the **hCCH-TOCSY** 3D spectrum (Figure 35) the magnetisation is transferred from the side-chain hydrogen nuclei to their attached ^{13}C nuclei. This is followed by isotropic ^{13}C mixing and finally transfer back to the side-chain hydrogen atoms for detection.

This spectrum was very useful since it showed, for each C-H strip, all the hydrogens atoms attached to all other carbon atoms in the aliphatic side-chain.

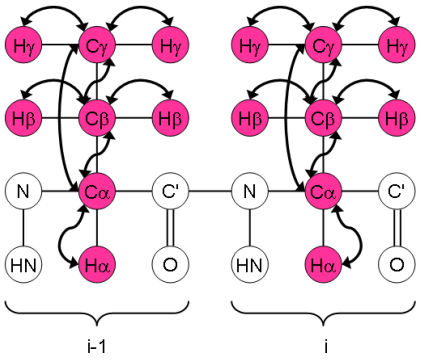

Figure 35: Scheme of the hCCH-TOCSY 3D experiment magnetization pathway.

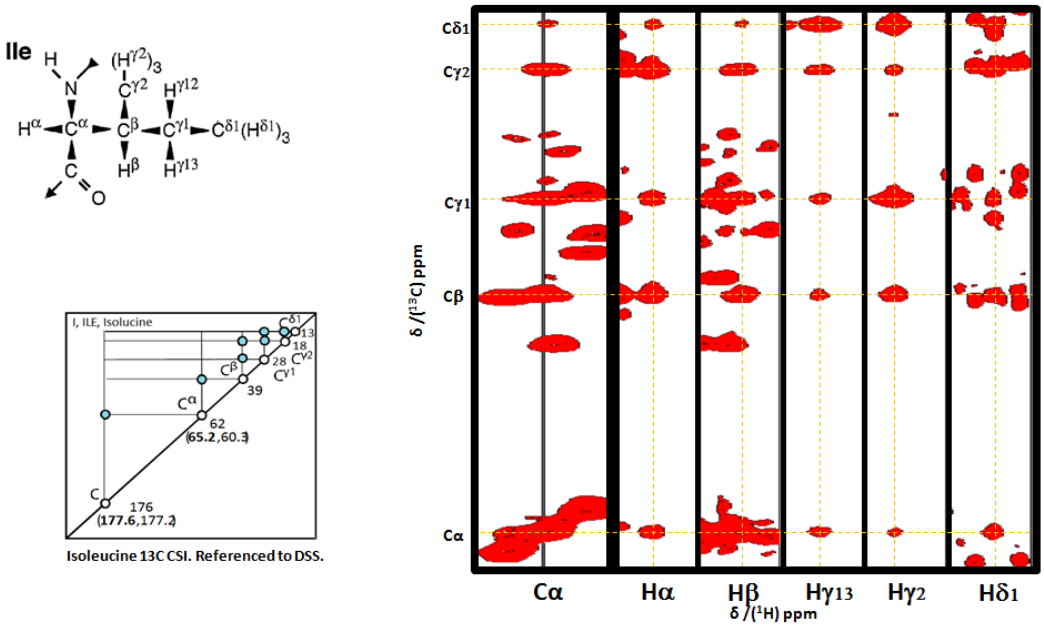

Figure 36: FeoA hCCH-TOCSY 3D spectrum. Isoleucine is here used as an example showing the correlations between carbon and proton resonances of the aliphatic chain.

With the use of chemical shift tables (Figure 36), it was possible to quickly assign all the side chain aliphatic protons on the **hCCH-TOCSY** spectra.

Finally it was time for the assignment of the aromatic side chain protons. This final task was 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** was connected to a carbon atom with the help of the **^{1}H-^{13}C-HSQC**.

In the **^{13}C-NOESY** the magnetization (Figure 37) is exchanged between all hydrogens via dipole-dipole coupling that cause the NOE effect. This phenomenon is distance dependent allowing us to correlate atoms through space. Then the magnetisation is transferred to neighbouring ^{13}C nuclei and back to ^{1}H for detection. Transfer either occurs to/from the aliphatic ^{13}C nuclei or to/from the aromatic ^{13}C nuclei (but not both) depending on the ^{13}C frequency used during the pulse sequence

This spectrum was used to obtain restraints for structure calculations. The **^{13}C-NOESY** can be centered either on the aliphatic or on the aromatic carbons. Here the spectrum was centered on the aromatic carbons for the assignment of the aromatic residues.

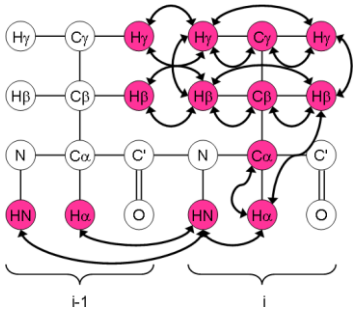

Figure 37: **Scheme of the ^{13}C-NOESY 3D experiment magnetization pathway and the corresponding FeoA ^{13}C-NOESY.**

In the **^{1}H-^{13}C-HSQC (**Figure 38**)** spectra the magnetisation is transferred from ^{1}H to ^{13}C and then back again for detection and all ^{1}H-^{13}C moieties, regardless of chemical type, are observed. This is the carbon equivalent of the **^{1}H-^{15}N-HSQC**.

All of the C-H correlations were useful to assign the aromatic peaks and as a basis for picking a 3D **^{13}C-NOESY** spectrum.

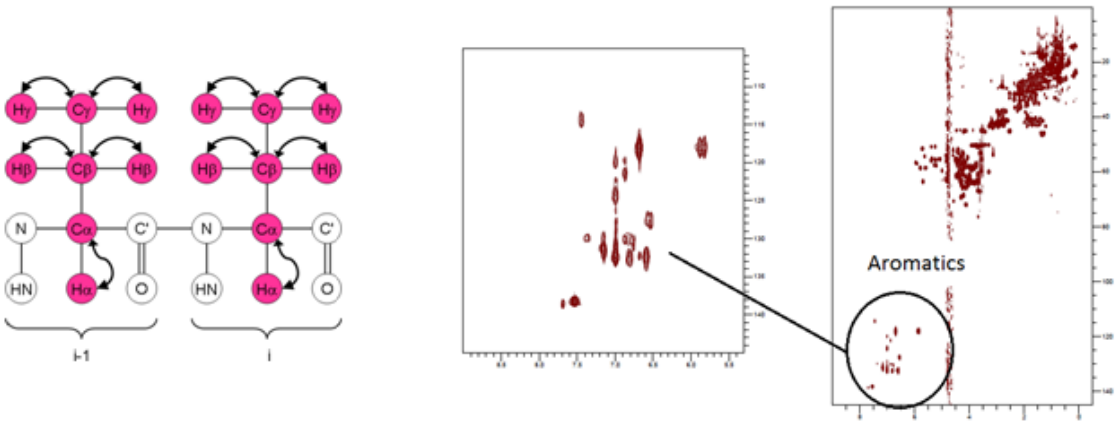

Figure 38: Scheme of the ^{1}H-^{13}C-HSQC 3D experiment magnetization pathway and the corresponding FeoA ^{1}H-^{13}C-HSQC.

A report of the completeness of those assignments was determined by CARA software (Table 1).

Some residues were not assigned because they are localized in the C and N-terminal. The residues present at this regions tend to be absent from the **^{1}H-^{15}N-HSQC** making it difficult to be identified. Others were not assigned because they are localized in areas of the protein that have no structure.

Shifts	Found	Missing	Complete (%)
^{1}H	418	71	85.481
^{1}H aromatics	18	10	64.286
^{1}H aliphatics	400	61	86.768
^{13}C	248	27	90.182
^{13}C aromatics	12	13	48.000
^{13}C aliphatics	236	14	94.400
^{15}N	67	31	68.367
Total	733	129	85.035
Total (without pseudo atoms shifts)	657	129	83.588
Backbone	273	21	92.857
Side chain	460	108	80.986

Table 2: Report of the completeness of the assignments of the FeoA protein.

3.3.3.-FeoA structure

The NOESY spectra provide short-range as well as long-range distance information between pairs of hydrogen atoms separated by less than 5 Å.

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** spectrum. In order to obtain a high resolution structure it was necessary to complete the assignment for a sufficient number of atoms in the sequence.

The analysis of the **^{1}H-^{1}H NOESY** spectrum was facilitated by the combined use of the information on the 3D **^{15}N-NOESY (**Figure 39**)** and **^{13}C-NOESY** spectra.

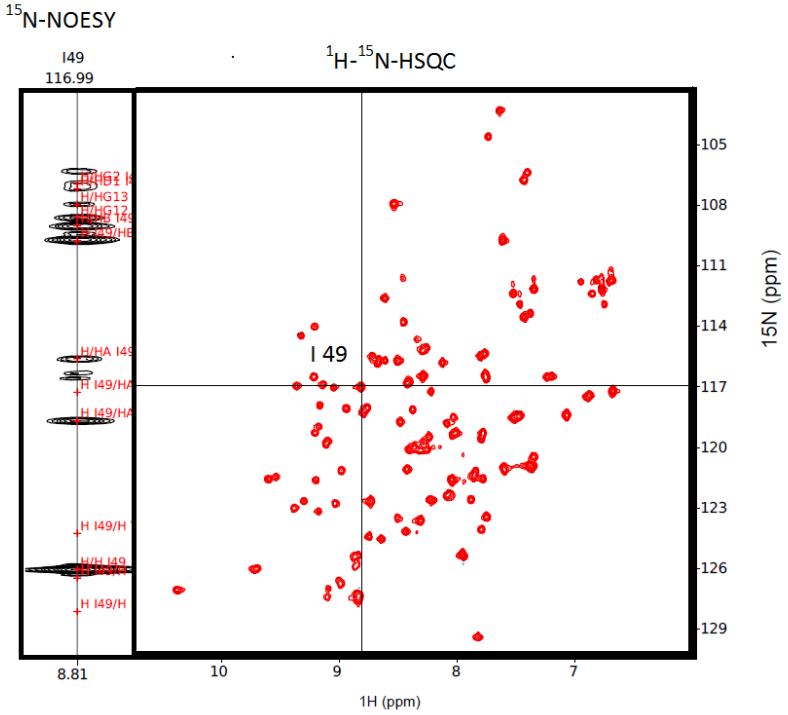

Figure 39: The FeoA ^{15}N-NOESY. Isoleucine 49 is here shown as an example.

After the assignment of the NOESY spectra the NOE intensities were converted into proton-proton distances through a calibration process where NOEs of known distances were used to calculate the calibration constant in the equation 1.1.

Equation 1.1: $NOE = \frac{K}{r^6}$

The NOE assignment and the structure determination was performed using an automated protocol of the programs UNIO and CYANA. The input for these calculations were the NOE spectra, and the assignment. We also used the program DANGLE through the interface of CCPN to calculate dihedral angle constraints based on the chemical shifts of the Hα,Cα,Cβ,CO atoms. This is possible due to the strong relationship between those chemical shifts and the dihedral angles of the peptide plane.

The automated procedure employed by UNIO is a multistep, iterative, process. It starts by the identification of all the NOESY peaks, a process called peak picking. Then the positions of these peaks were then compared with the assigned resonances list. At the first stage the peaks are assigned to the atoms with the closest chemical shift values to those of the NOE peaks. It is also possible to give more than one assignments for each peak. This process generates a list of distance restraints. These restraints are then used as constraints in a series of 100 simulated annealing calculations. The 20 structures with the least number of NOE violations are kept for evaluation.

However, at this stage the assignment contains many incorrect restraints due to the large overlap of ^{1}H chemical shifts. As a result the quality of the structure at this stage is poor. But it is good enough to be used as model for the more correct assignment of the NOEs in the second iteration of the calculation. In this step the NOE assignment is repeated but this time, instead of using only chemical shift similarity as a criterion, the previously calculated structure is used to assist in the NOE identification. This in turn leads to more correct assignment of the NOEs and to a better structure.

This process is repeated 7 times before the final structure is generated.

At this stage we obtained a bundle of structures with good RMSD (Figure 40). To improve the quality of the structure, we repeated the calculation this time using the dihedral angle restraints calculation from the chemical shifts as well as manually picked distances that correspond to hydrogen bonds. These distances were picked very conservatively, only selected for structrural elements where the NOEs had driven the structure to a specific secondary structure.

In Figure 41 we can see some of the restraints used for the FeoA structure calculation and their correlation to secondary structure.

Statistics of protein NMR structure determination	
Number of residues	75 (1-75)
Molecular weight [Da]	8373.75
Number of models	20
Target function [A^2]	8.03 +/- 0.13 (7.75..8.32)
Setup-given RMSD range	1-75
- Backbone RMSD [A]	0.51 +/- 0.16 (0.32..0.86)
- Heavy atom RMSD [A]	1.00 +/- 0.13 (0.85..1.29)
Optimal RMSD range	1-75
- Backbone RMSD [A]	0.51 +/- 0.16 (0.32..0.86)
- Heavy atom RMSD [A]	1.00 +/- 0.13 (0.85..1.29)
NOE restraints [#]	1020
- intraresidual (|i-j|=0)	256 (25.10%)
- sequential (|i-j|=1)	312 (30.59%)
- medium-range (1<|i-j|<5)	172 (16.86%)
- long-range (|i-j| >4)	280 (27.45%)
NOE restraints per residue	13.60
RMS NOE restraint violation [A]	0.0137
Dihedral restraints [#]	405
RMS dihedral restraint violation [^{0}]	3.4920
Ramachandran statistics	
- most favoured [%]	74.61
- additionally allowed [%]	21.95
- generously allowed [%]	3.36
- disallowed [%]	0.08

Figure 40: Final statistics of FeoA structure determination by NMR.

Figure 41: Experimental restraints for FeoA protein, including sequential, short- and medium-range NOEs and Hα, Cα, CO and Cβ secondary shits along with the secondary structure deduced from the data. The amino acid sequence and numbering are shown at the top. Sequential N-N and a-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 and a-N NOEs is indicated by solid lines. The chemical shift indices shown for 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.

In Figure 42 we can see the final structure of FeoA after all the steps involved in the structure determination were complete.

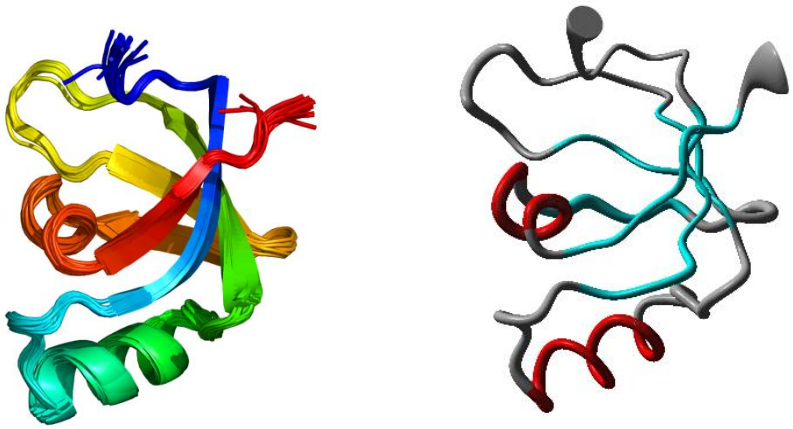

Figure 42: Solution structure of FeoA. a) Backbone ribbon display of the average structure of FeoA. b) The FeoA tertiary structure in sausage model.

3.3.4-NMR differences between dimer and monomer forms

As noted earlier the in the sample that we used, FeoA existed in two forms, a monomer and a dimer. During the assignment, we were able to identify the resonances of the monomeric form. The remaining peaks that corresponded to the dimer needed to be identified and correlated with those of the monomer.

Some chemical shifts did overlay presenting no changes but a significant number of resonances did change. The correlation of the monomeric and the dimeric peaks was possible with the use of the **HNCO** and the ^{15}**N-NOESY** spectra. While the CO resonances of each aminoacid were not identical in the monomeric and the dimeric forms, the pattern was very similar.

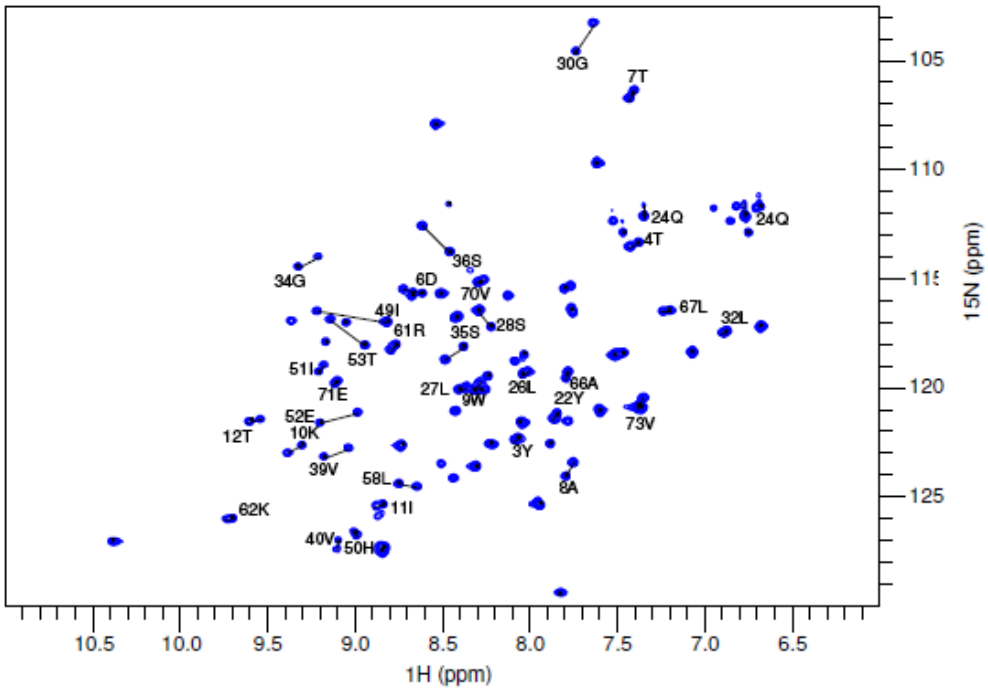

Figure 43: FeoA ^{1}H-^{15}N-HSQC showing the dimer monomer pairs. The identification of the peak is always closer to the monomer peak.

With the knowledge of the chemical shifts of each species, we were able to calculate the combined differences between the monomer and dimer chemical shifts according to equation 1.2. This allowed us to determine the combined chemical shit difference for ^{1}H and ^{15}N dimensions.

$$\text{Combined shift difference} = \sqrt{(\delta_{(1H)m} - \delta_{(1H)d})^2 + 0.156(\delta_{(15N)m} - \delta_{(15N)d})^2}$$

Equation 1.2- Combined chemical shift difference for ^{1}H and ^{15}N dimensions.

The weighing factor of 0,156 is used in the equation to normalize the chemical shift scales of ^{1}H and ^{15}N. The calculated chemical shift differences (Figure 44) were plotted

and the horizontal black line shows the residues which shifted the most (above 0.2) and are dues considered more relevant and significant results.

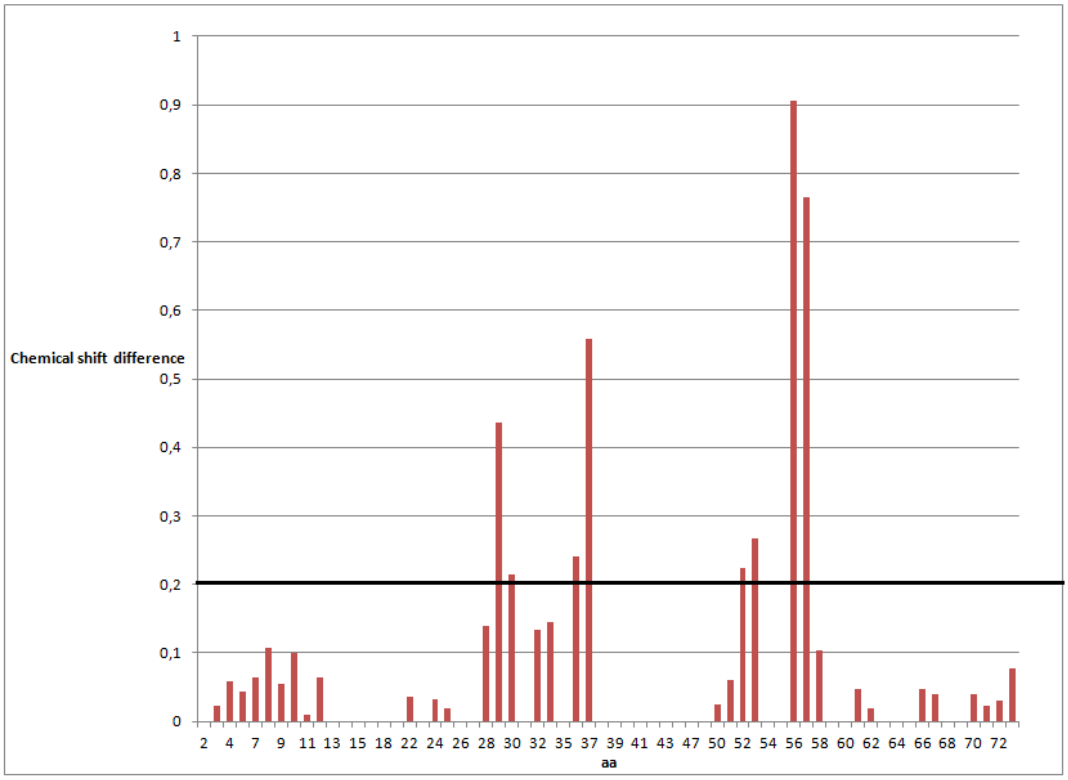

Figure 44: ^{1}H and ^{15}N combined shift difference between the ^{15}N-HSQC spectrum from monomer and dimer forms of FeoA protein. The horizontal black line at 0.2 combined shift difference aims to identify the residues that shifted the most.

In Figure 45 it is shown the number of restrains per residue.

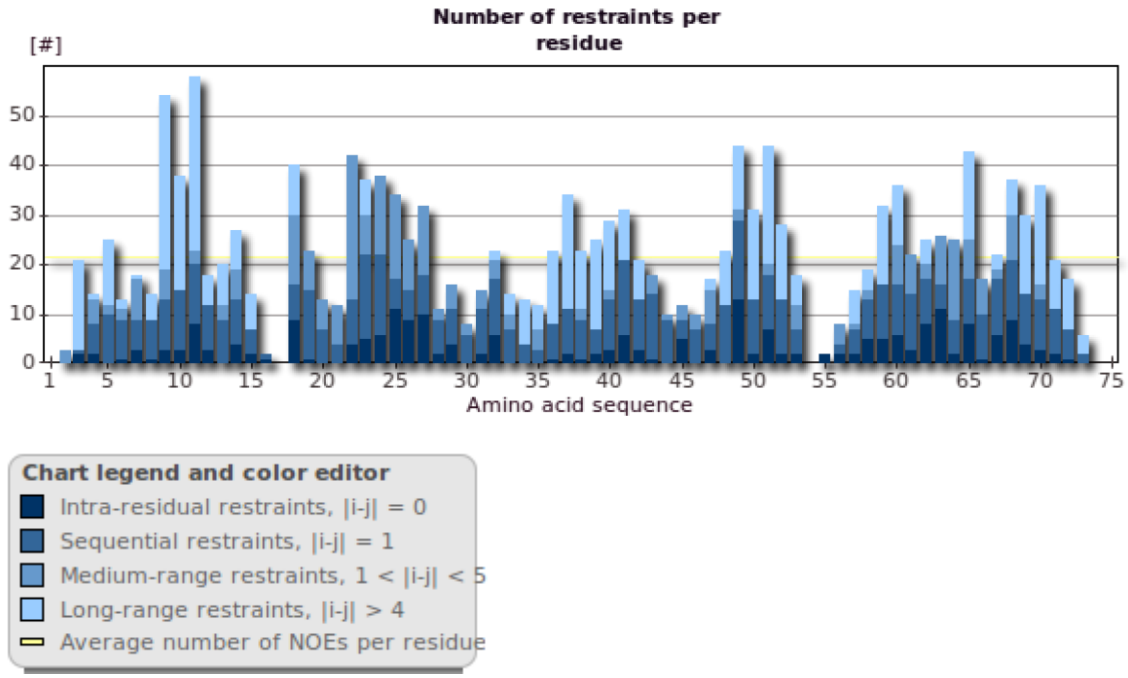

Figure 45: The number of restraints per residue in the FeoA sequence. The chart legend and color is shown above.

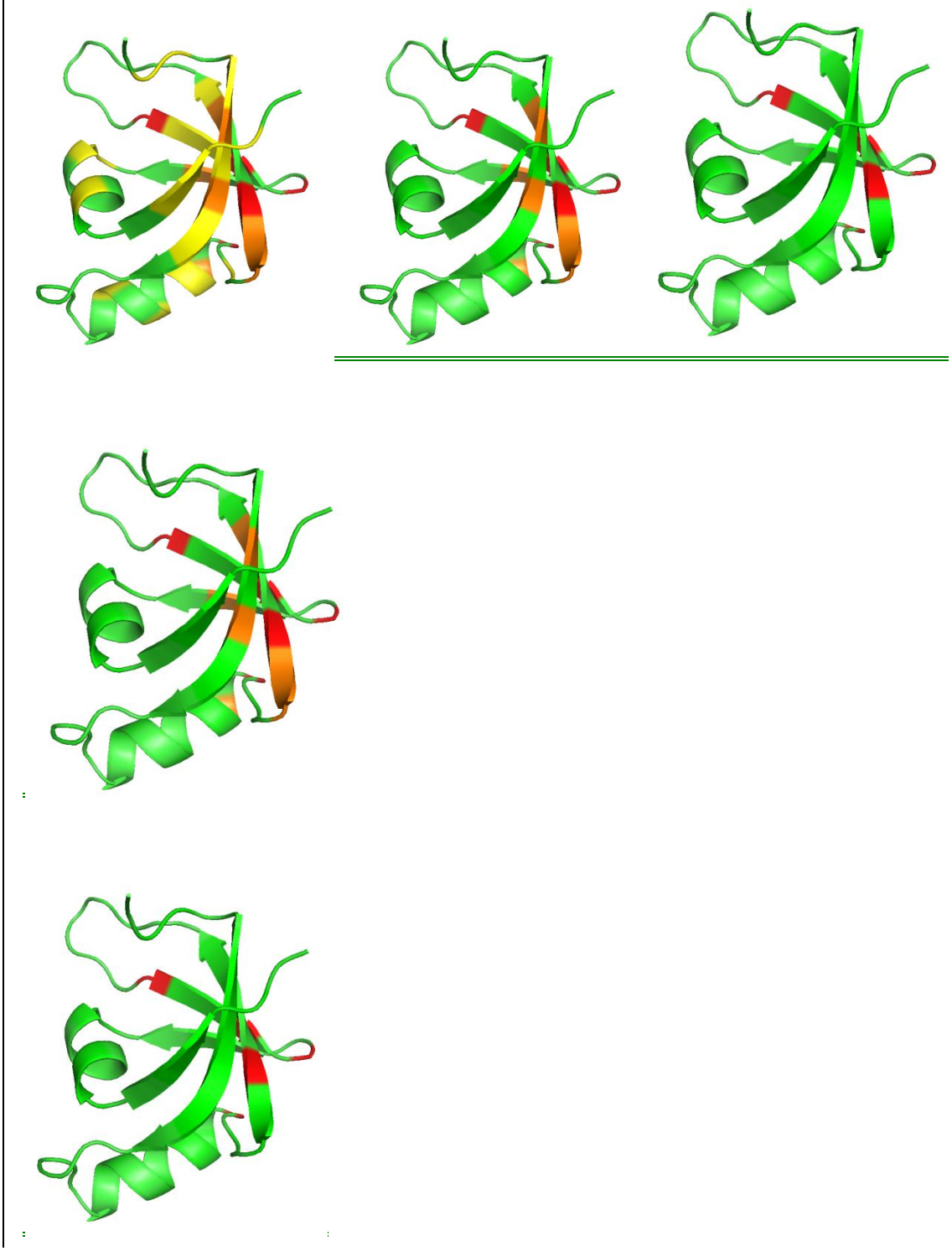

Figure 46: Structure of FeoA with the aminoacids affected by the dimer interaction shown in different colors. The yellow color represents those aminoacids that suffer chemical shift perturbation lower than 0.1 ppm. The color orange shows an intermediate level of perturbation of equal or higher than 0.1 ppm and finally the red color shows the strongest perturbation of equal or higher than 0.2 ppm.

4. Conclusions

4.1 The presence of the FeoA dimer in solution

During this project it was verified that the FeoA protein from *E.coli* can exist in solution both as a monomer and as a dimer. The existence of a 17 kDa FeoA dimer was proved by size exclusion chromatography and NMR spectroscopy.

In the literature the FeoA protein was never reported as a dimer in solution, although in the only paper that describes the FeoA structure by X-Ray Crystallography the purification scheme also included a size exclusion as a second purification step.**(15)**

In this same article the *Sm*FeoA protein exists mainly as a monomer when running gel-filtration column chromatography and analytical ultracentrifugation. The fact that they used the word mainly gives the opportunity to speculate that there were also other forms of the FeoA protein that were residual in concentration.

We can speculate that the FeoA from *Stenotrophomonas maltophilia* behaves differently in solution and doesn't form a dimer as easily as FeoA from *E.coli.* The buffer conditions used were also different than the ones used in this project and that can also account for the differences in the FeoA behavior.

The human Lck-SH3 protein that has the comparable SH3 structure as FeoA forms homodimers in solution induced by Zinc-ions. **(26)**

The SmFeoA has an SH3 domain-like fold that is cross-linked by zinc ions forming a dimer in the crystal structure but this same protein was never reported as a dimer at high zinc-ion concentrations. In the case of our project the FeoA from *E.coli* was titrated with increasing zinc-ion concentration by NMR and after 0.5 concentration equivalent the protein started to precipitate (data not shown). Prior to precipitation no shifts could be observed.

The eukaryotic SH3 protein Eps8 forms an intertwined dimer that exists in solution in a dynamic equilibrium. **(24)**

The dimer and monomer species can be separated on a fast protein liquid chromatography system. However, when peaks corresponding to the monomers and dimmers were collected and individually subjected to gel-filtration chromatography, they gave both dimeric and monomeric peaks. In solution both monomer and dimer are in dynamic equilibrium, with the monomer being the major species.

With the FeoA protein from *E.coli* is also not possible to isolate the monomer and dimer since they exist in a dynamic equilibrium. When we isolate the monomer form

and concentrate it for the NMR analysis the final sample presents both the dimer and monomer form, with the identity of prevalent species being subject to concentration.

In the case of FeoA from *E.coli* it seems that the overall protein concentration plays a role in the proportions of the dimer and monomer, where an increase in the overall concentration leads to a major formation of the dimer. In the case of our project the monomer form was never the major form present in solution as was reported in other publications regarding SH3 domain proteins.

In the literature, the Eps8 monomer form was isolated using a pH change. At pH 4 the major form is the monomer. The change in the pH value disrupts certain intermolecular salt bridges that are necessary to stabilize the dimeric form. The dimeric form is stabilized by hydrophobic interactions and salt bridges, the lowering of the pH protonates certain residues involved in those salt bridges disrupting them and exposing the hydrophobic surfaces forming the monomer. In the dimer the core interface between the two monomers is hydrophobic and it is surrounded by polar interactions and salt bridges.

The pH lowering of the FeoA buffer was never tested but it is something that can be done in the future to understand the nature of the dimer interaction. The pH 4 is a acidic environment that is compatible with NMR analysis.

The physiological relevance of a SH3 dimer has been proven in a lot of proteins both from eukaryotes and prokaryotes.

- The C-terminal domain of the KorB protein is a SH3 domain-like fold protein. This SH3 domain forms a dimer in solution. The dimer interface is also formed by hydrophobic interactions and salt bridges.

In this case the dimerization was proven to be physiologically relevant. The dimerization is important for specific high affinity operator binding, helping in the repressive role exerted by the N-terminal KorB. The SH3-SH3 interactions serves a biological role. Another example of a relevant SH3 dimerization is seen in the N-terminal domain of HIV-1 integrase **(28)** which activity depends on the dimeric state.

Although it is not known the physiological relevance of the FeoA homodimer from *E.coli.* it is of a significant importance that for the first time it was verified the dimer present in solution, an interaction that could be very important for the function or modulation of FeoA.

4.2-FeoA from *E.coli* has an SH3 domain-like fold

The FeoA structure from *E.coli* determined in this project has an SH3 domain-like fold similar to the structure determined by X-Ray Crystallography concerning the FeoA protein from *Stenotrophomonas maltophilia* **(15).**

The eukaryotic proteins with an SH3 domain are involved in signal transduction, in the case of other bacterial proteins that have an SH3 domain-like fold the functions are unknown or differ from that of the eukaryotic SH3 domain.

The determinants of the SH3 domain-like fold consist in some characteristics that have been described to differ from the eukaryotic SH3. The FeoA protein and others are referred as possessing an SH3 domain-like fold because their structure closely resembles the SH3 domain but cannot be traced to classical SH3 domains in terms of function or sequence similarity **(16).**

The FeoA protein from *Stenotrophomonas maltophilia* has only 9% sequence similarity to eukaryotic SH3 domains.

The apparent conservation of hydrophobic residues in key positions observed between these two groups of proteins constitutes the only tangible structural feature shared at the sequence level.

Conservation of a few key hydrophobic positions seems to be the only requirement for maintaining the fold without significant loss of modularity, making it perfect for adapting to a wide range of sequences.

Even between prokaryotic proteins with an SH3 domain-like fold the differences reveal further structural variations inside this kingdom.

4.3-FeoA Dimer

The FeoA from *E.coli* forms a dimer in solution. This dimer was analyzed by NMR where it was possible to determine the residues affected by this interaction. The dimer structure was not determined due to time constraints but a further analysis of the NOESY spectra to detect NOEs between the two monomers will in the future produce a dimer structure.

As was mentioned before the FeoA function was recently clarified. Bacteria two-hybrid assay determined that the FeoA protein directly and specifically binds to the

FeoB transporter in vivo. The FeoA-FeoB interaction appeared necessary for FeoB-mediated Fe (II) uptake **(19).**

It was also verified in this article that Leucine 26 is crucial to the interaction between FeoA and FeoB, where a single L26Q aminoacid substitution was sufficient to abolish the FeoA-FeoB interaction.

The leucine 26 is one of the residues affected by the FeoA dimerization, giving space to speculate that the dimerization could be a form of modulating the FeoA activity. The human protein Lck-SH3 forms a homodimer induced by zinc ions and this dimerization was proved to compete with the binding of proline-rich domains **(26).** We can speculate that the FeoA dimerization also competes with the interaction with FeoB presenting a possible way of modulating the FeoA activity.

5. Future Perspectives

At the end of this project that lasted over a year we were able to determine the structure of FeoA protein from *E.coli.* The function of this protein was not determined but this task has very good perspectives to be further developed at the biomolecular NMR group at ITQB.

This FeoA project was inserted into a bigger project that aims at understanding the whole Feo system. With this in mind the FeoB protein has been recently cloned and purified and it is ready to be used in studies of interaction with FeoA by NMR spectroscopy.

The function of FeoA is now clearer due to an article that came out during this project in June 2012 that states that FeoA protein is crucial to the import of Fe (II) by FeoB. This interaction was verified but not mapped something that can be done by NMR spectroscopy that not only proves interaction between molecules but can also map them.

Although we did not characterize the function of FeoA protein we verified that this protein exists in solution in two different forms the dimer and monomer forms. This was something reported for the first time since the only article publish about FeoA structure clearly stated that they detected FeoA mainly as a monomer in solution.

The nature of the FeoA association was only briefly understood in this work. For the future it is essential to understand this interaction by determining its dissociation constant and if possible the physiological relevance of this dimer.

Also the interaction studies with FeoB could also determine the physiological relevance of this homodimer.

6. Bibliography

(1) Posey,J., Gherardini, F., 2000 "Lack of a role for iron in the Lyme disease pathogen" Science; 288:1651-1653.

(2) Archibald, F., Duong, M., 1984"Manganese acquisition by *Lactlbacillus plantarum*" Journal Bacteriology; 158:1–8.

(3) Andrews, S., Robinson, A., Rodriguez-Quiñones,F., 2003 " Iron homeostasis" Microbiology Reviews 27 (2003) 215-237.

(4) Madigan, M., Martinko ,J., "Brock-Biology of Microorganisms" 2006 11th edition.

(5) Wandersman, C., Delepelaire,P., 2004 "Bacterial Iron sources: From Siderophores to Hemophores "Annu. Rev. Microbiol. 2004. 58:611–47.

(6) Carton,M., Maddocks, S., Gillingham,P., Craven, J., Andrews,S., 2006"Feo-Transport of ferrous Iron into bacteria" BioMetals 19:143-157.

(7) Grass,G., 2006 "Iron transport in *Escherichia coli*: All has not been said and done" BioMetals 19:159-172

(8) Kammler,M., Schon, C ., Hantke,K., 1993"Characterization of the ferrous iron uptake system of *Escherichia coli*"Journal of Bacteriology Oct.1993 p.6212-6219.

(9) Hantke,K., 2003" Is the bacterial ferrous iron transporter FeoB a living fossil?"Trends Mricrobiology 2003 May; 11(5):192-5.

(10) Koster,S., Wehner,M., Herrmann,C., Kuhlbrandt, W., Yildiz,O., 2009 "Structure and function of the FeoB G-domain from *Methanococcus jannaschii* " J Mol Biol. 2009 Sep 18; 392(2):405-19

(11) Hattori, M., Jin, Y., Nishimasu, H., Tanaka, Y., Mochizuki, M., Uchiumi, T., Ishitani, R., Ito, K. and Nureki, O. (2009)" Structural basis of novel interactions between the small-GTPase and GDI-like domains in prokaryotic FeoB iron transporter". Structure 17(10): 1345-55.

(12) Hung, K. W., Chang, Y. W., Eng, E. T., Chen, J. H., Chen, Y. C., Sun, Y. J., Hsiao, C. D., Dong, G., Spasov, K. A., Unger, V. M. and Huang, T. H. (2010). "Structural fold,

conservation and Fe(II) binding of the intracellular domain of prokaryote FeoB". J Struct Biol 170(3): 501-12.

(13) Koster, S., Wehner, M., Herrmann, C., Kuhlbrandt, W. and Yildiz, O. (2009). "Structure and function of the FeoB G-domain from Methanococcus jannaschii". J Mol Biol 392(2): 405-19.

(14) Gajiwala,K., Burley, S., 2000 " Winged helix proteins" Curr Opin Struct Biol. 2000 Feb;10(1):110-6

(15) Su, Y. C., Chin, K. H., Hung, H. C., Shen, G. H., Wang, A. H. and Chou, S. H. (2010). **"**Structure of *Stenotrophomonas maltophilia* FeoA complexed with zinc: a unique prokaryotic SH3-domain protein that possibly acts as a bacterial ferrous iron-transport activating factor". Acta Crystallogr Sect F Struct Biol Cryst Commun 66(Pt 6): 636-42.

(16) D´Aquino, J., Ringe, D., 2003 "Determinants of the Src Homology Domain 3-Like fold" Journal of Bacteriology p. 4081-4086.

(17) Ogura, K., Nobuhisa, I., Yuzawa, S., Takeya, R., Torikai, S., Saikawa, K., Sumimoto, H., Inagaki, F,. 2003 "NMR solution structure of the tandem Src homology 3 domains of p47phox complexed with a p22phox-derived proline-rich peptide" J Biol Chem. 2006 Feb 10;281(6):3660-8.

(18) Qiu, X., Pohl, E., Holmes, RK., Hol, WG.,1996 "High-resolution structure of the diphtheria toxin repressor complexed with cobalt and manganese reveals an SH3-like third domain and suggests a possible role of phosphate as co-corepressor" Biochemistry. 1996 Sep 24;35(38).

(19) Kim, H., Lee, H., Shin,D., 2012 " The FeoA protein is necessary for the FeoB transporter to import Ferrous iron" Biochem Biophys Res Commun. 2012 Jul 13;423(4):733-8.

(20) Teng, Q., " Structural Biology" 2005 Springer Science p.211-212.

(21) pET System Manual, 11th edition from Novagen.

(22) Stable Isotopes for Structural Biomolecular NMR, 2011 from Cambridge Isotope Laboratories.

(23) Ion Exchange Chromatography & Chromatofocusing- Principles and methods, 2004 from Amersham Biosciences

(24) Kishan,K;Newcomer,M; Rhodes,T; Guilliot,S; 2001"The effect of pH and salt bridges on structural assembly: Molecular strucutres of the monomer and interwined dimer of Eps8 SH3 domain" Protein Science 2001,10:1046-1055.

(25) Delbruk,H; Ziegelin,G; Lanka, E; Heinemann U; 2002 "An Src Homology 3-like domain is responsible for the dimerization of the Repressor protein KorB encoded by the promiscuous IncP plasmid RP4" The Journal of Biological Chemistry 2002, Vol277,No.6, 4191-4198.

(26) Romir,J; Lilie,H; Egerer-Sieber, C; Bauer,F; Sticht,H; Meller,A.(2007) J. Mol. Biol. 365, 1417-1428.

(27) Lodi,P; Ernst, J; Kuszewski,J; Hickman, A; Engelman, A; Craigie, R; Clore,G; Gronenborn, A.(1995)"The DNA-binding domain of HIV-1 Integrase has an SH3-like fold". Biochemistry 34, 9826-9833.

(28) Hung,K; Tsai,J; Juan,T; Hsu, Y; Hsiao, C; Huang, T.(2012) "Crystal structure of *Kp*NFeoB/*Kp*FeoC complex and the roles of FeoC in the regulation of Fe^{2+} transport by the bacterial Feo system" J. Bacteriol. 2012

(30) Gel Filtration-Principles and Methods, 2004 from Amersham Biosciences