António João Moreira Nabais Neca

Biochemistry Graduated

The *Desulfovibrio vulgaris* Hildenborough ORP: isolation, cluster reconstitution and NMR resonance assignment

Dissertation submitted in fulfillment of the requirements for the degree of Master of Science in Biochemistry

Supervisor: Sofia Rocha Pauleta, Principal Investigator, FCT/UNL

Jury:

Chair: Doctor José Ricardo Ramos Franco Tavares
Arguer: Doctor Carlos Alberto Gomes Salgueiro
Member: Doctor Sofia Rocha Pauleta

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

September 2015

António João Moreira Nabais Neca

Biochemistry Graduated

The *Desulfovibrio vulgaris* Hildenborough ORP: isolation, cluster reconstitution and NMR resonance assignment

Dissertation submitted in fulfillment of the requirements for the degree of Master of Science in Biochemistry

Supervisor: Sofia Rocha Pauleta, Principal Investigator, FCT/UNL

Jury:

Chair: Doctor José Ricardo Ramos Franco Tavares
Arguer: Doctor Carlos Alberto Gomes Salgueiro
Member: Doctor Sofia Rocha Pauleta

September 2015

Acknowledgments

In the first place, I would like to express my sincere gratitude to my supervisor Dra Sofia Rocha Pauleta for allowing me to be a part of this project and for the all the friendship and all the support, knowledge, patience and guidance given during my stay, and to professor Isabel Moura for all the friendship and for welcoming me into her laboratories. I would also like to thank the support from Fundação para a Ciência e Tecnologia for the research project (FCT-ANR/BBBMET/0023/2012).

I want to thank all the members of the work group, from the fourth and sixth floor, for the warm welcome and the great atmosphere created in work and non-work related activities. Especially to those who were more close to me, Cintia Carreira, Rute Nunes, Cláudia Nóbrega, Olga Mestre, Susana Ramos and Rui Almeida, for all the friendship, support, patience and encouragement and for all the fun we have had in the last year.

Last but not least I want to express my warm thanks to my family: to my parents, to my brother and to my sister for all the love, patience, encouragement, and help throughout my master degree and life in general. To all my friends who have walked this road with me I thank you for all the help, motivation and friendship given along my experience in the glorious Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa. And to my dog for the friendship and company in trough the all-nighters.

Abstract

DVU2108 is a protein belonging to the Orange Protein (ORP) family that in *Desulfovibrio vulgaris* Hildenborough forms a protein complex named the Orange Protein Complex. ORP is a conserved protein in anaerobic microorganisms and in *Desulfovibrio gigas* the homologous protein was isolated with a novel Mo-Cu cluster non-covalently attached to the polypeptide chain. The protein sample isolated by homologous expression of StrepDVU2108 in *Desulfovibrio vulgaris* Hildenborough displayed an absorption spectra similar to an iron-sulfur containing protein.

Nevertheless, the Apo form of DVU2108 (obtained by heterologous expression in *E.coli*) was able to incorporate a MoCu or a WCu cluster, when in the presence of tetrathiomolybdate or tetrathiotungstate and copper chloride, with a metal to metal ratio of 1Cu:2.3Mo ± 1.2 and 1Cu:1.49W ± 0.03, which is close to the preferable metal ratio for this protein found when titrating Mo or W to a equimolar solution of Apo-DVU2108 and copper (1Cu:2Mo/2W). The same metal stoichiometry was found in *Desulfovibrio gigas* ORP. The molar extinction coefficients for MoCu-DVU2108 are 10500 ± 280 $M^{-1}cm^{-1}$ at 334 nm and 5570 ± 150 $M^{-1}cm^{-1}$ at 483 nm, per molybdenum atom. For WCu-DVU2108 the extinction coefficients are 14560 ± 80 $M^{-1}cm^{-1}$ at 295 nm and 6990 ± 40 $M^{-1}cm^{-1}$ at 407 nm, per tungsten atom. The reconstituted proteins, MoCu-DVU2018 and WCu-DVU2108 showed higher stability with higher melting temperatures than the Apo form, 72.1 ± 0.6 ºC, 72.1 ± 0.7 ºC and 65.6 ± 0.2 ºC, respectively.

In a second part of this work, the resonance assignment of Apo-DVU2108 was accomplished at 94 %. The ^{1}H-^{15}N HSQC of this protein presented 118 resonances that were assigned to 101 of the total number of amides (113), including 9 side chains.

Keywords: *Desulfovibrio,* Anaerobic bacteria, Orange protein, NMR assignment, Novel Mo-Cu cluster

Resumo

DVU2108 é uma proteína que faz parte da família das Proteínas Laranja (ORP) que em *Desulfovibrio vulgaris* Hildenborough forma um complexo proteico denominado Complexo da Proteína Laranja. ORP é uma proteína conservada em microrganismos anaeróbicos em que a proteína homóloga em *Desulfovibrio gigas* foi isolada contendo um *cluster* de Mo e Cu único na natureza, ligado não covalentemente à cadeia polipeptídica. O espectro de absorção da amostra de proteína isolada por expressão homóloga de StrepDVU2108 em *Desulfovibrio vulgaris* Hildenborough exibia bandas características de uma proteína contento um *cluster* de ferro e enxofre.

No entanto, a forma Apo de DVU2108 (obtido por expressão heteróloga em *E. coli*) é capaz de incorporar *clusters* de MoCu ou de WCu, quando na presença de tetratiomolibdato ou tetratiotungstato e cloreto de cobre, respetivamente, com um rácio de metais de 1Cu:2.3Mo ± 1.2 e 1Cu:1.49W ± 0.03, que é próximo do rácio preferido para esta proteína determinado pela titulação de Mo ou W a uma solução equimolar de Apo-DVU2108 e cobre (1Cu:2Mo/W). O mesmo rácio de metais é encontrado na ORP de *Desulfovibrio gigas*. Os coeficientes de extinção molar obtidos para a proteína MoCu-DVU2108 foram de 10500 ± 280 $M^{-1}cm^{-1}$ a 334 nm e 5570 ± 150 $M^{-1}cm^{-1}$ a 483 nm, por átomo de molibdénio. Para a proteína WCu-DVU2108 os coeficiente de extinção molar obtidos foram 14560 ± 80 $M^{-1}cm^{-1}$ a 295 nm e 6990 ± 40 $M^{-1}cm^{-1}$ a 407 nm, por átomo de tungsténio. As proteínas reconstituídas, MoCu-DVU2018 e WCu-DVU2108, demonstraram ser mais estáveis tendo temperaturas de desnaturação mais elevadas do que a forma Apo, 72.1 ± 0.6 ºC, 72.1 ± 0.7 ºC e 65.6 ± 0.2 ºC, respetivamente.

Na segunda parte deste trabalho, a atribuição de ressonâncias da Apo-DVU2108 foi de 94%. O ^{1}H-^{15}N HSQC desta proteína apresentava 118 ressonâncias que foram atribuídas a 101 do total número de amidas (113), incluindo 9 cadeias laterais.

Termos chave: *Desulfovibrio,* Bactérias anaeróbicas, Proteína laranja, Atribuição por RMN

Table of Contents

List of Figures

List of tables

List of abbreviations

∅	Diameter
ε	Molar Extinction Coefficient
λ	Wavelength
Abs	Absorbance
AMP	Adenosine Monophosphate
APS	Adenosine-5'-Phosphosulfate
ATP	Adenosine triphosphate
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
CCPN	Collaborative Computing Project for NMR
COG	Cluster of groups of orthologous proteins
CSI	Chemical shift index
DMF	*N,N*-dimetilformamida
DSC	Differential Scanning Calorimetry
DTT	Dithiothreitol
EBP	Enhancement Binding Protein
EXAFS	Extended X-ray absorption fine structure
FID	Free Induction Decay
HSQC	Heteronuclear Single Quantum Correlation
ICP	Inductively coupled plasma mass spectrometry
IHF	Integration Host Factor
INEPT	Insensitive Nuclei Enhanced by Polarization Transfer
NADH	Nicotinamide Adenine Dinucleotide
NCBI	National Center for Biotechnology Information
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser effect spectroscopy
$OD_{600\ nm}$	Optic Density at 600 nm
ORP	Orange Protein
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
PPi	Pyrophosphate
RCI	Random Coil Index
SDS-PAGE	Polyacrilamide Gel Electrophoresis with Sodium Dodecyl Sulfate
SRB	Sulfate Reducing Bacteria
TALOS+	Torsion Angle Likeliness Obtaine from Shift and Sequence Similari
TCEP	Tris(2-carboxyethyl) phosphine
TCI Probe	Triple Resonance Probe
Tris	Tris (hydroxymethyl) aminomethane
TTMo	Tetrathiomolybdate
TTW	Tetrathiotungstate
UV	Ultraviolet

Microorganisms

D.	*Desulfovibrio*
DvH	*D. vulgaris Hildenborough*
E. coli	*Escherichia coli*
M.	*Methanobacterium*
T.	*Thermotoga*

1. Introduction

1.1 Sulfate-reducing bacteria

Sulfate-reducing bacteria (SRB) are anaerobic prokaryotes that combine the oxidation of organic compounds or molecular hydrogen (H_2) with the reduction of sulfate (SO_4^{2-}) to produce energy necessary for cell synthesis and growth. It is known that SRB use a wide range of organic compounds of low-molecular mass (such as, ethanol, lactate, pyruvate) which can be incompletely oxidized, producing acetate and CO_2, or completely oxidized, producing solely CO_2.[1–3] When a strain has the ability to do the last it is usually named a complete oxidizing sulfate reducer.[3]

Despite their name, SRB can use many other electron acceptors, such as thiosulfate, sulfite and sulfur and even nitrate and nitrite. This versatility coupled with their capability to grow in different environmental conditions allows them to be ubiquitous in nature and found in engineered environments where sulfate is present.[2]

Sulfate reduction is more complicated than molecular oxygen due to the larger number of enzymes required and because sulfur states lower than +6 are quite reactive and can go through interconversion or autoxidations.[1] Sulfate is transported through the cytoplasm membrane via an ion gradient, as its reduction occurs in the cytoplasm, whether simultaneously with protons (fresh-water species) or with sodium ions (salt dependent species). [1,3]

Free sulfate is not easily reduced, the redox potential of the pair sulfate-sulfite is too negative ($E^{0'}$ = -516 mV) to be reduced by the electron mediators in the cell, ferredoxin ($E^{0'}$=-398 mV) and NADH ($E^{0'}$=-314 mV). Therefore, the first step in sulfate reduction is its activation to adenosine-5'-phosphosulfate (APS) by ATP sulfurylase (Eq. 1). Pyrophosphate is hydrolyzed to 2-phosphate by a pyrophosphatase which favors the formation of APS (Eq. 2).

$$(1) \qquad SO_4^{2-} + ATP + 2H^+ \rightarrow APS + PPi \qquad \Delta G^{0'} = +46 \text{ kJ/mol}$$

$$(2) \qquad PPi + H_2O \rightarrow 2Pi \qquad \Delta G^{0'} = -22 \text{ kJ/mol}$$

APS is then converted into sulfite or bisulfite and AMP by APS reductase (Eq. 3). The APS-sulfite redox couple plus AMP have an $E^{0'}$ of -60 mV which is not an obstacle to the cell.

$$(3) \qquad APS + 2e^- \rightarrow HSO_3^- + AMP$$

This leads to the final step in sulfate reduction, the conversion of sulfite or bisulfite to sulfide catalyzed by sulfite reductase ($E^{0'}$=-116 mV) (Eq. 4). Both substrates have a free electron pair at the sulfur and so they are more reactive than the initial substrate sulfate.[1–3]

$$(4) \qquad HSO_3^- + 6e^- + 8H^+ \rightarrow H_2S + 6H_2O$$

APS reductase and sulfite reductase are not membrane-bound or membrane associated enzymes like the cytochrome oxidases in aerobic respiration, being this one more aspect in which the two processes differ.[3]

As one can see, this SRB characteristic gives them an important role in the sulfur cycle. Sulfur is one of the most abundant elements on the planet, which can be found in the form of pyrite (FeS_2) and gypsum ($CaSO_4$) in rocks and sediments or as sulfate in seawater. The last is one of the most abundant and thermodynamically stable sulfur species in the biosphere, making sulfate reduction the core of the sulfur cycle.

Sulfate reduction is also involved in organic carbon in marine sediments, thus SRB also have a role in the carbon cycle.[1,2] Furthermore, sulfate is present in waste water of some industrial processes that use sulfuric acid. This leads to the occurrence of sulfate reduction by SRB, which is undesirable because the end product sulfide is a toxic, odorous and corrosive chemical compound. Sulfide has also a great impact in the environment, as it can react with oxygen and iron minerals.[1,2,4]

The sulfate reduction pathway can involve electron transport from a metal surface and in such cases SRB would have a role in microbial influenced corrosion.[4] Nonetheless they also have biotechnological interest as they can be used to remove sulfur dioxide from flue gas and sulfate and metals from waste water. Whereas metal sulfates (cadmium, nickel, cobalt, zinc, copper and iron) are extremely soluble, the corresponding metal sulfides are not. Thus, using this sulfate reduction one can recover metals by precipitation.[2,3]

1.2 *Desulfovibrio vulgaris*

The *Desulfovibrio* (*D.*) genus, a member of the δ-proteobacteria, has been the chosen model for studying many aspects of the metabolism and biochemistry of SRB, especially *Desulfovibrio gigas* and *Desulfovibrio vulgaris* that are by far the most studied strains. In fact, the first crystal structures of SRB enzymes were obtained from *D. gigas* (hydrogenase and aldehyde oxireductase) and the complete genome of *D. vulgaris* Hildenborough has been sequenced, and many genetic tools have been designed to engineer this organism.[1–4]

Desulfovibrio strains feature a vast range of enzymes and proteins that seem to be oxygen-sensitive. Despite this characteristic, a large number of SRB, including several *Desulfovibrio* species which have superoxide dismutase activity, have been shown to survive when exposed to

oxygen. This ability to endure aerobic conditions is rather limited but guarantees the survival of the organism.[3,4]

In *D. vulgaris* the electrons for sulfate reduction are provided from organic acid and alcohol oxidations producing acetic acid as an end product and therefore this organism is an incomplete oxidizing sulfate reducer. Although lactate is the preferred substrate, *D. vulgaris* is known to use other compounds such as pyruvate, ethanol, malate and fumarate but not riboses or hexoses. This strain's genome also contains pathways for the reduction of other terminal acceptors other than sulfate (oxygen, nitrite and metal ions). This alternate electron acceptor capacity may be used to prevent inhibition of sulfate reduction by these compounds.[4]

1.3 The Orange Protein

An interesting ORange colored Protein (ORP) was isolated from *D. gigas*[5] and sequence analysis revealed that its homologous gene is present in other *Desulfovibrio* species, such as *D. vulgaris* Hildenborough and *D. alaskensis* G20, which have 48% and 42% homology, respectively.[6] A special characteristic of ORP is its novel heterometallic cluster composed by copper, molybdenum and sulfur. Copper and molybdenum are often used as catalytic and structural elements in several enzymes and other molecules but are seldom found together.[5] Analysis by extended X-ray absorption fine structure (EXAFS) indicated that this cluster has a 2:1 ratio of molybdenum and copper and each metal is connected to four sulfur atoms, $[S_2MoS_2CuS_2MoS_2]^{3-}$, and is non-covalently bound to the polypeptide chain (Figure 1.1).[5,7]

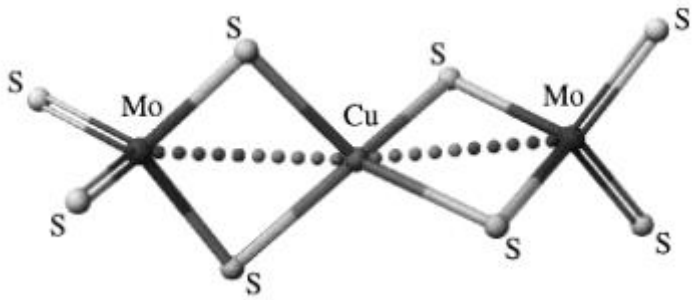

Figure 1.1 – Structure of the $[S_2MoS_2CuS_2MoS_2]^{3-}$ cluster found in the orange protein from D. gigas[7].

The structure of ORP in its Apo-form has been determined for homologues from *Thermotoga (T.) maritima* (TM1290 – pdb cpde 1RDU)[8], *Methanobacterium (M.) thermoautotrophicum* (MTH1175 – pdb code 1EO1)[9] and *D. gigas* (pdb code 2WFB)[6,10], although their role in each organism has not yet been determined. These ORP homologous are part of a family of conserved hypothetical proteins (COG1433). The structures of MTH1175 and TM1290 consists of six and five β-sheets, respectively, flanked by three α-helices, two on one side and the remaining on the opposite side. This α/β structure topology and fold is similar to the Ribonuclease HI family.[8,9]

Figure 1.2 and Table 1.1 show the sequence alignment of *D. vulgaris* Hildenborough ORP homologue with the other known homologues and some of their biochemical properties and percentage of sequence homology.

Figure 1.2 – Sequence alignment of D. vulgaris Hildenborough ORP homologue with the other known homologues

Table 1.1 – Some biochemical properties of ORP proteins obtained with EXPASY ProtParam and percentage of sequence homology with D. vulgaris Hildenborough *ORP obtained with NCBI-BLAST*[11]

ORP homologues	MW (kDa)	PI	AA	Cys	Homology with DVU2108 (%)
***D. vulgaris* Hildenborough**	12.5	5.7	119	2	100
D. gigas	11.8	6.5	117	0	48
***D. alaskensis* G20**	12.4	5.4	119	0	59
M. thermoautotrophicum	13.2	9.5	124	0	35
T. maritima	12.5	4.9	119	0	36

1.4 The Orange Protein Operon in *Desulfovibrio vulgaris* Hildenborough

Approximately a third of *D. vulgaris* Hildenborough genome contains hypothetical and conserved hypothetical proteins of unknown function[4,12] and several of them may have a role in the anaerobic way of life of SRB since they are conserved in anaerobes. The analysis of genomes from aerobic and anaerobic microorganisms and their comparison through clusters of groups of orthologous proteins (COGs)[13] enabled the identification of 33 COGs that are specific to anaerobic way of live, in which five correspond to proteins of unknown function.[14]

The ORP homologue from *D. vulgaris* Hildenborough, encoded by the gene *DVU2108*, belongs to one of those five COGs, namely COG1433 which is similar in sequence to proteins, NifB and NifX, that may have a role in the biosynthesis of the MoFe cofactor in the nitrogen fixation.[15] This protein has 119 residues and along with six other genes completes the *orp* gene cluster which is arranged in two divergent operons: *orp1* (*DVU2107-DVU2108-DVU2109*) and *orp2* (*DVU2103-DVU2104-DVU2105*). Between these two operons there is a gene (*DVU2106*) that encodes a σ^{54}-dependent transcription factor that not only activates the transcription of both operons but also down regulates its own expression. It is also down regulated in the presence of oxygen (Figure 1.3).

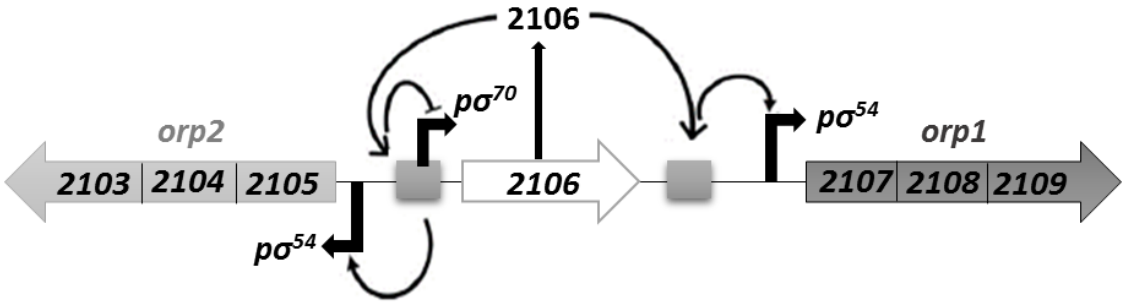

Figure 1.3 – Arrangement and regulatory mechanisms of the orp gene cluster in D. vulgaris Hildenborough. Adapted from [16]

In the promotor regions of *orp1* and *orp2* there are imperfect palindrome sequences where DVU2106 specifically binds, and downstream of those sites are the binding sequences for σ^{54}-RNA polymerase. The interaction of these two proteins is essential for the transcription of both operons.[16] Between their binding sites there is yet another protein binding sequence specific for the integration host factor protein (IHF) (Figure 1.4). IHF are proteins that bend the DNA allowing efficient and functional interaction of σ^{54}-RNA polymerase with an enhancement binding protein (EBP) like DVU2106[17–19].

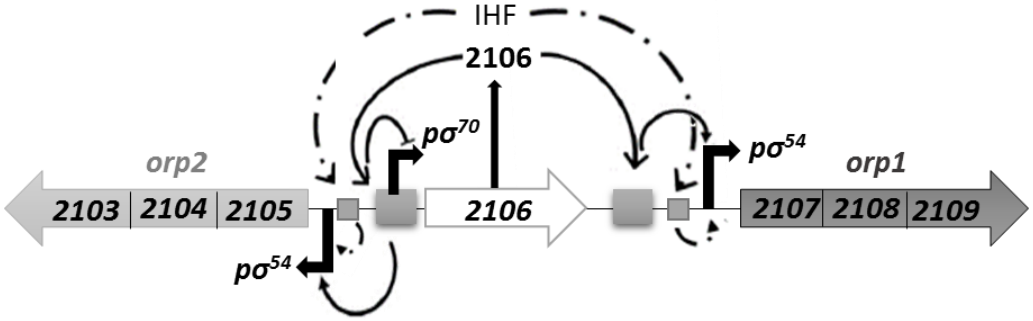

Figure 1.4 – Arrangement and regulatory mechanisms of the orp gene cluster in D. vulgaris Hildenborough with IHF binding sequences location. Adapted from [20]

In the absence of DVU2106, *orp1* and *orp2* expression are greatly affected.[20] Transcription of regulator DVU2106 is unaffected as its expression is σ^{54}-independent. Instead, in its promotor region there is a binding site for σ^{70}-RNA polymerase, which -10 element overlaps the DVU2106 imperfect palindrome binding sequence. This results in a direct competition between σ^{70} and DVU2106, wherein the last displaces the first, and thus down regulating its own expression.[16] Therefore σ^{54}-RNA polymerase, DVU2106 and IHF are involved in *orp* gene cluster transcription regulation. This gene cluster might be involved in cell division since its inactivation affects bacterial cell morphology.[16]

A phylogenetic analysis shows that homologues of *DVU2108* genes tend to be clustered with *DVU2103* and *DVU2104* homologues in most archaeal and bacterial genomes. In *D. vulgaris* Hildenborough their expression increases when the growth conditions change from syntrophic to sulfate reduction, thus it is important to understand its physiological role[21].

In vivo pulldown experiments revealed that DVU2103, DVU2104, DVU2105, DVU2108 and DVU2109 proteins form a multiprotein complex in *D. vulgaris* Hildenborough, and it was proposed that DVU2108 (ORP) interacts directly with DVU2103 and DVU2104.[16] These proteins have 41% of sequence similarity to each other and they both belong to COG1149 which includes MinD superfamily p-loop ATPases with an additional ferredoxin domain.[16] DVU2103 is a Fe-S protein with two 4Fe-4S clusters in its structure (unpublished data of the S.Pauleta research lab) and its N-terminal has homology with COG2894 which comprises septum formation inhibitor-activating ATPases. DVU2107, DVU2105 and the C-terminal domain of DVU2109 belong to the same COG that DVU2108 (COG1433), the N-terminal domain of DVU2109 belongs to COG0489 which comprises ATPases involved in chromosome partitioning. [16]

1.5 Protein Nuclear Magnetic Resonance

Nuclear Magnetic Resonance (NMR) allows us to detect atomic nuclei by taking advantage of their nuclear spin property. Nuclear spin creates a magnetic dipole that can be seen as a bar magnet. Depending on its mass and charge number a nuclei will have different nuclear spin quantum numbers, for example, nuclei with an odd mass number have a half integral spin (I=± 1/2, ± 3/2, etc.) and can be detected by NMR spectroscopy. When in the presence of a magnetic field this spin quantum number defines the number of quantum states (2I + 1). For the ^{1}H nuclei the spin number is 1/2 which generates two energy states and thus the dipole can either align with the field (lowest energy state) or align against it (highest energy state).[22,23] In the absence of an external magnetic field these two states are degenerated, they have the same energy, the applied magnetic field will thus induce a separation to different and quantized energy levels (Figure 1.5). The energy difference depends on the magnetic field strength and has a very small value which leads to a very small nuclei population difference between the two energy levels (aligned or unaligned).

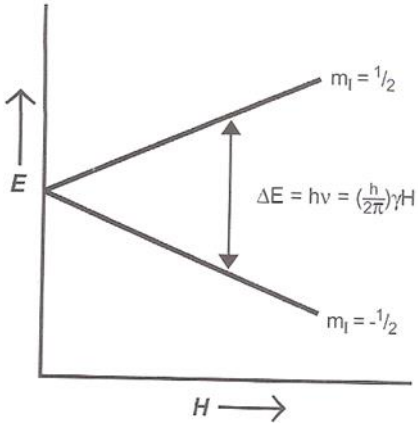

Figure 1.5 – Schematic representation of energy versus magnetic field for a nuclear spin with a nuclear quantum number of 1/2.[22]

If a nuclei is irradiated with an electromagnetic pulse of radio-wave frequency some of them will absorb the energy and align themselves against the field thus generating higher population at the higher energy state. When the pulse stops, the nuclei will return to equilibrium (lowest energy state) irradiating energy, as a radio frequency wave and that is detected by the NMR spectrometer. The free induction decay (FID) of the nuclei returning to equilibrium can be transformed into a frequency spectrum through a Fourier transform.

The number of electrons and nuclei surrounding each individual nucleus is different from one another. Electrons will interact with the applied magnetic field with a magnetic moment that

opposes the field, thus creating a lower magnetic field locally. This is called shielding and it is the reason why the same type of nuclei have a different chemical shift on the NMR spectra depending on their local environment.

Since the frequency at which a nucleus resonates depends on the strength of the applied magnetic field, different spectrometers would have different spectra for the same molecule, which would make it harder to compare spectra. Frequency can then be expressed as a relative change to a given standard, depending on the type of nucleus, enabling a spectrum representation independent of the applied magnetic field. This is called the chemical shift (δ) and is given in parts per million (ppm) according to equation 5. Where v is the frequency of the nucleus and v_{ref} is the frequency of the standard compound.[22–24]

$$\delta = \frac{\nu - \nu_{ref}}{\nu_{ref}} 10^6 \quad (5)$$

Figure 1.6 shows a ^{1}H spectrum of ubiquitin with the regions of specific protons.

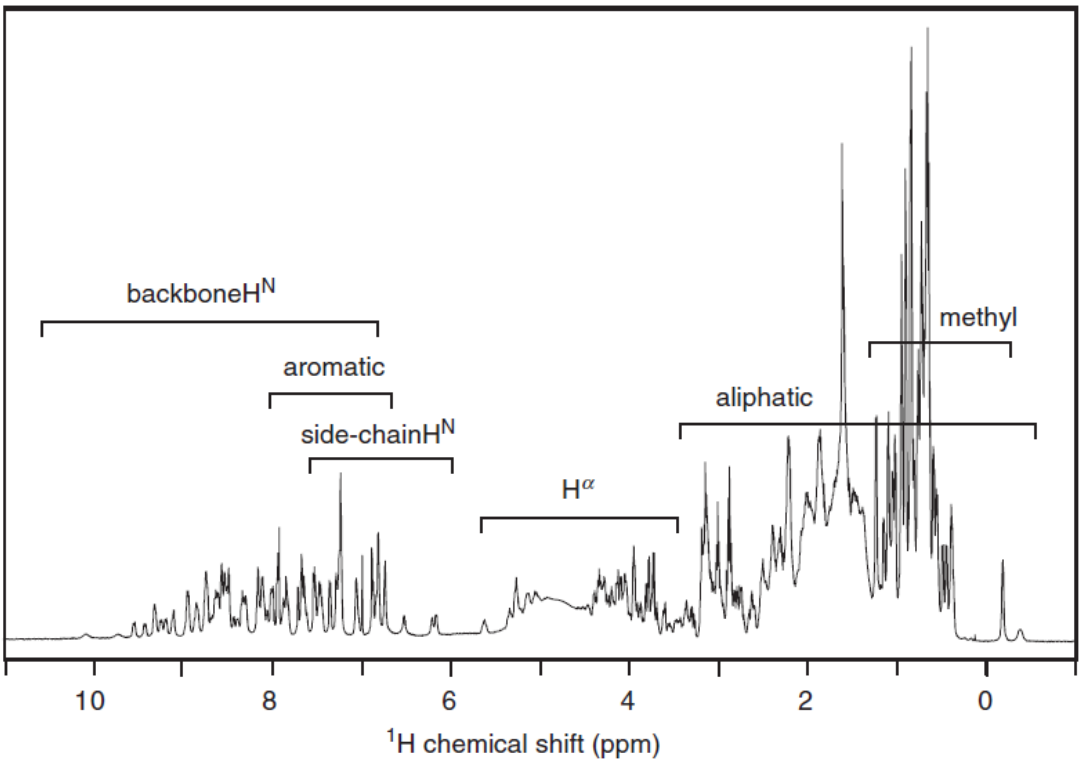

Figure 1.6 – Chemical shift ranges observed for the various types of ^{1}H resonances in ubiquitin. [24]

In any NMR spectroscopic study the first thing to do is to assign each resonance to the specific nucleus of the molecule under investigation. In the case of proteins, as you can see in Figure 1.6, there is a high overlap of the same type of protons in specific ppm regions and thus it can be difficult to assign each resonance to its corresponding nucleus and residue. The strategies used

for resonance assignment with isotopic labeled proteins, ^{15}N and/or ^{13}C, help us overcome this issue facilitating the sequential assignment of resonances. Nitrogen and carbon isotopes are not usually found in native proteins, however, with bacterial growth using ^{15}N and ^{13}C enriched compounds their abundance in protein structure can be enhanced.

We can use multiple radio frequency pulses to detect different kind of nucleus at the same time and thus obtain a wide range of spectra with one or more dimensions. One of the most useful spectra for protein investigation is the Heteronuclear Single Quantum Correlation (HSQC) spectrum (Figure 1.7).

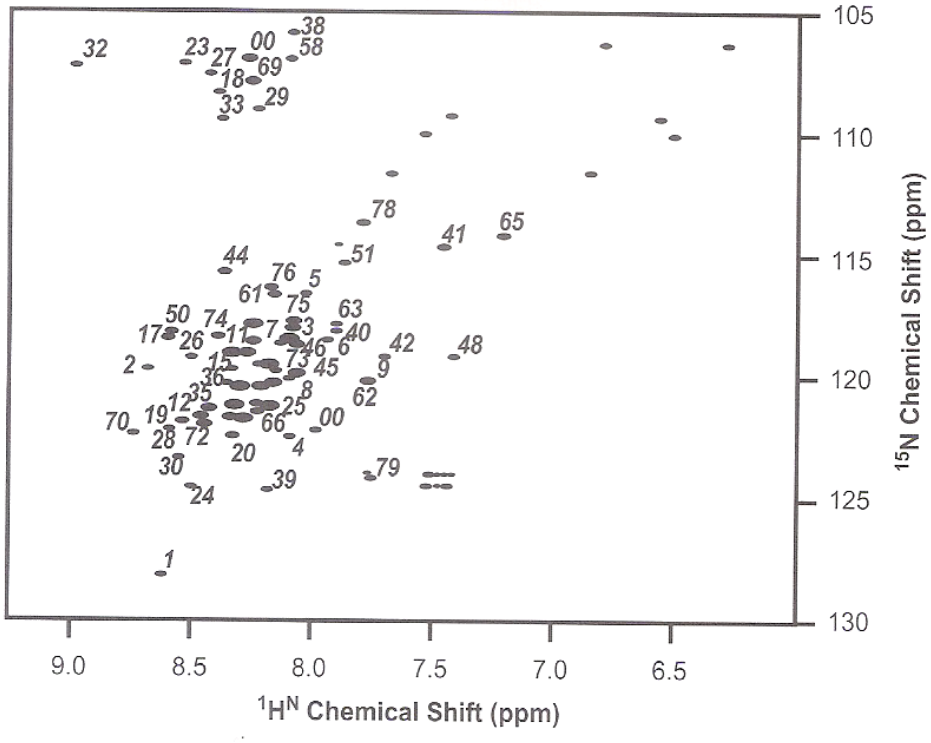

Figure 1.7 – Example of a protein HSQC spectrum. [22]

This spectrum as information of the ^{1}H and ^{15}N that are covalently attached to each other, each signal in the spectrum represents a ^{1}H-^{15}N pair. Every residue, except for proline and the first residue, has such a pair in a peptide bond and produces a resonance. Side chains that contain amides will also produce a resonance signal in this kind of spectrum. 2D ^{1}H-^{15}N HSQC provides useful information for the assignment of the resonances to specific residues and is the base of almost all other spectra used for protein sequential assignment.[22,24]

Protein NMR can be used to study protein interactions whether with other proteins or other type of molecules, structure determination and dynamics, among other applications.

1.6 Differential Scanning Calorimetric

Differential Scanning Calorimetric (DSC) is a thermodynamically tool that assesses the energy uptake in terms of heat by a sample when temperature is being increased or reduced in a

controlled environment. This feature allows DSC to study the biochemical reaction of a molecular transition by a molecule from one conformation to another. The transition temperature (or melting temperature-Tm) is determined for samples in solution, solid or mixed phases such as suspensions. DSC measures the enthalpy changes in a sample by the heat excess it radiates or absorbs when compared to a reference during a change in temperature.[25,26]

There are two types of DSC depending on the mechanism of operation: heat-flux DSC and power compensate DSC. In power compensate DSC sample and reference cells are separately heated and maintained at the same temperature, the energy difference necessary to maintain it on both cells is a measure of the enthalpy changes in the sample. When sample and reference cells are heated by a single furnace and are connected by a low resistance heat flow thermoelectric disk it is called heat flux DSC.[25–29] Figure 1.8 shows a DSC scan from which we can extract the Tm, the change in enthalpy (ΔH) and the change in heat capacity (ΔC_p) of the unfolding process.

DSC can be used to analyze the structural transitions of proteins, nucleic acids, lipids, carbohydrates, and also in nanoscience where it is used to analyze the thermal properties of nanostructures.[26,28]

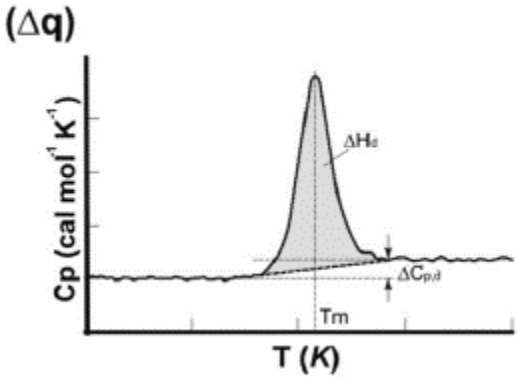

Figure 1.8 – DSC scan in which shows the excess heat changes from a sample compared to a reference as a function of temperature.[26]

2. Materials and methods

2.1 Protein expression and purification

2.1.1 Apo-DVU2108 for NMR studies and Reconstitution experiments

Please note that the cloning, heterologous expression optimization and purification of unlabelled, ^{15}N and $^{13}C/^{15}N$ labelled DvH Apo-DVU2108 was performed by Marta S.P. Carepo, Rui Soares and Sofia R. Pauleta.

The PCR isolated DVU2108 gene was cloned into the NdeI/XhoI restricted pET 21-c expression vector (Novagen) and the Apo-DVU2108 was heterologously expressed in the *E. coli* strain, BL21(DE3) (Invitrogen). Expression of uniformly ^{15}N or $^{13}C/^{15}N$-labelled Apo-DVU2108 was carried out by growing the cells in M9 minimum medium containing either 1 g/L $^{15}NH_4Cl$ or 1 g/L $^{15}NH_4Cl$ and 4 g/L [$^{13}C_6$]-glucose, respectively, as the sole nitrogen and carbon sources. The cells were grown at 37ºC, at 250 rpm, with the protein production induced at an OD_{600nm} of 0.6 with 0.5 mM IPTG, and carried for 16h at room temperature. The cells were harvested by centrifugation and the pellet was resuspended in 50 mM Tris-HCl at pH 7.6, containing protease inhibitors (Complete, Roche). The cell-free soluble extract was obtained by breaking the cells with a French-Press, the cell debris and membrane fraction were removed by low-speed centrifugation followed by an ultracentrifugation at 138000 g.

The purification of the DVU2108 was carried out in two-steps. The soluble cellular extract was diluted in cold water prior to being loaded onto a DEAE-FF (GE HealthCare, ∅ 26 x 100 mm), equilibrated with 10 mM Tris-HCl, pH 7.6. The unbound proteins were removed with the equilibration buffer and DVU2108 was eluted with a gradient between 0 and 500 mM NaCl in 10 mM Tris-HCl, pH 7.6. The fractions containing DVU2108, as judged by SDS-PAGE (12.5% acrylamide Tris-tricine buffer system), were concentrated using a Vivacell (Sartorius Stedim Biotech) apparatus over an YM5 exclusion membrane, at 4 ºC. This concentrated fraction was then loaded onto a size-exclusion column Superdex 75 (GE HealthCare, ∅ 16 x 600 mm), equilibrated with 50 mM Tris-HCl, pH 7.6, 150 mM NaCl. The fractions containing the pure DVU2108 were pooled and concentrated as before, and its buffer was exchanged to 20 mM sodium phosphate, pH 7.0, 100 mM NaCl using a PD10 desalting column (GE HealthCare). The purified heterologously produced DVU2108 was stored at – 80 ºC until further use.

Protein concentration was determined using a modified version of the Lowry Method[30] using BSA (Bovine Serum Albumin) as the standard protein. The yield of the heterologous expression was 50 mg of protein per L of minimum medium, and the purified protein did not present any bound

metal (Zn, Mo, Cu) as determined by ICP (REQUIMTE lab Analysis) and also by the inspection of its UV-visible spectrum.

2.1.2 StrepDVU2108 from a *Desulfovibrio vulgaris* Hildenborough strain

Desulfovibrio vulgaris Hildenborough strain containing a version of DVU2108 gene linked to a Strep-Tag at its N-terminal inserted in the genome, and containing a thiamphenicol resistance mark were grown, in anaerobic conditions, in POSTGATE C medium (see SI 7.1) supplemented with 0.15 mM of antibiotic concentration for 24h at 32ºC. Another approach was made with a POSTGATE C medium supplemented with $CuCl_2$ (0.8µM) and Na_2MoO_4 (0.1µM).
Before sterilization, 10ml and 100ml medium had its atmosphere exchanged from oxygen to nitrogen for 10 min and 30 min, respectively, and then autoclaved at 120ºC for 20 min. For volumes of 1L and 2L the procedure was slightly different. It first underwent an autoclave cycle for 5 min at 90ºC after which, and while still warm, atmosphere was exchanged from oxygen to nitrogen during 1h per liter. Finally, it was autoclaved again at 120ºC for 20 min. The antibiotic was added at the time of the inoculation. Inoculation was carried out in a flame induced sterile environment and with sterile needles and syringes. All inoculums were made with 10% of a previous growth.

For StrepDVU2108 reviving, 10 ml medium volumes were used until it reached a 24h growth. After that, the expression of StrepDVU2108 was carried in three steps. First, two pre-inoculums of 100ml each were prepared. From these, twelve 100ml inoculums were set to grow and finally, 12L of medium were inoculated with those cultures. After cell collection the medium was autoclaved at 120ºC for 30 min for sterilization.

Outside of the anaerobic box, *D. vulgaris* Hildenborough cells where collected by centrifugation at 11305 g in an Avanti J-26 XPI (BECKMAN COULTER) with a JA-10 rotor (BECKMAN) for 20min at 4ºC and then resuspended in 100 mM Tris-HCl, pH 8.0, 150 mM NaCl prior to being lysed by four passages in a French Press (1100 psi). Cellular debris and soluble extract where separated by centrifugation at 45913 g in an Avanti J-26 XPI (BECKMAN COULTER) with a JA-25.50 rotor (BECKMAN) for 1h at 4ºC. Inside the anaerobic box, supernatant was loaded onto a 5 ml StrepTrap HP column (GE Healthcare), equilibrated with 100 mM Tris-HCl, pH 8.0, 150 mM NaCl. Unbounded proteins were washed with the same buffer and StrepDVU2108 was eluted with 100 mM Tris-HCl, pH 8.0, 150 mM NaCl and 2.5 mM desthiobiotin. The fractions containing the protein of interest were identified by SDS-PAGE analysis and concentrated using a Vivacell (Sartorius Stedim Biotech) apparatus over an YM5 exclusion membrane, at 4ºC, and under a flush of argon.

Protein concentration was determined using a modified version of the Lowry Method[30] using BSA (Bovine Serum Albumin) as the standard protein. Metal concentrations were determined by ICP

(REQUIMTE lab Analysis) for Mo, Cu and Fe. UV-visible spectroscopy analysis was performed in an UV-1800 spectrophotometer (Shimadzu).

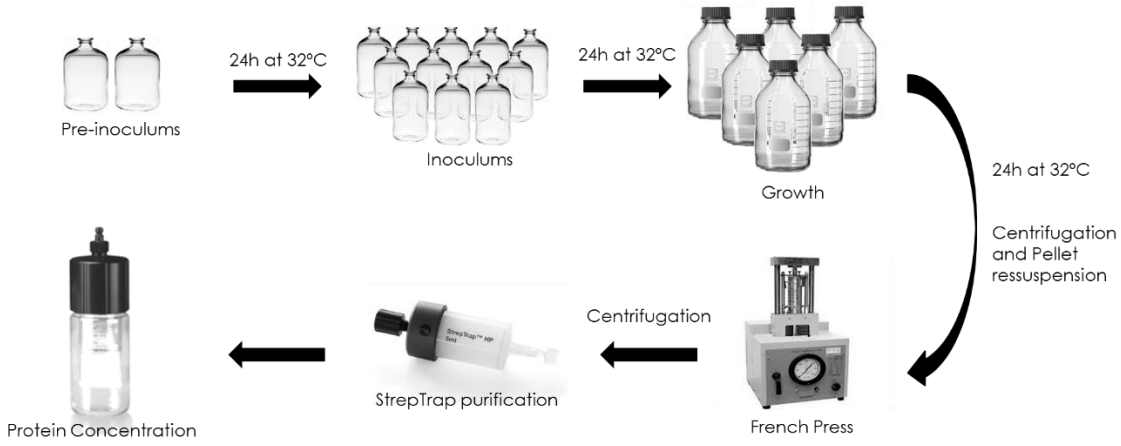

Figure 2.1 – Scheme for the expression and purification steps of StrepDVU2108

2.2 Apo-DVU2108 mediated cluster reconstitution with a Mo-Cu cluster or a W-Cu cluster

Tetrathiomolybdate (TTM), tetrathiotungstate (TTW) and copper solutions were prepared in 50mM Tris-HCl pH 7.6, 150mM NaCl, 20% DMF (SIGMA) using $(NH_4)_2MoS_4$ (SIGMA), $(NH_4)_2WS_4$ (SIGMA) and $CuCl_2 \cdot 2H_2O$ (Merck) always fresh and prior to being used.
The copper solution was added to 30 µM of Apo-DVU2108 followed by addition of TTM or TTW solution in the corresponding ratio. The mixture was them incubated for 15 min at room temperature. The initial metal to protein ratio used was 2.2 Mo: 1 Cu: 1 Apo-DVU2108 and 2.1 W: 1 Cu: 1 Apo-DVU2108.

Size exclusion chromatography was performed in order to separate the protein from the unbound metals, thus the mixture was applied onto a PD10 Column (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.6, 150 mM NaCl and eluted in the same buffer.

The metal content of the reconstituted-DVU2108 was determined by ICP (REQUIMTE lab Analysis) using the Reagecom ICP multielements as a standard solution in a concentration range of 0–3 ppm for each metal (W/Mo and Cu), except for Cu content in Mo reconstituted-DVU2108 which was determine by 2,2'-biquinoline method. Protein concentration was determined using a modified version of the Lowry Method[30] using BSA (Bovine Serum Albumin) as the standard protein. The reconstituted protein was analyzed by UV-visible spectroscopy in an UV-1800 spectrophotometer (Shimadzu).

2.3 DVU2108 reconstitution ratio analyzed by titration with TTW and TTMo

For these experiments $CuCl_2$ was added to a stirring solution containing 30 µM of Apo-DVU2108 to a 1:1 ratio, to which small aliquots (5 µL) of 0.51 mM TTMo or 0.48mM TTW solution were sequentially added until the Cu:Mo/W ratio reached 1:4. Spectra were acquired in an UV-1800 spectrophotometer (Shimadzu) after each addition and a 1.5 min stirring time and also before and after $CuCl_2$ was added. Metal solutions were prepared in the same way as for the reconstitution experiments, see 1.2 Materials and Methods. The buffer used was also same.
Control experiments were performed in the absence of the protein in solution.

2.4 NMR spectroscopy

For the NMR experiments, three protein samples were prepared, the unlabeled, the uniformly ^{15}N-labelled, and the $^{13}C/^{15}N$-labeled Apo-DVU2108 following the procedure described above. Samples were 1.0 mM in protein concentration in 20 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM sodium azide, 1 mM DTT and 10 % $2H_2O$.

NMR experiments were carried out at 298 K on Bruker AvanceIII 600MHz spectrometer equipped with a TCI cryoprobe, located at UCIBIO, REQUIMTE, Chemistry Department, Science and Technology Faculty, NOVA University of Lisbon. The sequence assignments of the protein backbone resonances were obtained using 2D ^{1}H-^{15}N–HSQC, 3D HNHA, 3D HNCO, 3D HN(CA)CO, 3D CBCA(CO)NH, 3D CBCANH, 3D HBHA(CBCACO)NH, 3D (H)C(CCO)NH and 3D H(CCCO)NH spectra[31]. For the side chain assignment 2D ^{1}H-^{13}C–HSQC and 3D (H)CCH-TOCSY experiments were performed[31]. The ^{1}H spin systems of the aromatic rings of His residues were identified using a 2D ^{1}H-^{15}N HSQC type spectrum with an INEPT constant optimized for ^{2}J detection. The assignment of the aromatic side chains was performed with the analysis of the ^{1}H-^{13}C HSQC together with 3D ^{13}CNOESY[31].

(The following spectra were acquired during my Master thesis: 3D H(CCCO)NH, 3D (H)C(CCO)NH, 3D ^{13}CNOESY and the 2D ^{1}H-^{15}N HSQC for the ^{1}H spin systems identification of the aromatic rings of His residues, while the remaining spectra were previously acquired by S. R. Pauleta)

Spectra processing was made using TopSpin™ software and spectra analysis and sequential assignment was performed with CARA™ software.

Table 2.1 – Spectra acquisition parameters

	Pulse Sequence	Number of Scans (NS)	Sweep Width (SW) (ppm)			Number of Points (TD)			O1 (ppm)	O2 (ppm)	O3 (ppm)
			F1	F2	F3	F1	F2	F3			
Backbone Assignment											
2D ^{1}H-^{15}N HSQC	hsqcetfpf3gpsi2	2	14.03	34		1024	128		4.7	118.49	
3D HNHA	hnhagp3d	16	14	16.02	34	128	2048	48	4.7	51.99	118.49
3D HNCO	hncogp3d	4	14.03	21.99	34	2048	80	48	4.7	172.97	118.49
3D HN(CA)CO	hncacogp3d	4	14.03	21.99	34	2048	26	48	4.7	172.97	118.49
3D CBCA(CO)NH	cbcaconhgp3d	8	14.03	74.96	34	2048	80	48	4.7	38.99	118.49
3D CBCANH	hncacbgp3d	8	14.03	74.96	34	2048	80	48	4.7	41.99	118.49
3D HBHA(CBCACO)NH	hbhaconhgp3d	8	16.02	16.02	34	80	2048	48	4.7	39.99	118.49
3D (H)C(CCO)NH	ccconhgp3d	16	14.03	72.03	34	1024	80	48	4.7	39.99	118.49
3D H(CCCO)NH	hccconhgpwg3d2	16	14	14.03	34	80	1024	48	4.7	39.99	118.49
Side Chain Assignment											
3D (H)CCH-TOCSY	hcchdigp3d2	8	14.03	72.03	72.03	1026	128	64	4.7	39.99	39.99
2D ^{1}H-^{13}C HSQC	hsqcetgpsi2	4	16.02	169.9		2048	256		4.7	74.99	
Aromatic Side Chain Assignment											
3D ^{13}CNOESY	noesyhsqcetgp3d	16	14	80.03	16.02	200	48	2048	4.7	39.99	118.49
Aromatic Rings of His Residues											
2D ^{1}H-^{15}N HSQC	hsqcetfpf3gpsi2	104	24.03	200.01		2048	256		4.7	174.97	

2.5 Differential Scanning Calorimetry

Differential Scanning Calorimetry (DSC) was carried out on a Nano-DSC system (TA instruments), in 300 µL cells. A control experiment was made with only 50 mM TrisHCl, pH 7.6, 150mM NaCl buffer on both sample and reference cells with 8-altering heating and cooling cycles at a rate of 1ºC/min from 10 to 100ºC and a pressure of 3 atm, followed by an isothermal cycle to 20ºC and zero pressure. The sample reading was carried with buffer in the reference cell and protein sample on the sample cell at a heating rate of 1ºC/min and with 3 atm of pressure. Protein concentrations used were 0.747 mg/mL for Apo-DVU2108, 0.656mg/mL and 0.747mg/mL for Mo- and W-reconstituted DVU2108, respectively.

Apo-DVU2108 stock solution was bufferexchanged using a NAP-5 column (GE Healthcare) to 50 mM TrisHCl, pH 7.6, 150mM NaCl. Reconstituted-DVU2108 proteins were prepared the same way as before but with an initial protein concentration of 150 µM. Since DVU2108 has two Cys residues which are not involved in a sulfur bridge but could form one upon denaturation, 1 mM of tris(2-carboxyethyl)phosphine (TCEP), as reducing agent was added to the Apo-DVU2108 sample. For the reconstituted proteins TCEP was not added since in its presence the metal cluster was lost over time.

DSC data were analyzed with the NanoAnalyze™ Software (TA instruments), from which the thermal transition temperature (Tm) and the change in enthalpy (ΔH) of the unfolding process were obtained.

3. Results

3.1 DVU2108 expression and purification in a *Desulfovibrio vulgaris* Hildenborough strain

The DVU2108 homologue, ORP, was isolated from *D. gigas* in its native form and was shown to have a copper-molybdenum cluster in its structure.[5] The Apo form of this protein can incorporate this metal cluster when molybdenum and copper are available[32] which raises the question if the native form has in fact a molybdenum-copper cluster or if metals were present in the medium and the protein incorporated the metal cluster during the isolation process. In order to study the native DVU2108 and elucidate if it has a copper-molybdenum cluster we have expressed and purified it in *Desulfovibrio vulgaris* Hildenborough (DvH) under anaerobic conditions. The strain was genetically modified to knockout the *DVU2108* gene and to introduce in the chromosome a version of *DVU2108* linked to a StrepII-Tag at its N-terminal to facilitate its purification. This strain was grown in POSTGATE C medium with and without copper and molybdenum supplementation. The isolation and basic biochemical characterization of DVU2108 from both growths are reported below.

3.1.1 Standard POSTGATE C medium

In the first growth, the genetically modified DvH/StrepDVU2108 was grown in the POSTGATE C medium (12 L in total, grown in 2 L anaerobic flasks at 30ºC). In order to prepare the inoculum of the 12 L, it was necessary to grow two pre-inoculums from which twelve inoculums would be prepared. The first two 100 mL pre-inoculums grew to an $OD_{600\ nm}$ of 0.766 and 0.792, the twelve 100 mL inoculums grew to an average $OD_{600\ nm}$ of 1.025, and for the 12 L growth, the averaged $OD_{600\ nm}$ at the end of the growth was 0.727. Each of these growths had a duration of approximately 24 h. During this process, the culture was also monitored by microscopy to examine whether there was any changes in the morphology or purity of the culture. As can be observed by the images in Figure 3.1, there was no changes to report in the size and shape of the cells and no contamination could be detected. The wet cell mass obtained in this growth was ~20g.

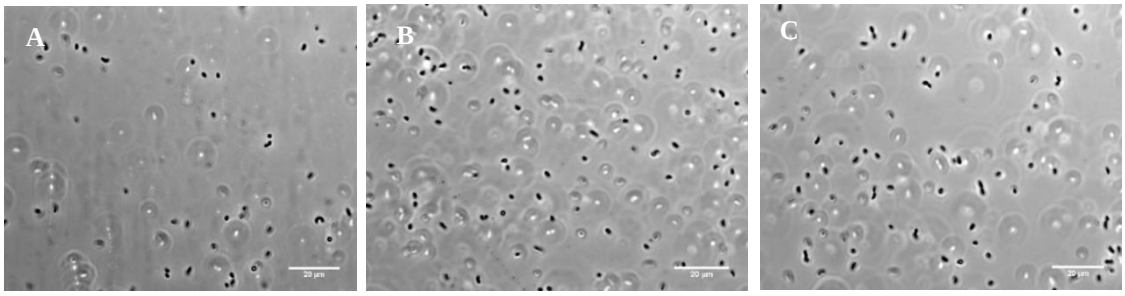

Figure 3.1 – Optic microscopic images of samples from the pre-inoculum (A), inoculum (B) and end of 12L growth (C) with 100x objective.

After cell harvesting and cell lyses followed by another centrifugation, the supernatant, containing the protein of interest, was purified under anaerobic conditions in a Coy anaerobic chamber. Since DVU2108 had an N-terminal StrepII-tag, a chromatographic step using a highly cross-linked agarose (6 %) matrix with StrepTactin as a ligand was performed, and the protein was eluted with 100 mM Tris-HCl, pH 7.6, 150 mM NaCl, and 50 mM desthiobiotin. The fractions obtained were analyzed in a SDS-PAGE (Figure 3.2).

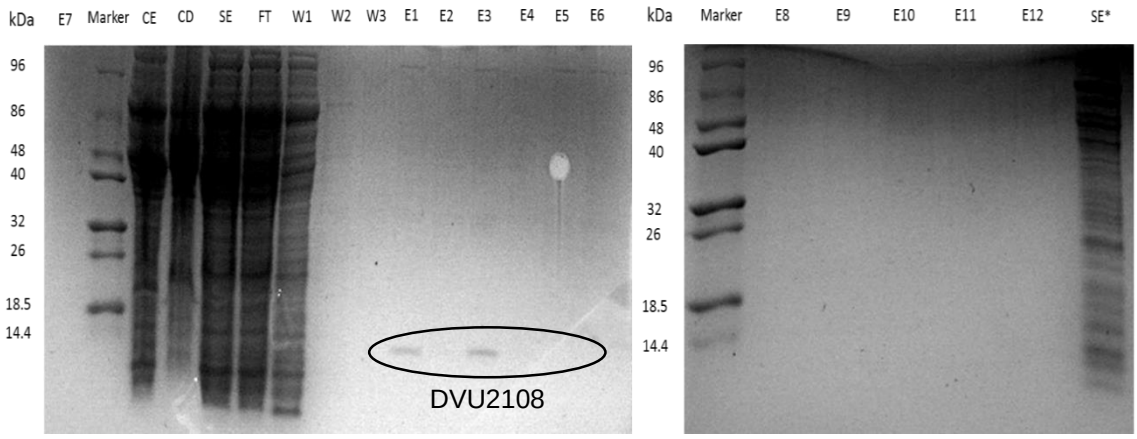

Figure 3.2 – 12.5% SDS-PAGE after coomassie staining. Gel was run during 1h at 150V. (Marker – protein marker; CE – cellular extract; CD – cellular debris; SE – Soluble extract; FT – Flow through; W1-3 – Washes; E1-12 – elution samples; SE - Soluble extract lane with less loaded quantity.) Black ellipse shows* DVU2108 *bands.*

As mentioned in the Introduction (Table 1.1) DVU2108 has a molecular mass of 12.5 kDa. Therefore, the analysis of the SDS-PAGE indicates that fractions E1-3 had a band with the expected size. Moreover, the color of E3 was slightly yellow/brown. A UV-visible spectrum was acquired for this fraction, and both protein and metals (Mo, Cu, and Fe) were quantified (Table 3.1)

The fraction E1 - E5 were concentrated to 1 ml, and its protein and metal content were also quantified. The UV-visible spectra of fraction E3 and of the concentrated fractions E1-5 are shown in Figure 3.

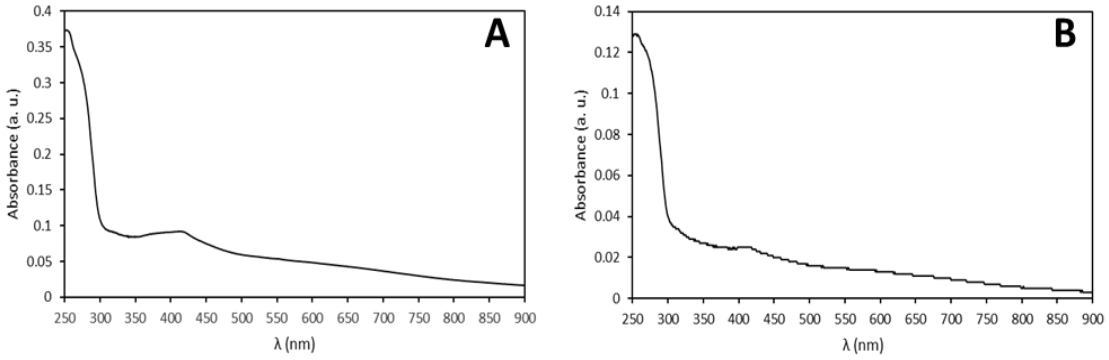

*Figure 3.3 – UV-visible spectra of fraction E3 (**A**) (without dilution) and concentrated protein fraction E1-E5 (**B**) (diluted 1:9) in 100 mM Tris-HCl, pH 7.6, 150 mM NaCl.*

Both spectra present a distinct broad absorption band at about 400 nm and another around 550 nm that could be of a Fe-S protein[33]. At this point, we can only speculate that this absorption features could either correspond to a Fe-S cluster that is bound to StrepDVU2108 or to one of the contaminant proteins, as StrepDVU2108 is not pure (Figure 3.4)

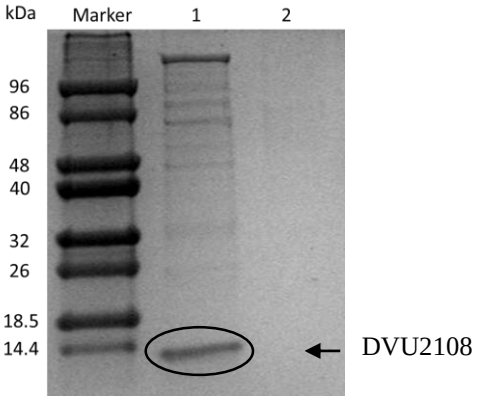

Figure 3.4 – 12.5% SDS-PAGE after Coomassie staining. Gel was run during 1h at 150V. (Marker – protein marker; Lane 1 – protein sample after concentration; Lane 2 – Flow through after concentration). Black ellipse shows DVU2108 band.

In fact, the analysis of SDS-PAGE of StrepDVU2108 indicate that there are a significant number of contaminants that elute together with StrepDVU2108 even in the presence of 150 mM NaCl, and thus another chromatographic step should be performed, such as a gel filtration chromatography (as most contaminant proteins have a higher molecular mass and DVU2108 homologues behave as monomers in solution). However, the amount of protein sample that remained was not enough to perform this additional step.

Table 3.1 shows protein and metal concentrations obtained for both fraction E3 and concentrated protein (the amount of protein sample was not enough for duplicates in ICP analysis and protein quantification, thus errors cannot be determined for metal and protein concentrations)

Table 3.1 –Protein and metals quantification of StrepDVU2108 fraction E3 and concentrated fractions E1-5

	[Protein] (mg/ml)	[Mo] (µM)	[Cu] (µM)	[Fe] (µM)
Fraction E3	-	0.09	2.98	9.40
StrepDVU2108 E1-5	0.30	0.45	13.50	41.39

The metal quantification indicates that the amount of Mo is almost negligible, and that of Cu was lower than Fe, which again suggests that concentrated StrepDVU2108 fractions E1-5 contains an iron protein. Normalizing to the amount of Mo, the sample presents a metal molar ratio of 1Mo:30Cu:92Fe.

3.1.2 POSTGATE C medium supplemented with 0.8 µM Cu and 0.1 µM Mo

Given the previous results another approach was made to grow the DvH/StrepDVU2108 by supplementing the growth medium with 0.8 µM $CuCl_2 \cdot 2H_2O$ and 0.1 µM $Na_2MoO_4 \cdot 2H_2O$, concentrations that would have no major effect on the DvH growth.[34,35]

The two pre-inoculums grew to an $OD_{600\ nm}$ of 0.636 and 0.812, the twelve inoculums grew to an average $OD_{600\ nm}$ of 1.001, and for the 12 L growth the averaged $OD_{600\ nm}$ was 0.599. Like in the previous growth the bacteria exhibited normal size and shape for this strain (Figure 3.5). The wet cell mass obtained was ~18 g, which is less than obtained in the first approach, which correlates with the lower $OD_{600\ nm}$ observed throughout the growth.

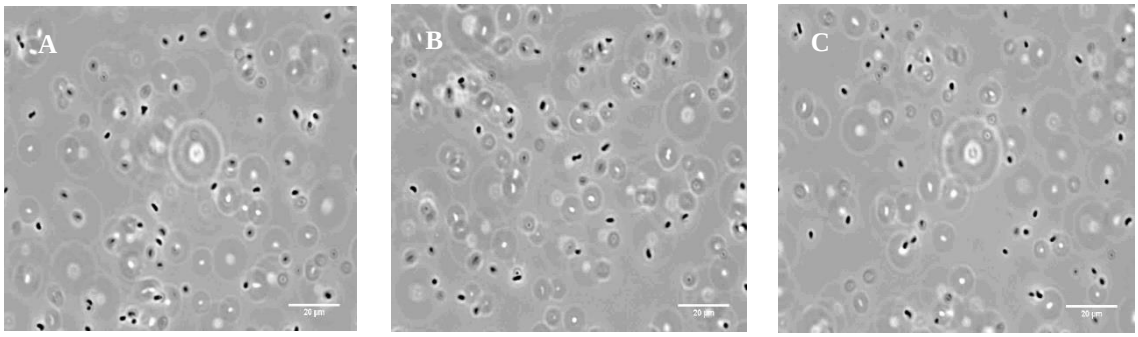

Figure 3.5 – Optic microscopic images of samples from the pre-inoculum (A), inoculum (B) and end of the 12 L growth (C) with 100x objective.

The SDS-PAGE analysis of protein expression is shown in Figure 3.6. This time we eluted the protein in fewer fractions with higher volume (5 mL) since in the first experience the final fractions showed no protein whatsoever. Once again there was a fraction with a different color than the others, in this case fraction E2, and so a spectrum was acquired (Figure 3.7 - A).
Isolation of StrepDVU2108 was accomplished although with probably the same contaminants as before that we can better detect in this second purification attempted.

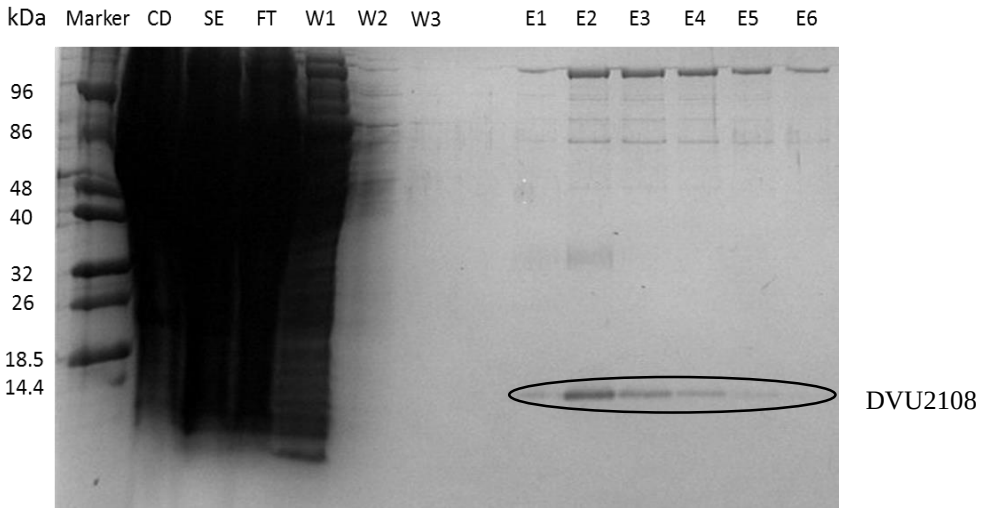

Figure 3.6 – 12.5% SDS-PAGE after Coomassie staining of fractions obtained during purification of StrepDVU2108 from a growth with Cu/Mo supplementation. Gel was run during 1h at 150V. (Marker – protein marker; CD – cellular debris; SE – Soluble extract; FT – Flow through; W1-3 – Washes; E1-6 – elution fractions.) Black ellipse shows DVU2108 bands.

The spectrum of fraction E2 is similar to the one obtained for fraction E3 from the previous purification, but absorption bands are much more intense. Due to the presence of similar contaminant proteins, the question remains as if these absorption features are due to the contaminant proteins or the fact that DVU2108 is a Fe-S protein.

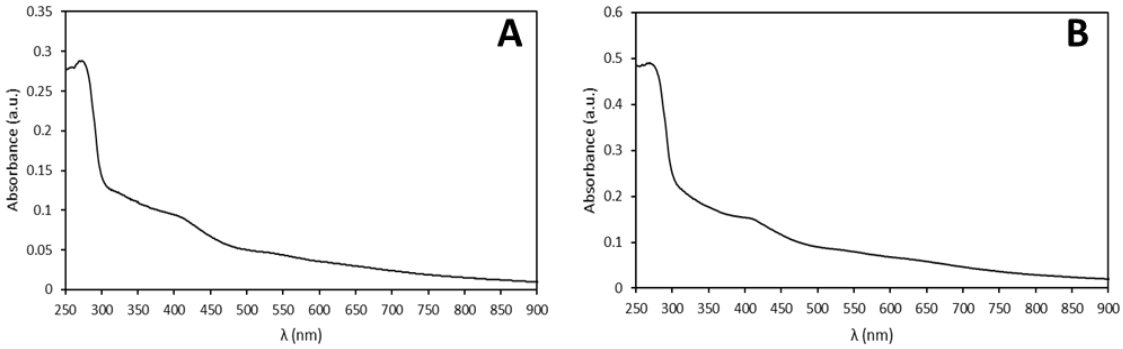

*Figure 3.7 – UV-visible spectra of Fraction E2 (**A**) and concentrated protein fractions E1-5 (**B**) in 100 mM Tris-HCl, pH 7.6, 150 mM NaCl.*

Fraction E1 to E5 were concentrated to 2.5 mL which was then analyzed by SDS-PAGE (Figure 3.8 - A) and a spectrum was acquired (Figure 3.7 - B). Protein and metal concentrations are shown in Table 3.2 (the amount of protein sample was not enough for duplicates in ICP analysis and the errors on these quantifications could not be determined). For this concentrated protein fraction, a PAGE was also performed (Figure 3.8 - B).

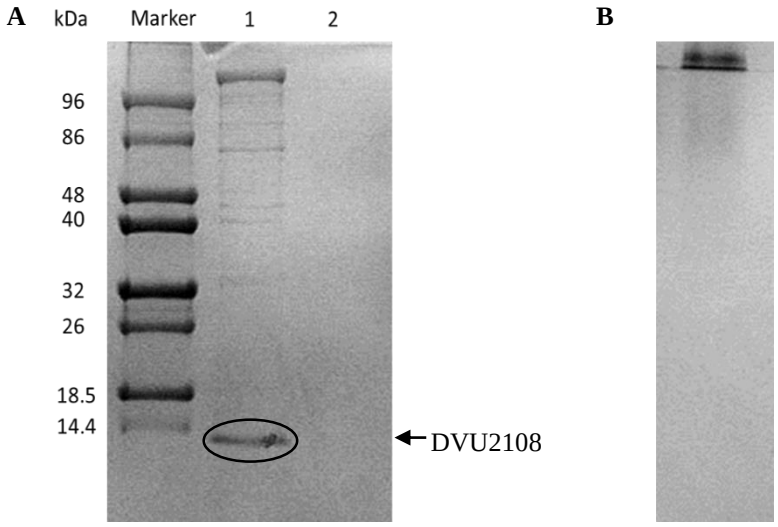

*Figure 3.8 – **A** -12.5% SDS-PAGE after Coomassie staining. Gel was run during 1h at 150V. (Marker – protein marker; Lane 1 – protein fraction E1-E5 after concentration; Lane 2 – Flow through after concentration) **B** - 12.5% PAGE after Coomassie staining. Gel was run during 1h at 150V*

Once again the StrepDVU2108 after the first purification step is not pure. This was expected given the first SDS-PAGE to analyze the elution fractions. The PAGE analysis was inconclusive, as most protein did not migrate into the gel and only a very diffuse and very low intensity band is observed probably due to protein complex formation (large complex that cannot enter in the resolution gel, band diffusion might be due to different forms of DVU2108).

In this purification, the concentrated sample had higher protein and molybdenum concentrations but lesser copper and iron. This means we have fewer clusters per protein, although we cannot

really tell why that happened. We expected a higher copper and molybdenum but only the last increased and it was not a great increase.

Table 3.2 – Protein and metals concentration of StrepDVU2108 concentrated fractions E1-5

	[Protein] (mg/ml)	[Mo] (µM)	[Cu] (µM)	[Fe] (µM)
StrepDVU2108 E1-5	0.391 ± 0.006	0.73	1.96	14.85

Surprisingly in this case the Cu and Fe content seems lower (normalizing to the concentration of Mo, the sample presents a metal molar ratio of 1Mo:2.7Cu:20Fe), but again the metal in higher concentration was Fe, with the amount of Mo being very small.

3.2 Apo-DVU2108 metal cluster reconstitution

The novel metal cluster composed by copper, molybdenum and sulfur is a distinctive characteristic of ORP isolated from *D. gigas*.[5] Recent studies showed that it is possible to reconstitute an Apo form of ORP by adding copper chloride ($CuCl_2$) and tetrathiomolybdate (TTMo) in the right proportions to a protein solution.[32] Since DVU2108 is a homologous protein of *D. gigas* ORP we have used the same strategy to reconstitute the Apo-DVU2108 with this metal cluster, even though it was not possible to isolate the native protein with the cluster.

Apo-DVU2108 was purified prior to the start of this Master project. Nevertheless, it was necessary to quantify the protein (Table 3.3), perform a SDS-PAGE analysis (Figure 3.9) and acquired an UV-visible spectrum (Figure 3.10) to determine its extinction coefficient and verify its purity.

Table 3.3 – Apo-DVU2108 concentration and extinction coefficient. Absorbance and concentration value corresponds to a 1:20 dilution of protein sample.

[Apo-DVU2108] (µM)	Absorbance at 278nm	ε_{278nm} (M^{-1} cm^{-1})
165 ± 5	0.845	5130 ± 160

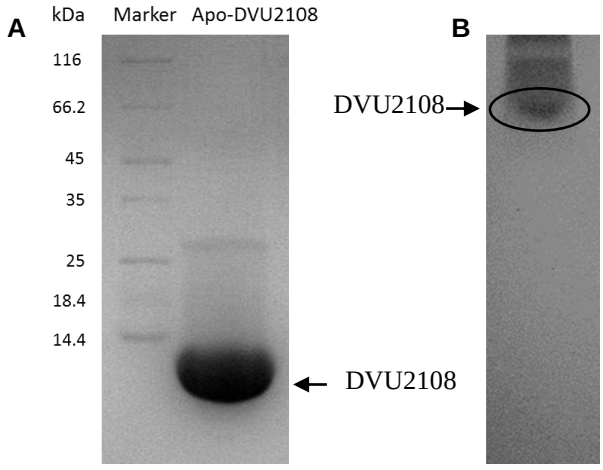

*Figure 3.9 – **A** -12.5% SDS-PAGE of Apo-DVU2108 sample, after Coomassie staining. Gel was run during 1h at 150V. **B** – 10% PAGE after Coomassie staining. Gel was run during 1h at 150V*

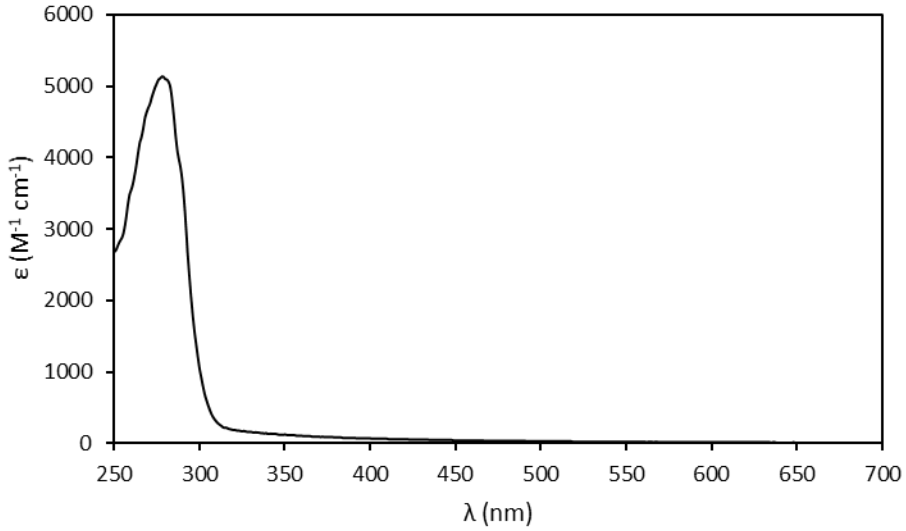

Figure 3.10 – UV–visible spectrum of Apo-DVU2108 in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl.

Analysis of the SDS-PAGE shows that the protein is pure. Apo-DVU2108 has a molecular mass of 12456 Da, from its polypeptide chain (determined in EXPASY-Protparam), and a single band is observed below the 14.4 kDa marker accordingly. The spectrum shows a single absorbance band at about 280 nm which is characteristic of most proteins mostly due to the presence of tryptophan and tyrosine residues and also phenylalanine[36], and no absorption band is observed in the visible part of the spectrum, which is indicative of the absence of bound metals.

3.2.1 Apo-DVU2108-assisted metal cluster reconstitution with copper (II) chloride and tetrathiomolybdate

To a solution containing 30 μM of Apo-DVU2108 were added 30 μM $CuCl_2$ prior to the addition of 66 μM tetrathiomolybdate (TTMo) in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF, in order to have a 1:1 protein-copper ratio and to ensure a 1:2 copper-molybdenum ratio, and the slight excess in TTMo concentration was to reduce the Cu(II) to Cu(I), which is the oxidation state of Cu in the heteronuclear cluster of *D. gigas* ORP.[32] This solution was them incubated for 15 min at room temperature and the unbound metals were separated with a size exclusion chromatography.

The UV-visible spectrum of the Mo-Cu reconstituted DVU2108 after removal of the unbound metals is shown in Figure 3.11. The spectrum shows two absorption maxima at 334 nm (10500 ± 280 $M^{-1}cm^{-1}$) and 483 nm (5570 ± 150 $M^{-1}cm^{-1}$) as well as the characteristic shoulder around 433 nm and is similar to the one obtained for its homologous Orange Protein from *D. gigas* (SI 7.2).[32] Extinction coefficients were determined relative to molybdenum concentration.

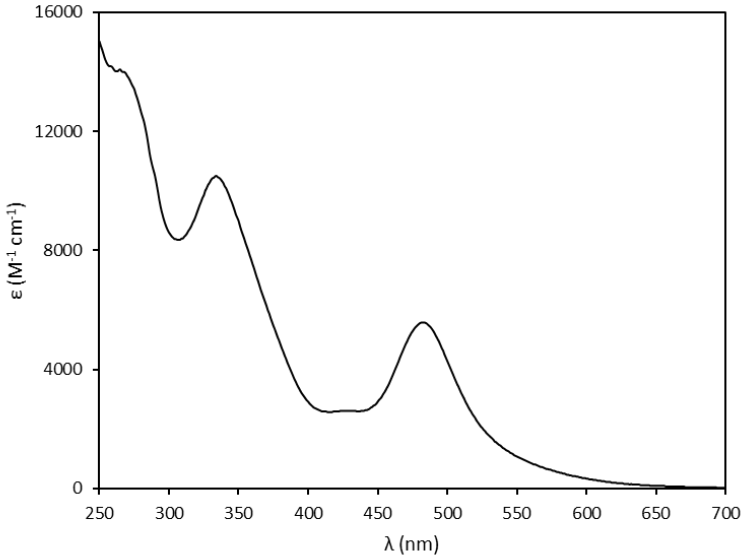

Figure 3.11 – UV–visible spectrum of DVU2108 reconstituted with $CuCl_2$ and TTMo in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl. Extinction coefficient were determined relative to molybdenum concentration.

The stoichiometry of copper per molybdenum atoms and the amount of cluster per protein after reconstitution are listed in Table 3.4 and were close to the expected, since the native *D. gigas* ORP has one cluster per protein (1:1 ratio) and the cluster has a copper atom for each two molybdenum atoms (1:2 ratio). The differences observed can be due to errors associated with metal, namely copper, and protein quantifications.

Table 3.4 – Protein and metal concentrations and ratios for MoCu-DVU2108

[Protein] (μM)	[Cu] (μM)	[Mo] (μM)	DVU2108:Cu	DVU2108: Mo	Cu:Mo
25.3 ± 0.6	19 ± 6	46 ± 1	1:0.8 ± 0.2	1:1.82 ± 0.01	1:2.3 ± 1.2

Both absorption maxima and extinction coefficient values are slightly different from the ones of *D. gigas* ORP, but within the error intervals and are compared in Table 3.5.

Table 3.5 – Extinction coefficient and absorption maxima of MoCu-DVU2108 and reconstituted with TTMo and $CuCl_2$ and native D. gigas ORP.

	MoCu-DVU2108		*D. gigas* MoCu-ORP[32]		Native *D. gigas ORP*[32]	
Maximum absorption (λ_{max})	334 nm	483 nm	338 nm	480 nm	338 nm	480 nm
ε (M^{-1} cm^{-1})	10500 ± 280	5570 ± 150	11400 ± 800	6300 ± 360	11000	5500

3.2.2 Apo-DVU2108-assisted metal cluster reconstitution with copper (II) chloride and tetrathiotungstate

To test if, similarly to *D. gigas* ORP[32], Apo-DVU2108 can also be reconstituted with a similar cluster containing tungsten instead of molybdenum the same procedure was used but with tetrathiotungstate (TTW) instead of tetrathiomolybdate (TTMo).

The approach was identical to the one described for MoCu-DVU2108 with the only difference being the concentration of tetrathiotungstate, 0.63 μM instead of 0.66 μM, which proved to be enough to reconstitute Apo-DVU2108 with this type of cluster.

The UV-visible spectrum of the W-Cu reconstituted DVU2108 after removal of the unbound metals is shown in Figure 3.12. Similarly to what was obtained for the Mo-Cu reconstituted protein, the spectrum is similar do the spectrum obtained for the WCu *D. gigas* ORP (SI 7.3), with absorption maxima at 295 nm (14560 ± 80 $M^{-1}cm^{-1}$) and 407 nm (6990 ± 40 $M^{-1}cm^{-1}$). This time extinction coefficients were determined relative to the concentration of tungsten. Cluster number per protein ratio obtained for WCu-DVU2108 was close to the expected but copper:tungsten ratio was slightly lower than the expected ratio of one copper atom per two tungsten atoms. These results are listed in Table 3.6.

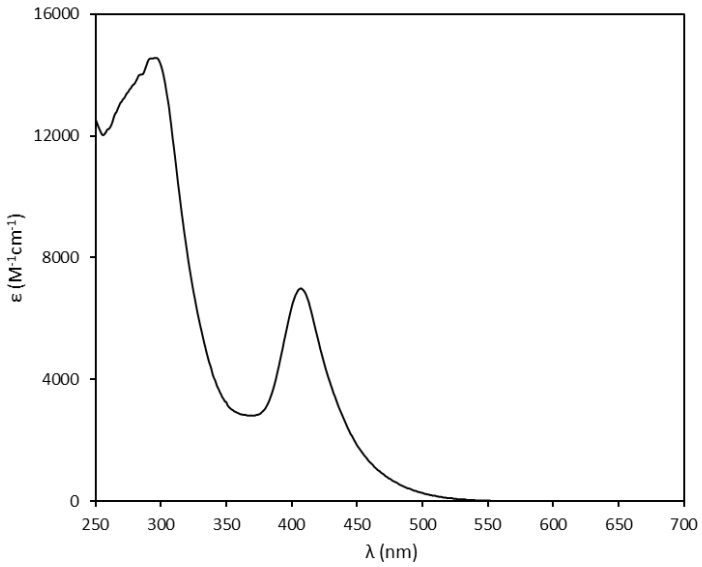

Figure 3.12 – UV–visible spectrum of DVU2108 reconstituted with $CuCl_2$ and TTW in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl. Extinction coefficient where determined relative to tungsten concentration

Table 3.6 - Protein and metal concentrations and ratios for WCu-DVU2108

[Protein] (μM)	[Cu] (μM)	[W] (μM)	DVU2108:Cu	DVU2108:W	Cu:W
35 ± 2	33.6 ± 0.4	50.1 ± 0.2	1:0.96 ± 0.05	1:1.43 ± 0.08	1:1.49 ± 0.03

Once again there were differences in the absorption spectrum between the WCu reconstituted DVU2108 and the reconstituted *D. gigas* homologue which are listed in Table 3.7.

Table 3.7 – Extinction coefficient and absorption maxima comparison of WCu-DVU2108 to D. gigas ORP reconstituted with TTW and Cu using a similar procedure.

	WCu-DVU2108		*D. gigas* WCu-ORP[32]	
Maximum absorption (λ_{max})	295 nm	407 nm	300 nm	405 nm
ε ($M^{-1} cm^{-1}$)	14560 ± 80	6990 ± 40	11300	5200

3.3 Metal Titration of Apo-DVU2108

Apo-DVU2108 was titrated in the presence of $CuCl_2$ with TTMo or TTW to determine the metal stoichiometry of the reconstituted protein, and assess whether the protein has a preference for a specific metal ratio. A control experiment was performed in the absence of protein.

3.3.1 Apo-DVU2108 titration with TTMo

A solution containing 30 µM Apo-DVU2108 and 30 µM $CuCl_2$ in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF was titrated with small aliquots of 0.51 mM TTMo and spectra was acquired after each addition. Spectra of the main Cu:Mo ratios additions are shown in Figure 3.13.

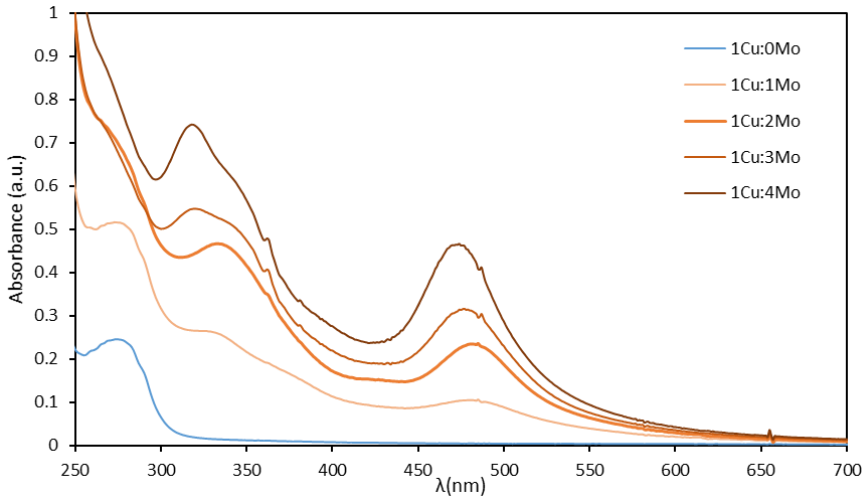

Figure 3.13 – Titration of 30 µM Apo-DVU2108, in the presence of 30 µM $CuCl_2$, with TTMo in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF. Spectra of titration at specific Cu:Mo ratios. Blue: 1Cu:0Mo; Orange: increasing Mo concentration from lighter to darker shades of orange.

As we can see in this Figure, when the ratio of Cu:Mo reaches 1:3, the maximum absorption bands shift to lower wavelengths, which are similar to the ones of the TTMo solution itself, that has the maximum absorption bands at 317 and 468 nm (Figure 3.14)

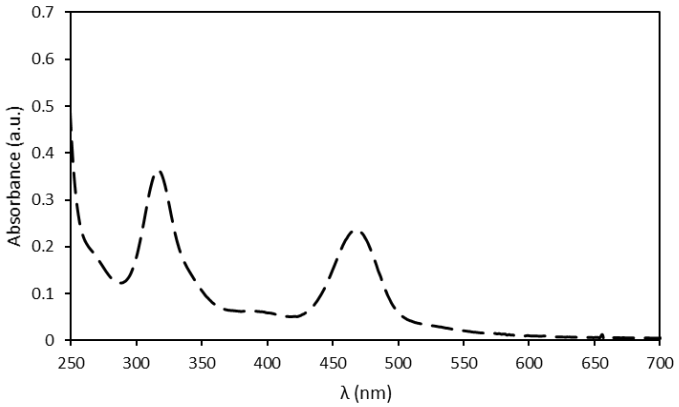

Figure 3.14 – Spectrum of 25.5 µM of tetrathiomolybdate solution in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF.

We have also analyzed the absorbance of the shoulder region in MoCu-DVU2108, at 433 nm, and subtracted the contribution of excess TTMo solution considering the absorbance of 5 µL of that solution. This absorbance is represented as a function of [Mo]/[Cu] ratio in Figure 3.15, and indicates that there is a stabilization at around 2:1 metal ratio. In addition to that, when we examine the wavelength at which the maximum absorbance occurs along the titration (Figure 3.16), we observe a similar behavior: the wavelength of the maximum absorbance reaches the expected values at 2:1 Mo:Cu ratio followed by a stabilization and then a decrease to lower values.

These results confirm that the metal cluster ratio preferred by the reconstituted protein is 2Mo:1Cu.

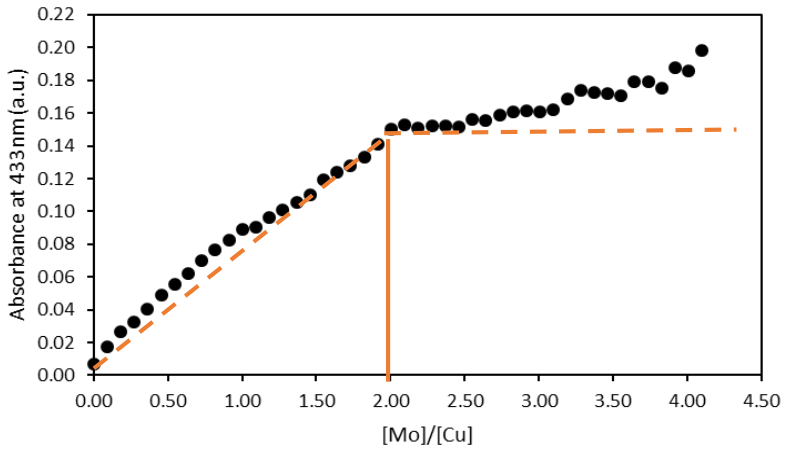

Figure 3.15 – Absorbance variation at 433 nm at each [Mo]/[Cu] ratio

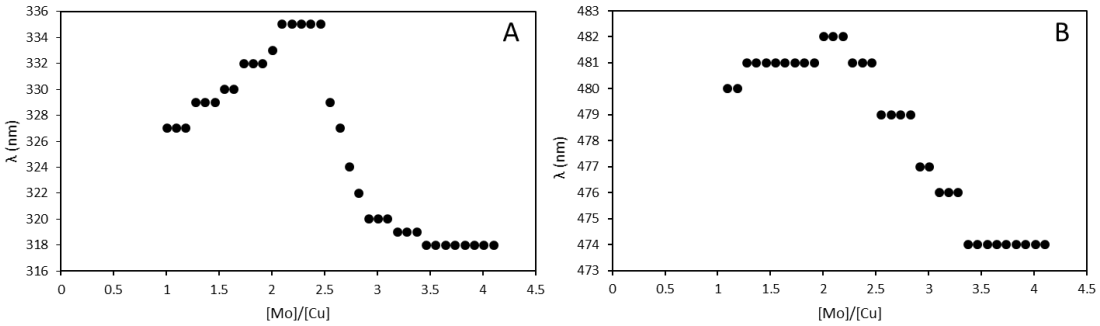

Figure 3.16 – Wavelengths at which is observed maximum absorption along the titration, for maximum absorption band at around 334 nm (A) and wavelengths around 482 nm (B)

The titration in the absence of Apo-DVU2108 shows very different spectra (Figure 3.17) when compared to the ones obtained in its presence (Figure 3.13). In solution TTMo is able to complex with Cu(II) and this results in a total loss of UV-visible spectra from both metal solutions at a 1:1 ratio.[37] After that the observed spectra corresponds to the TTMo spectra due to its excess in solution in relation to copper.

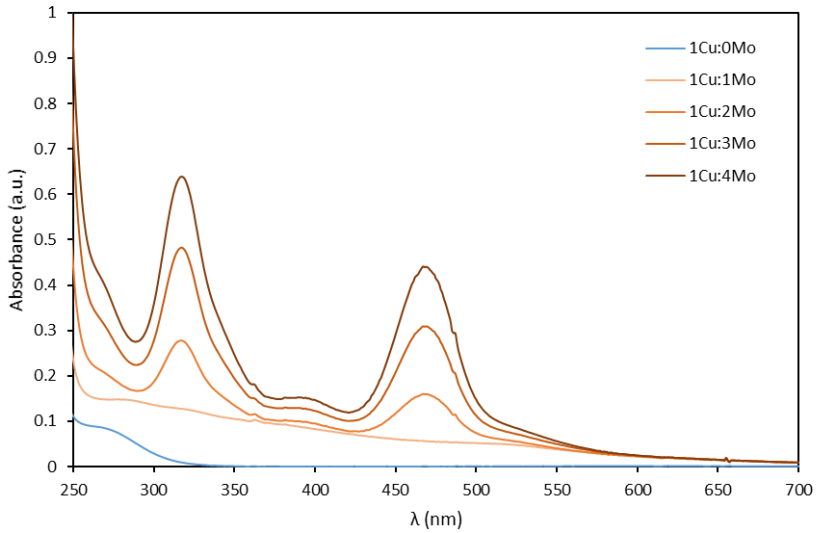

Figure 3.17 – Titration of 30 µM $CuCl_2$ with TTMo in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF. Spectra of titration at specific Cu:Mo ratios. Blue: 1Cu:0Mo; Orange: increasing Mo concentration from lighter to darker shades of orange.

In Figure 3.18 we present the spectra of the metal solutions involved in this experiment along with spectra of 1:2 metal ratio with and without DVU2108, to better analyze the differences observed. We can observe that in the absence of DVU2108, the wavelengths for maxima absorption observed for 1Cu:2Mo ratio are very similar to the TTMo solution whereas in the presence of the protein they are broader and at higher wavelengths.

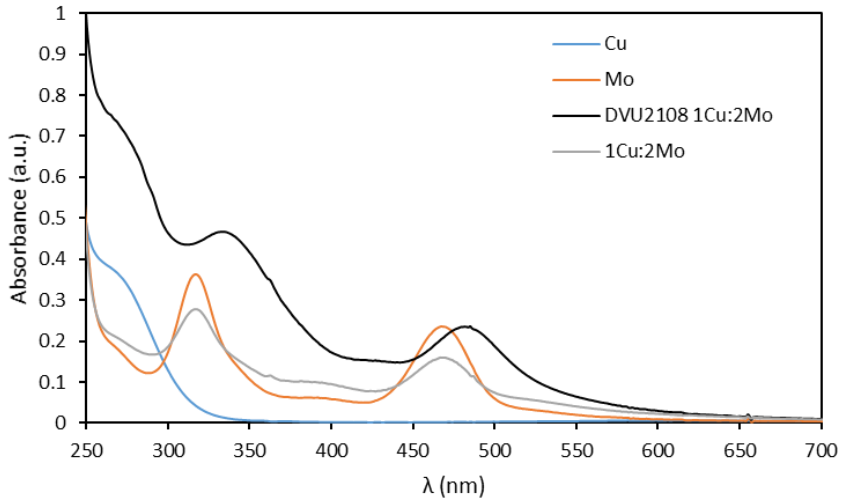

Figure 3.18 – Spectra of metal solutions, diluted 1:20, ($CuCl_2$ – Blue; TTMo – Orange) used in the titration experiments with TTMo, 1Cu:2Mo cluster proportion in the presence (black) and in the absence of DVU2108 (grey), in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF.

3.3.2 Apo-DVU2108 titration with TTW

The approach used in the TTW titration was identical to the one described before for TTMo, in which a solution containing 30 µM Apo-DVU2108 and 30 µM $CuCl_2$ in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF was titrated with small aliquots of 0.48 mM TTW and spectra was acquired after each addition. Spectra of the main Cu:W ratios are shown in Figure 3.19.

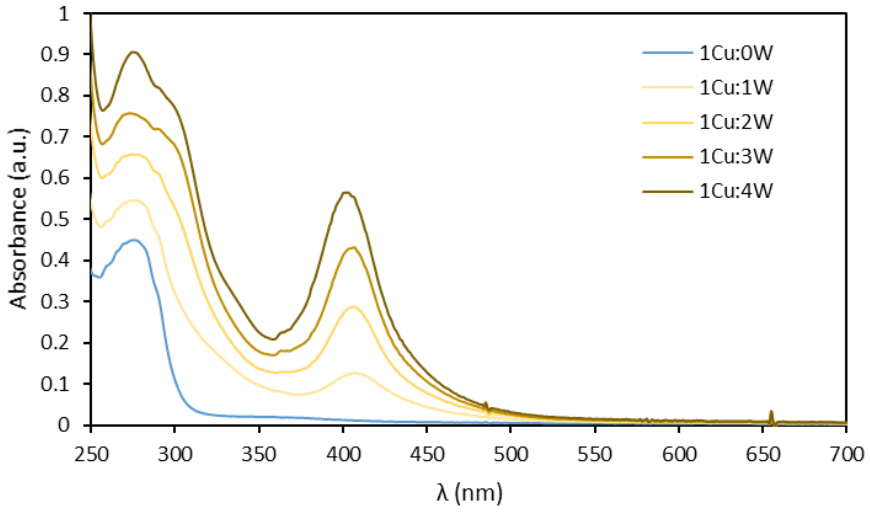

Figure 3.19 – Titration of 30 µM Apo-DVU2108, in the presence of 30 µM $CuCl_2$, with TTW in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF. Spectra of titration at specific Cu:W ratios. Blue: 1Cu:0W; Yellow: increasing TTW concentration from lighter to darker shades of yellow

Figure 3.19 shows that from 1Cu:3W ratio there is an excess of TTW, which shifts the wavelength at which absorbance reaches a maximum from 407 nm to 394 nm and there is also an increase in absorbance at 277 nm. These wavelengths are characteristic of the TTW solution (Figure 3.20).

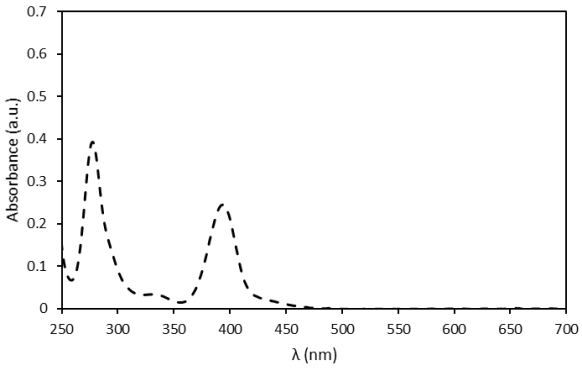

Figure 3.20 – Spectrum of 24 µM of tetrathiotungstate solution in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF).

When we analyze the absorbance values at 407 nm and subtract the contribution of excess TTW solution, according to the absorbance for 5 µL, we clearly observe a stabilization beginning at about 1Cu:2W ratio (Figure 3.21 - A). If we analyze the wavelengths around 407 nm for which there is a maximum absorbance we observe that the wavelengths shift to lower values after a 2.5 metal ratio (Figure 3.21 – B). The shifts at the maximum absorption band at around 295 nm was not analyzed since it is too close to the absorbance band of the protein (280 nm). These results show that when we replace molybdenum with tungsten the preferred cluster metal ratio in DVU2108 is maintained (1:2).

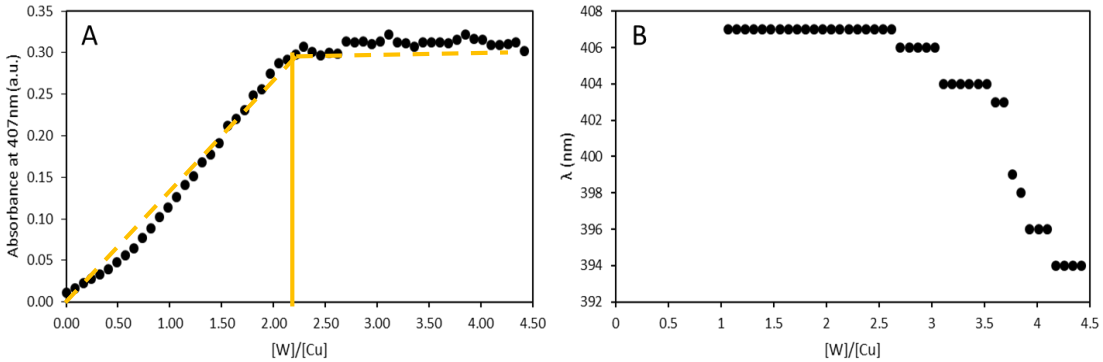

*Figure 3.21 – **A** - Variation of absorbance at 407nm at each [W]/[Cu] ratio. **B** – Wavelengths around 407 nm at which is observed maximum absorption along the titration.*

Similarly to what was observed in the TTMo titration, the spectra in the absence of protein in the TTW titration are very different from the ones in its presence (Figure 3.22). After 1:1 ratio of copper and tungsten there is an excess of TTW in solution and the observed spectra corresponds mainly to the TTW solution. In the absence of protein, until a 1:1 Cu/W ratio it is observed a complete loss of UV-visible spectrum, similar to what occurred when TTMo was used (Figure 3.17).

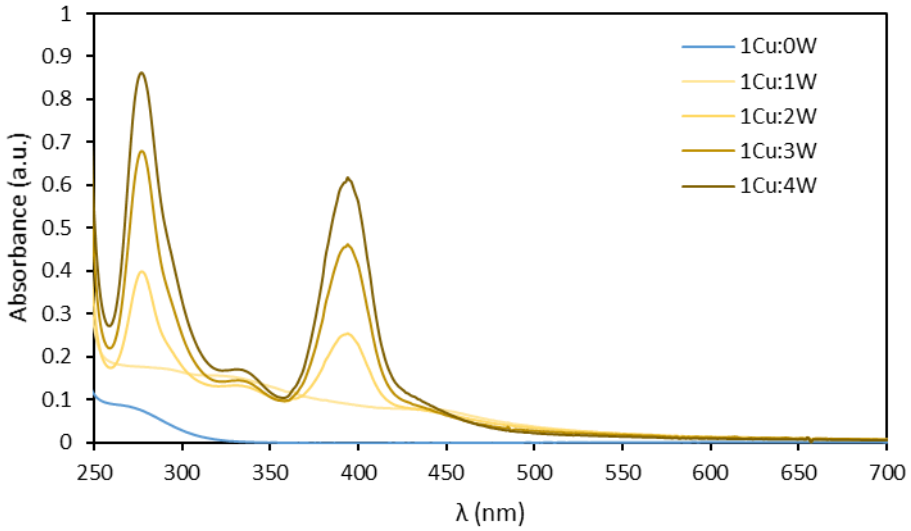

Figure 3.22 – Titration of 30 µM $CuCl_2$ with TTW in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF. Spectra of titration at specific Cu:W ratios. Blue: 1Cu:0W; Yellow: increasing TTW concentration from lighter to darker shades of yellow

In Figure 3.23, it is shown the spectra of the metal solutions used in this experiment along with spectra of 1Cu:2W ratio with and without protein. Once again we can observe that in the absence of protein the maximum absorption bands shift to lower wavelengths, namely around 407 nm, and become identical to the Cu:TTW solution spectrum at the same ratio (1:2).

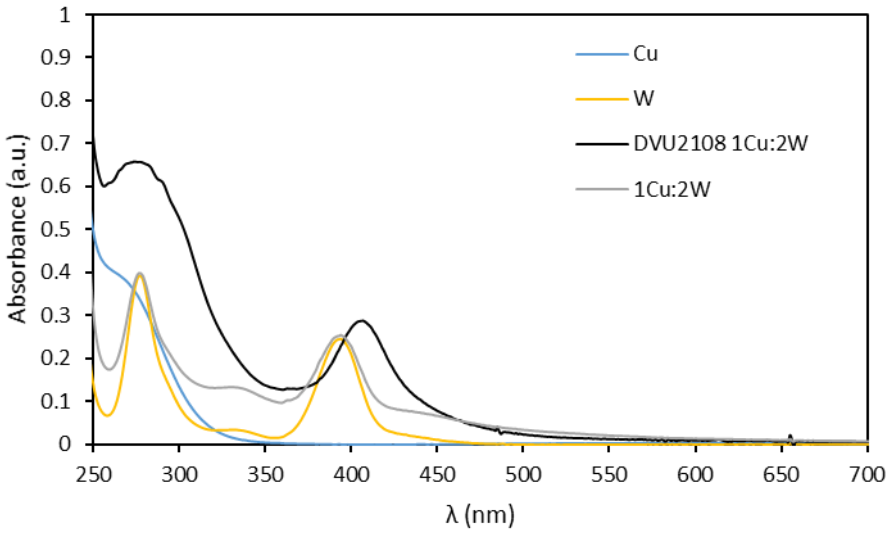

Figure 3.23 – Spectra of metal solutions, diluted 1:20, ($CuCl_2$ – Blue; TTW – Yellow) involved in the titration experiments with TTW, 1Cu:2Mo cluster proportion in the presence (black) and in the absence of DVU2108 (grey), in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF.

3.4 Differential Scanning Calorimetry

Studies made for *D. gigas* ORP showed that a W-reconstituted protein was more stable than a Mo-reconstituted in the NMR experiments.[32] To analyze this hypothesis for DVU2108 we have

used Differential Scanning Calorimetry, which as the name suggests, is a calorimetric technique that can determine the melting temperature (Tm) of the denaturation process, as well as, the energy involved in this process (ΔH).

In Figures 3.24 and 3.25, the results obtained for each protein are presented, Apo-DVU2108, MoCuDVU2108 and WCuDVU2108, respectively. Protein concentrations used in this experiment were 0.747 mg/mL for the Apo-DVU2108 and 0.656 and 0.747 mg/mL for the Mo and W-reconstituted proteins, respectively.

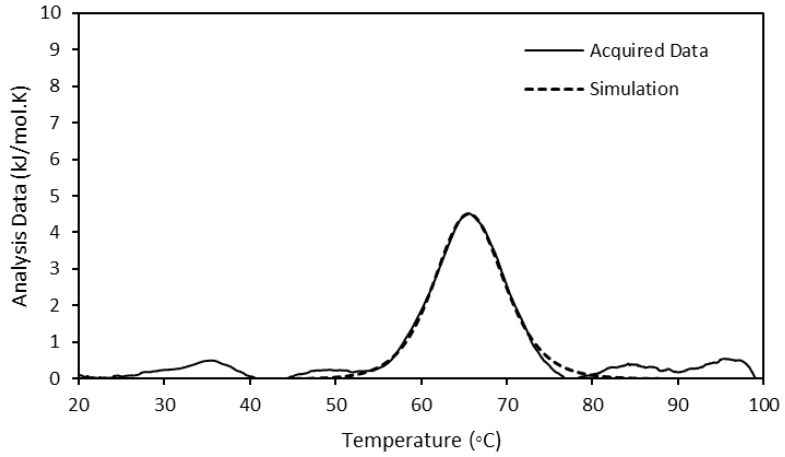

Figure 3.24 – Apo-DVU2108 (0.747 mg/mL) DSC analysis in 50 mM TrisHCl, pH 7.6, 150mM NaCl. Acquired data (full line); Simulation (dashed line)

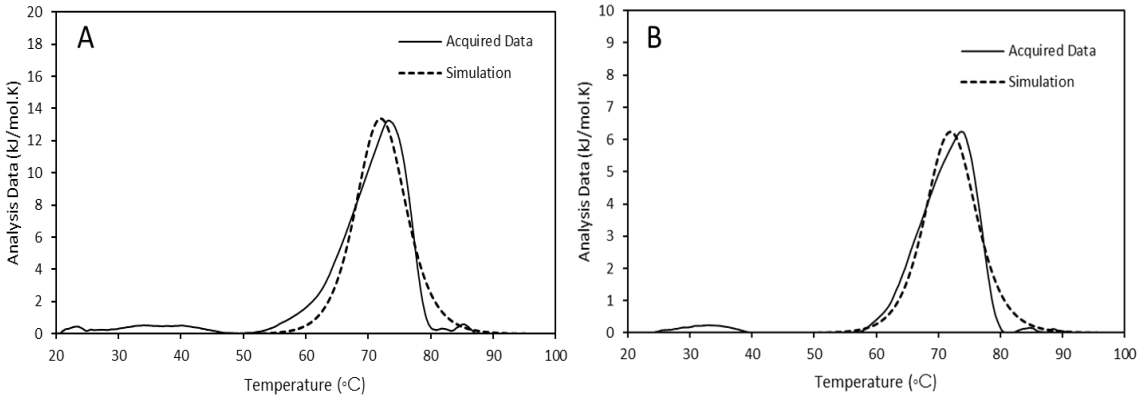

*Figure 3.25 – **A** - MoCu-DVU2108 (0.656 mg/mL) DSC analysis in 50 mM TrisHCl, pH 7.6, 150mM NaCl. Acquired data (full line); Simulation (dashed line). **B** - WCu-DVU2108 (0.747 mg/mL) DSC analysis. Acquired data (full line); Simulation (dashed line).*

Table 3.8 – DSC results for Apo-DVU2108 and reconstituted proteins (MoCu-DVU2108 and WCu-DVU2108)

Variable	Apo-DVU2108	MoCu-DVU2108	WCu-DVU2108
Tm (ºC)	65.6 ± 0.2	72.1 ± 0.6	72.1 ± 0.7
ΔH (kJ/mol)	352 ± 17	370 ± 69	363 ± 73

These results show that reconstituted DVU2108, whether with TTMo or TTW, are more stable than its Apo form since they both display higher Tm (with an increase of 6ºC). We can therefore state that not only do they endure higher temperatures without conformation loss but they also need higher energy to unfold. We can note that MoCu-DVU2108 needs a slightly higher energy to promote denaturation.

The model used for the simulation in all three proteins was the two stated scaled. This model takes in account one transition from one state to the other[38], which in this case is the native state transitioning to the denatured state.

A PAGE analysis was performed for the samples used in the DSC analysis and is shown in Figure 3.26. When we compare all the samples, we can see that for the reconstituted proteins there is still some amount of Apo-DVU2108 since there is a slight band identical to the Apo-DVU2108 band in lane 3. This analyzes indicates that the results obtained could be different from the real values for these proteins but are close since the amount of Apo-protein seems small, and metal cluster might be lost during electrophoresis

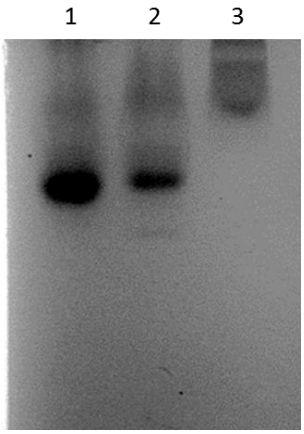

Figure 3.26 – 10% PAGE after Coomassie staining. Gel was run during 1h at 150V. (Lane 1 – MoCu-DVU2108; Lane 2 – WCu-DVU2108; Lane 3 – Apo-DVU2108)

3.5 NMR assignment of *Desulfovibrio vulgaris* Hildenborough Apo-DVU2108

NMR is a technique with multiple applications, including structure determination of proteins. In order to determine the solution structure of a protein using NMR, we first need to obtain its "fingerprint", that is, a Heteronuclear Single Quantum Correlation spectrum (HSQC), that is unique for each protein for each sample conditions.

A 2D ^{1}H-^{15}N-HSQC gives us the correlation between the proton and the nitrogen atoms in amides whether involved in the peptide bond of each protein residue, except for proline, or in the side chains of asparagine, glutamine and tryptophan. For Apo-DVU2108, and according to its amino acid sequence (Figure 3.27), we can determine how many backbone HN resonances we expect to observe in its HSQC spectrum. This protein has 119 residues with 5 of them being prolines. Since the first residue is not involved in a peptide bond it will not produce a HSQC resonance, thus we expect 113 resonances corresponding to amide protons, and 17 resonances corresponding to the side chains of 4 asparines, 4 glutamines and 1 tryptophan. Figure 3.28 shows the 2D ^{1}H-^{15}N-HSQC of Apo-DVU2108 prior to the chemical shift assignment. In this

session the procedure used for the sequential assignment of this protein will be explained step by step.

```
1   MRIVITAQGNTLDSPLDPRFGRARSFIVCDTETGATHAVDVSANMNLAQGAGIQAAQMAA
61  DAGAEAVITGHVGPKAFTALNRGHIAVYLCDLATPREALAAFIEGKLRPADTADREGHW
```

Figure 3.27 – DVU2108 119 long amino acid sequence

There is a wide range of strategies one can use to determine the structure of a protein due to the variety of NMR spectra and depending on protein labelling.[39] Here we have used a ^{13}C, ^{15}N-labelled Apo-DVU2108 and a combination of two- and three-dimensional spectra which allows the backbone resonance assignment of proton, nitrogen and carbon chemical shifts: 2D ^{1}H-^{15}N HSQC, 3D HNCO, 3D HN(CA)CO, 3D CBCANH 3D CBCA(CO)NH.[31] The program used for spectra analysis and backbone and side-chain assignment was CARA™.[40]

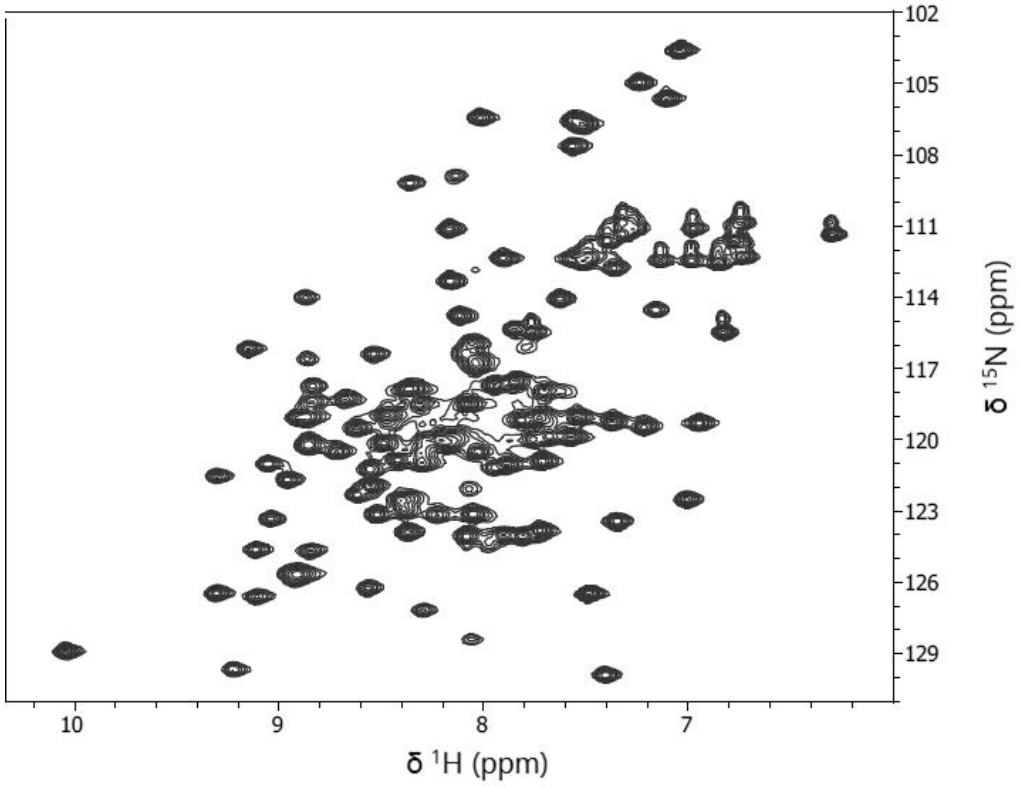

Figure 3.28 – 2D ^{1}H-^{15}N HSQC spectrum of 1 mM Apo-DVU2108 in 20 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM sodium azide, 1 mM DTT and 10 % $^{2}H_2O$. Spectrum acquisition was carried out at 298 K on a Bruker AvanceIII 600 MHz spectrometer equipped with a TCI cryoprobe

The protein sequential assignment relies on the resonance frequency of $H^N(i)$ and $N(i)$ observed in the 2D ^{1}H-^{15}N HSQC since this spectrum is the foundation of 3D spectra used for this purpose.[22]

The 3D HNCO spectra correlates the $H^N(i)$ and N(*i*) of a residue (*i*), with the carbonyl group of the previous residue (*i-1*), C(*i*-1), of a spin system (Figure 3.29). In Figure 3.30, it is shown a 5 residue long fragment, each with its respective previous carbonyl resonance, which is the one that is marked and centered (the line in the middle of each strip should pass through the center of the spherical peak). At this stage we still cannot know the residue type nor their arrangement in terms of sequence therefore more information is required.

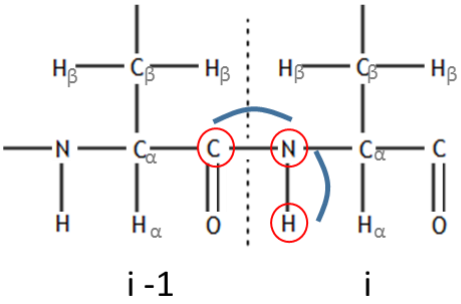

Figure 3.29 – 3D HNCO magnetization transfer

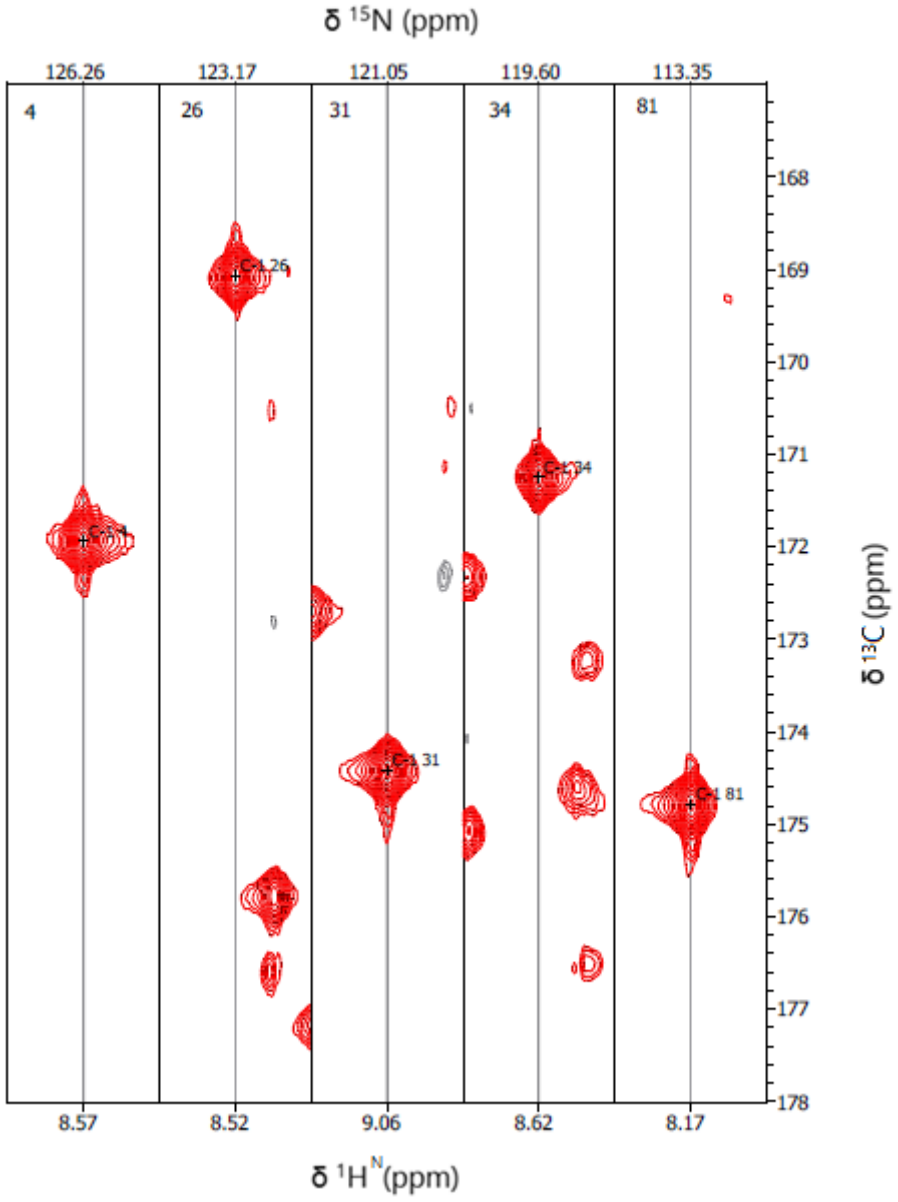

Figure 3.30 – 3D HNCO carbon dimension of spin systems 4, 26, 31, 34, 81 belonging to a sequential fragment.

To that purpose, we have analyzed a 3D HN(CA)CO spectra that correlates the $H^N(i)$ and N(*i*) with the C(*i*) of the same residue, and thus we can identify the carbonyl group of the residue corresponding to the chosen spin system (Figure 3.31). Due to the similarity in size of the *J* couplings constants used for magnetization transfer in ^{13}C-, ^{15}N-labelled proteins of C(*i*-1) and C(*i*), both resonances are detected in this kind of spectrum.[41] Figure 3.32 shows the carbon dimension of the 3D HN(CA)CO spectrum for the same fragment.

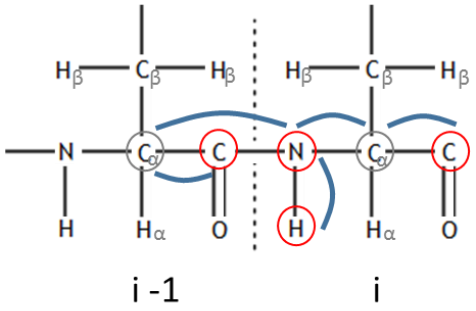

Figure 3.31 – 3D HN(CA)CO magnetization transfer

Just as a remark, the name of the experiment is formed by the nuclei involved and reveals the pathway of magnetization transfer. Spins in parenthesis correspond to nuclei through which the magnetization is transferred but there is no chemical shift evolution.[41]

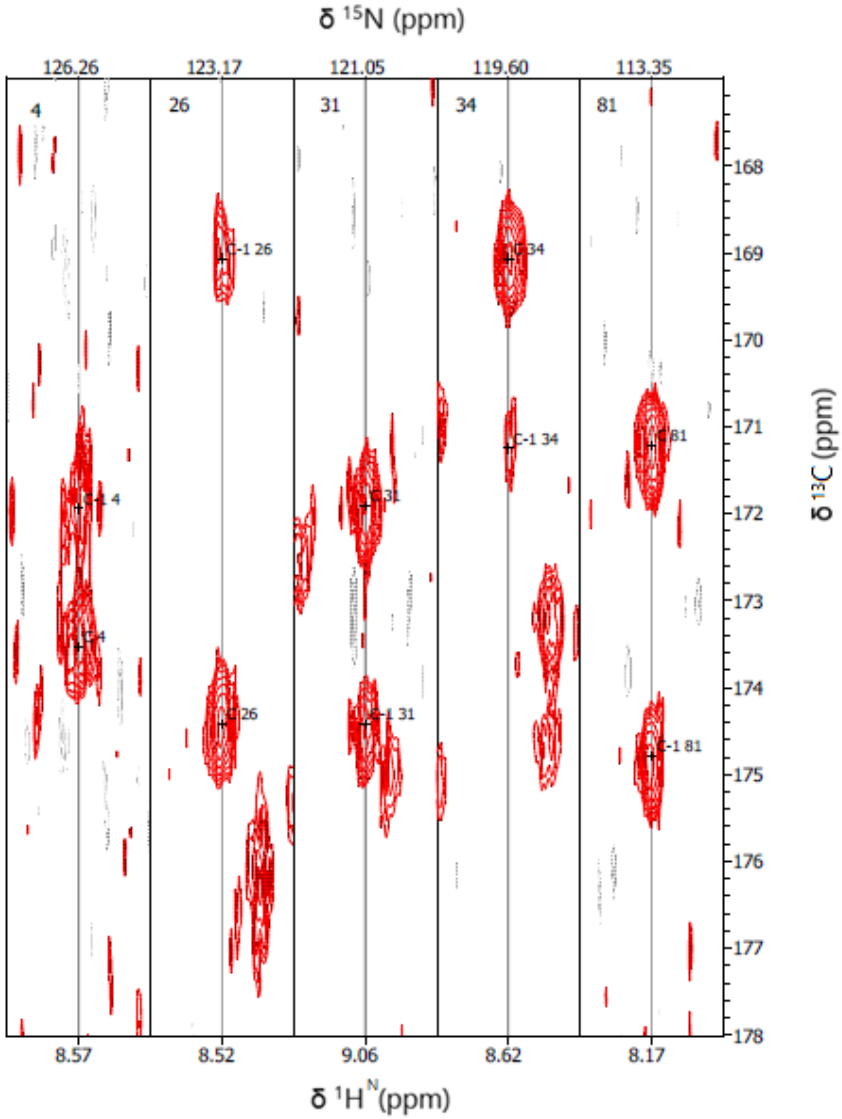

Figure 3.32 – 3D HN(CA)CO carbon dimension of spin systems 4, 26, 31, 34, 81 belonging to a sequential fragment

Taking into account that this is a five-residue fragment in a 119 residue long protein, one might think that sequential assignment would be possible just with this information, but carbonyl groups have a narrow ppm range, not differing considerably for each residue type, which would lead to uncertain alignment. Thus, once again, more data is required.

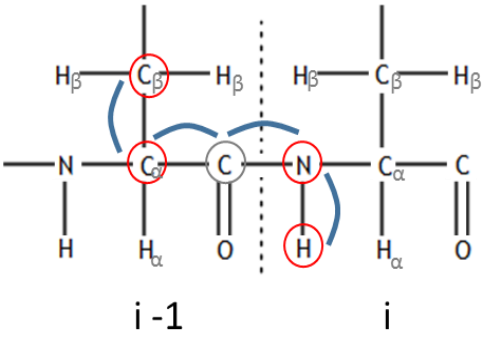

Figure 3.33 – CBCA(CO)NH magnetization transfer

Apart from glycine all amino acids have both alpha (CA) and beta (CB) carbon atoms in their structure. Chemical shift information from these atoms can help to overcome the ambiguity of the previous experiments since they have a much wider ppm range when compared to carbonyl groups, and differ significantly for each residue type. In addition, the chemical shift values of CA and CB vary with secondary structure.

For instance, serine and threonine have the CB resonance at lower field than the CA unlike all other residue types, a distinctive property that makes the identification of these residues simpler.

A 3D CBCA(CO)NH experiment allows us to obtain for each spin system with a detectable H^N, the chemical shifts of CA(*i*-1) and CB(*i*-1) of the previous residue. (Figure 3.33).

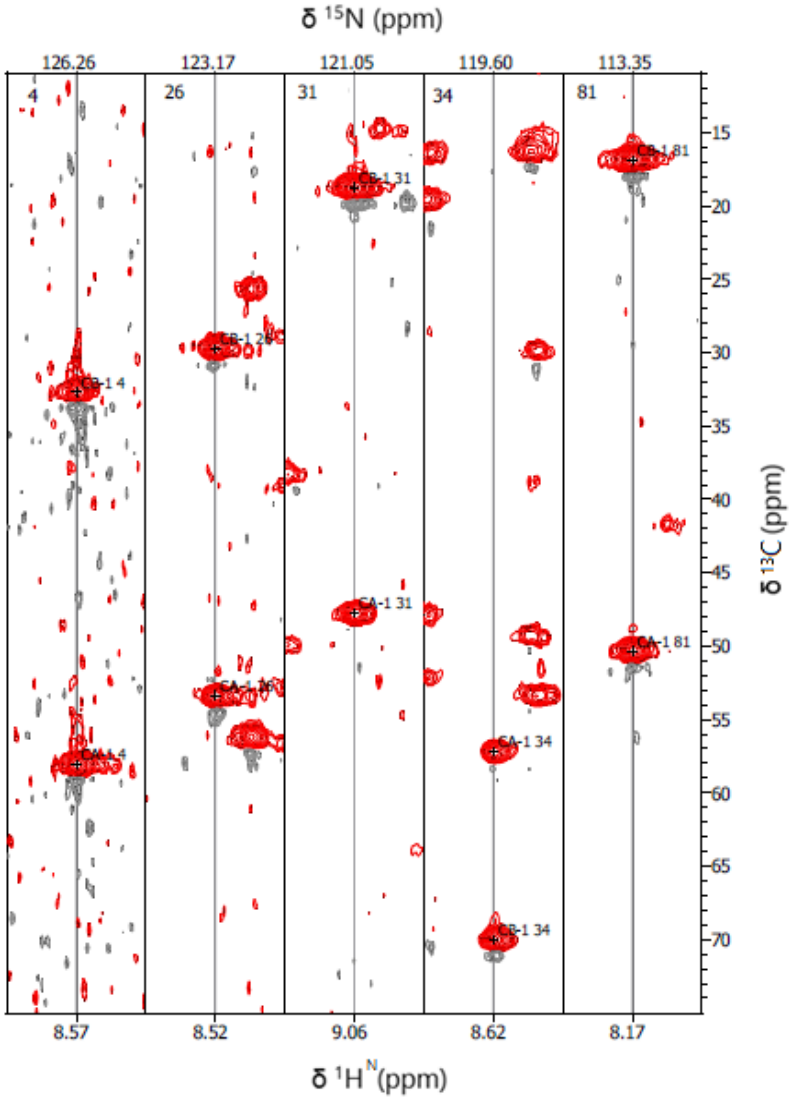

Figure 3.34 – 3D CBCA(CO)NH carbon dimension of spin systems 4, 26, 31, 34, 81 belonging to a sequential fragment

Similar to the strategy used for the carbonyl atom, we have the chemical shift of both CA and CB of the previous residue, so now we need the chemical shifts of the same atoms but of the residue itself. This information can be obtained with a 3D CBCANH spectra. In this spectra, the CB resonance is a negative peak whereas CA resonance is positive.

Just as the 3D HN(CA)CO detected the previous carbonyl group in addition to the one of itself, this spectrum will detect the CA(*i*-1) and CB(*i*-1) in addition to CA(*i*) and CB(*i*) (Figure 3.35).

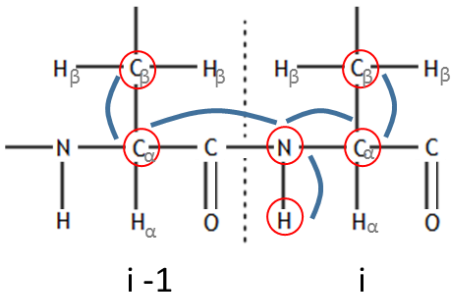

Figure 3.35 – 3D CBCANH magnetization transfer

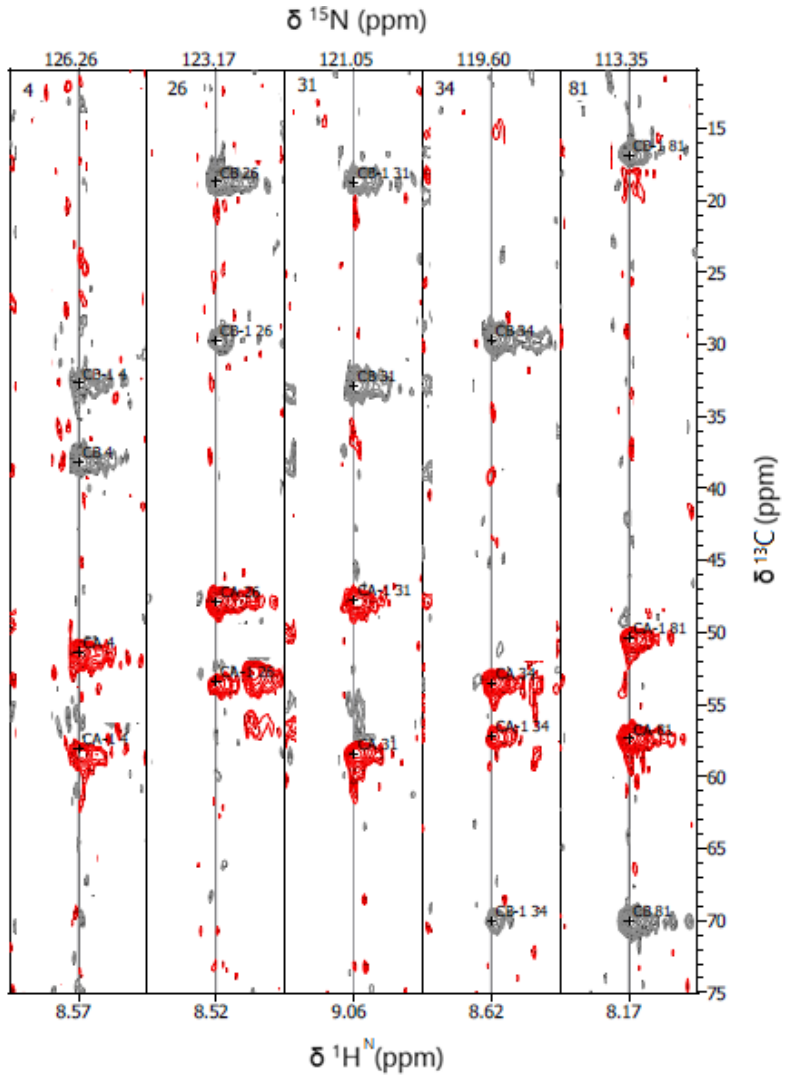

Figure 3.36 – 3D CBCANH carbon dimension of spin systems 4, 26, 31, 34, 81 belonging to a sequential fragment

At this point we have, for each backbone spin system detected in the 2D ^{1}H-^{15}N HSQC, the chemical shifts of the following atoms: $H^{N}(i)$, N(i), CO(i-1), CO(i), CA(i-1), CB(i-1), CA(i) and CB(i). So now we have the information needed to start the sequential assignment. Figure 3.37 shows the alignment of the five-residue long fragment that now can be positioned in the sequential order 81-34-26-31-4.

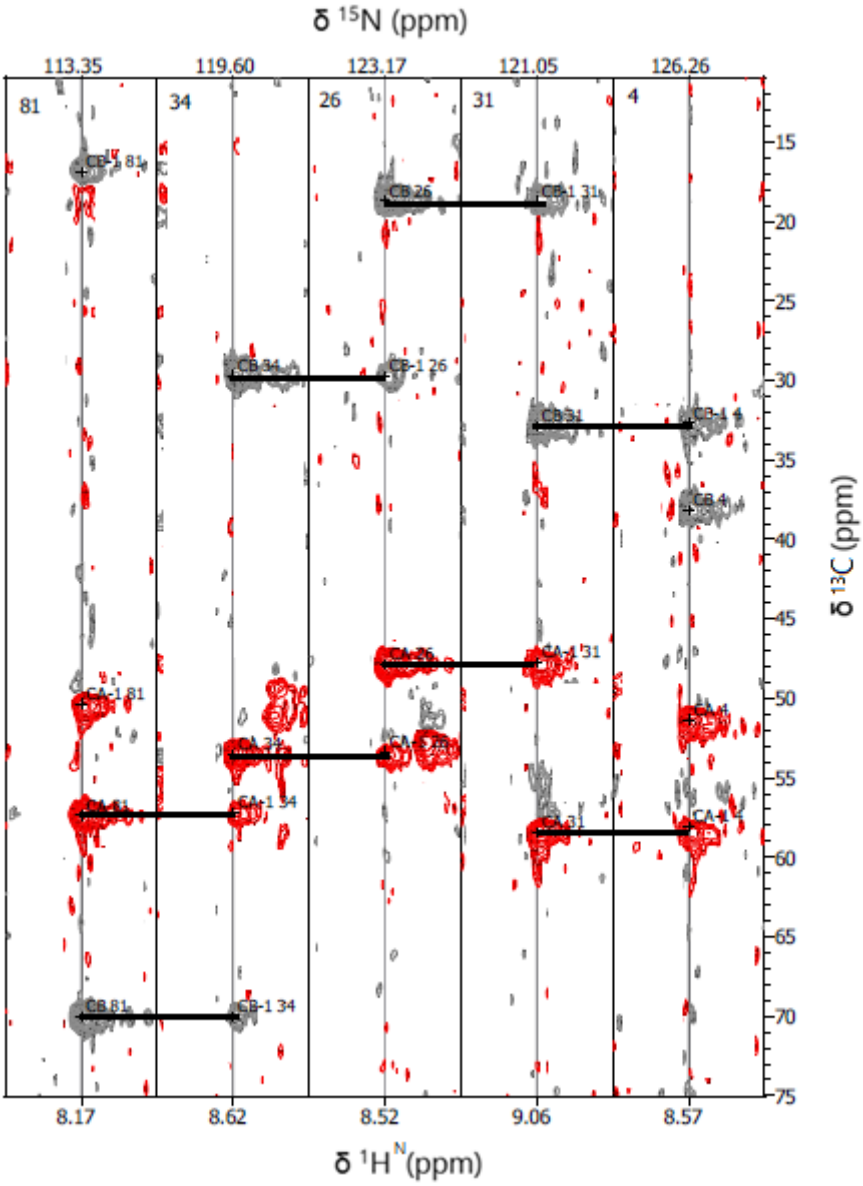

Figure 3.37 – 3D CBCANH spectra of spin systems 4, 26, 31, 34, 81 belonging to a sequential fragment already aligned

Now that the sequential alignment is done, we need to assign each spin system to its residue type. CARA™ matches the chemical shifts obtained of the assigned atoms (C, CA and CB) to the residue database in order to predict the amino acid type. Fragment assign is ranked in relation to the matching quality by green bars and fitness percentage. Larger fragments will give better predictions (Figure 3.38).

(#81)	#81	#34	#26	#31	#4	(#4)	
Sequence/Fragment							Fitness
A35	T36	H37	A38	V39	D40		74%
A98	L99	A100	A101	F102	I103		48%
A56	Q57	M58	A59	A60	D61		47%
L80	N81	R82	G83	H84	I85		42%
	M1	R2	I3	V4	I5		42%
G83	H84	I85	A86	V87	Y88		40%

Figure 3.38 – Assignment predictions

For the chosen fragment the assignment prediction given by CARA™ seems accurate. And thus the fragment 81-34-26-31-4, can be assigned to T36-H37-A38-V39-D40 of Apo-DVU2108. The assigned fragment is shown in Figure 3.39.

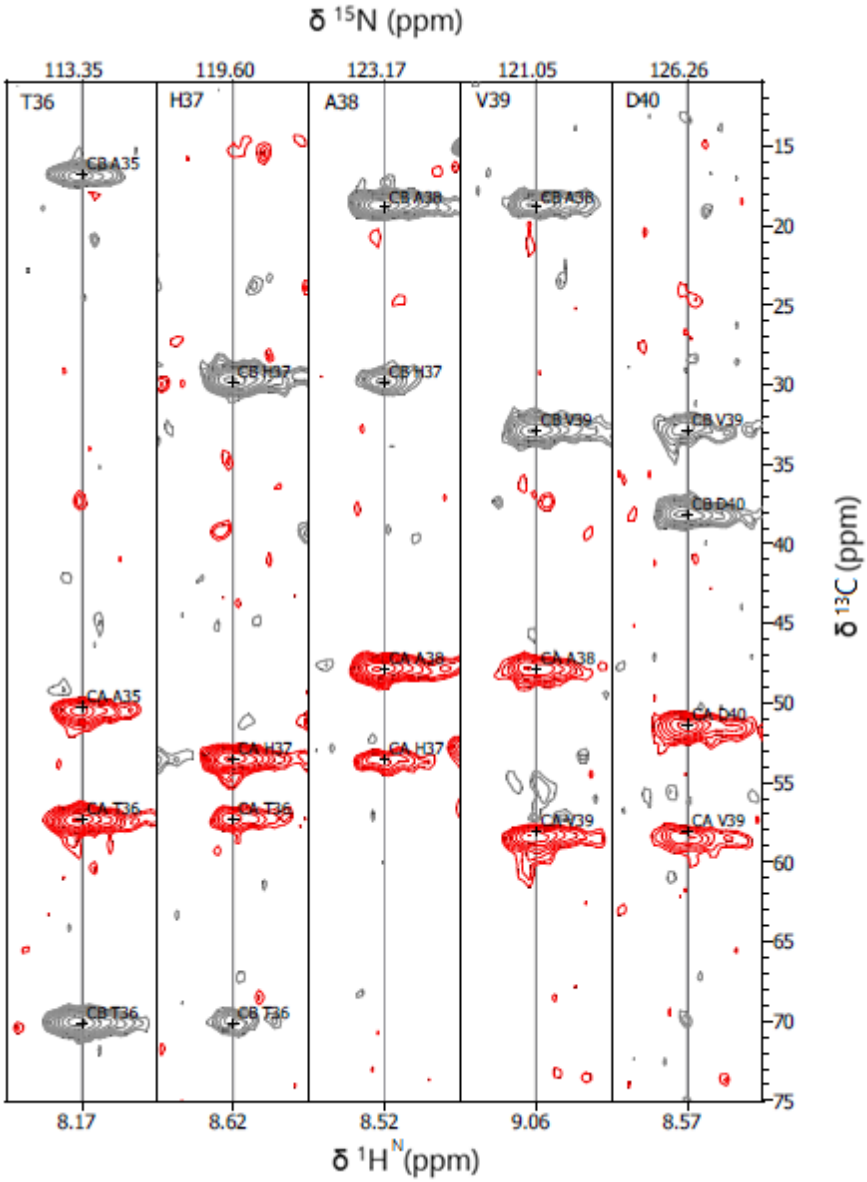

Figure 3.39 – 3D CBCANH carbon dimension of the fragment composed by T36, H37, A38, V39, D40.

Even with all this information sometimes CARA™ cannot correctly assign the spin systems to the right residue, as there can still exist some ambiguities. In order to overcome this issue we can acquire and analyze other spectra, such as 3D HNHA, 3D HBHA(CBCACO)NH, 3D (H)C(CCO)NH, 3D H(CCCO)NH and 3D (H)CCH-TOCSY[31]. The combination of the 3D HNHA and 3D (H)CCH-TOCSY spectra provides additional information on the side chain residues which may solve some ambiguities that could exist. The complete assignment of protons and carbons of the aliphatic side chains can be assigned through the analysis of the 3D (H)CCH-TOCSY spectra. For that, we first used the 3D HNHA spectra which will provides us with the chemical shift of the alpha proton (HA(*i*) linked to the alpha carbon) (Figure 3.40) of each residue.

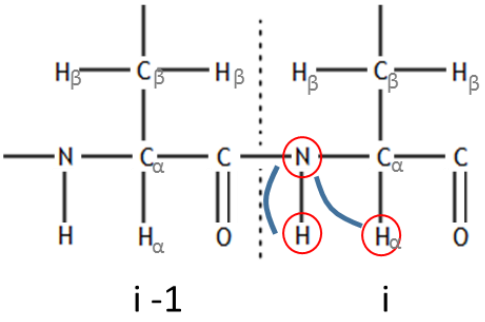

Figure 3.40 – 3D HNHA magnetization transfer

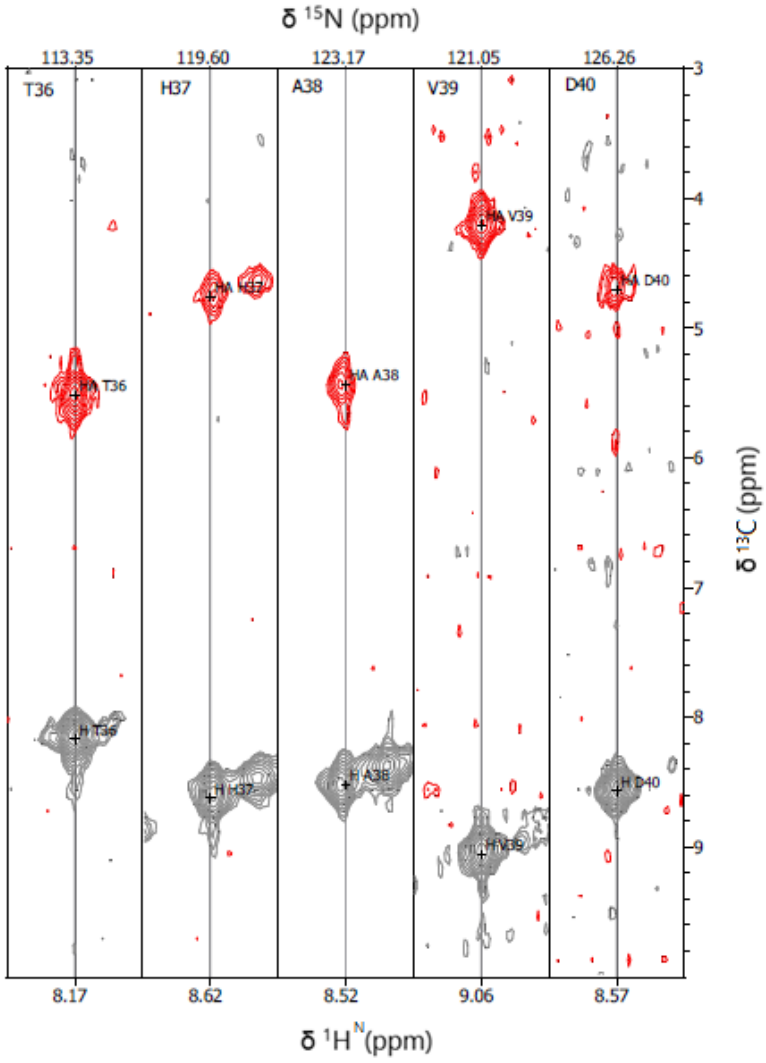

Figure 3.41 – 3D HNHA carbon dimension of the fragment composed by T36, H37, A38, V39, D40

With a 3D (H)CCH-TOCSY spectrum we can link the CA(*i*) and HA(*i*) chemical shifts that we already have, to the rest of the side chain nuclei. Using this strategy, it is possible to better identify the residue type of each spin system improving the assignment prediction or confirm it, and obtain the chemical shift of the side chain protons that is essential for structure determination. Figure 3.42 shows the 3D (H)CCH-TOCSY spectrum for Valine 39 whose structure is shown in Figure 3.43.

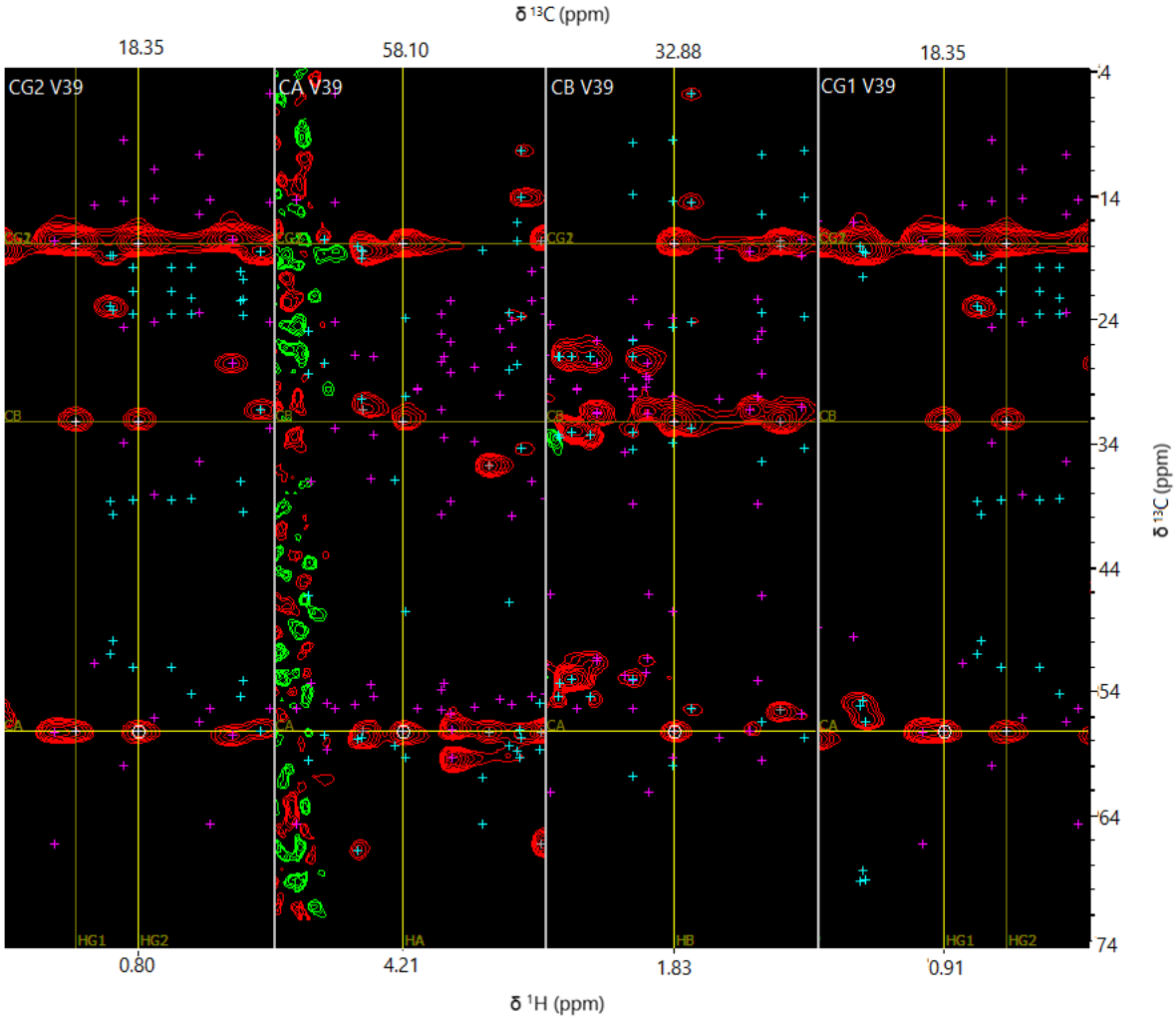

Figure 3.42 – 3D (H)CCH-TOCSY spectra of Valine 39

Since the 3D (H)CCH-TOCSY detects the carbon and proton for each aliphatic residue (Figure 3.44), all protons have correlation peaks with all the carbons of that specific residue.

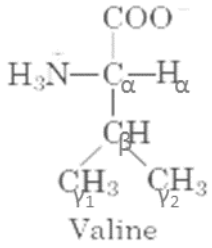

Figure 3.43 – Valine structure. Adapted from [36]

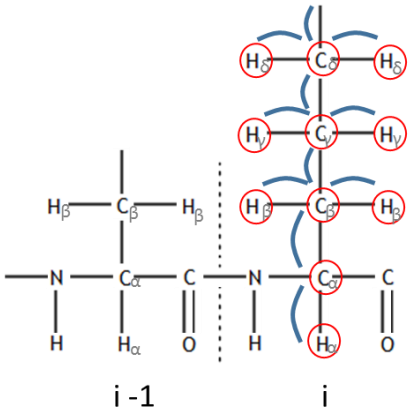

Figure 3.44 – 3D (H)CCH-TOCSY magnetization transfer

After doing this for every spin system we will hopefully have assigned all the residues, except for the first residue and prolines, with the final result being a 2D ^{1}H-^{15}N HSQC such as the one presented in Figure 3.45.

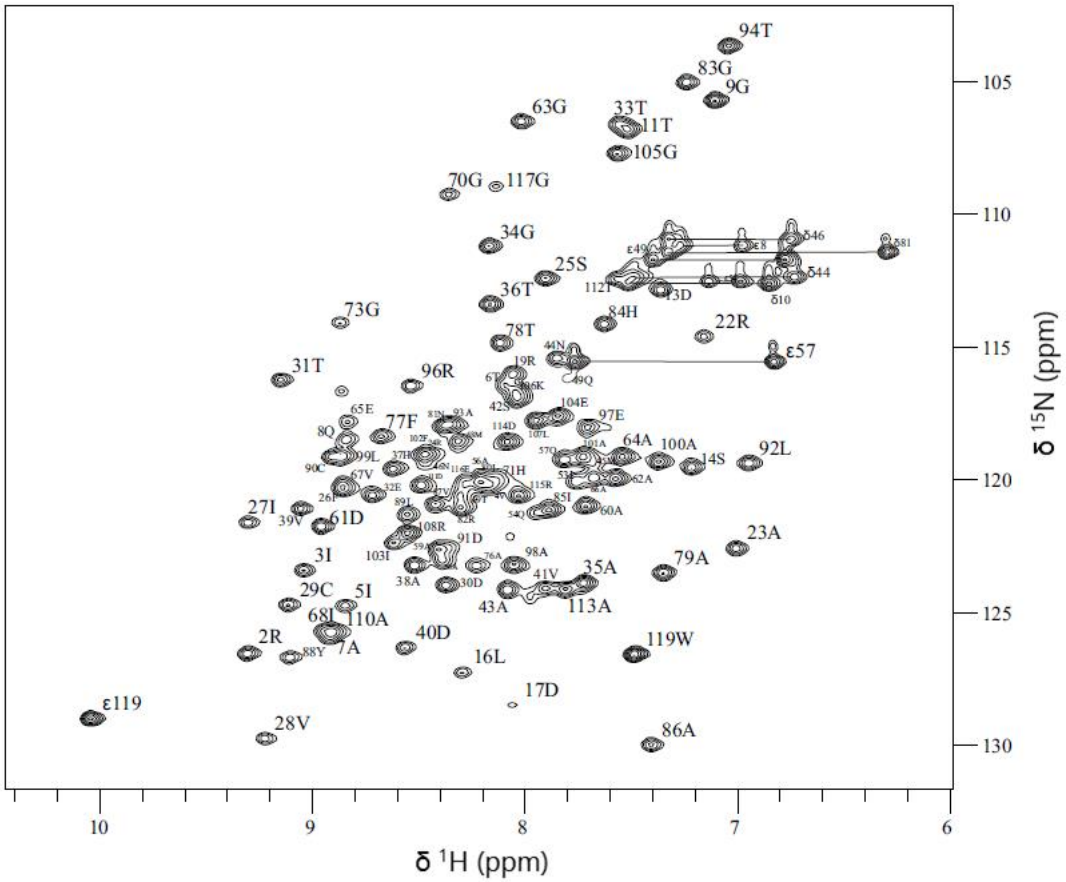

Figure 3.45 – 2D ^{1}H-^{15}N HSQC spectrum of 1 mM Apo-DVU2108 in 20 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM sodium azide, 1 mM DTT and 10 % $^{2}H_2O$. Spectrum acquisition was carried out at 298 K on Bruker AvanceIII 600 MHz spectrometer equipped with a TCI cryoprobe

Aside from the amides that have no HSQC signal, some can undergo exchange with solution buffer and therefore are not detected and cannot be assign. However, we can assign their other atoms without the $^{1}H/^{15}N$ resonance if we have the assignment made for their succeeding residue, as for these residues we have the *i-1* chemical shifts of C, CA, CB, HA, which will allow us to assign them to their respective residue.

In order to perform structure calculations we need all the chemical shifts we can obtain. Up to now, we have the backbone and aliphatic side chains, only lacking the aromatic side chains of histidines, phenylalanines, tyrosines and tryptophan and the amides of asparagines, glutamines and the NH of the tryptophan ring.
Tryptophan side chain amide has a distinctive resonance usually around 10 ppm at the ^{1}H dimension. The remaining amides have a signal in the HSQC spectrum and can be assign if we already have all the other atoms assign for those residues, especially the CB(*i*) and CG(*i*) for glutamines and the CA(*i*) and CB(*i*) for asparagines. Since this amides are also linked to a carbonyl group we can use the 3D HNCO and 3D CBCA(CO)NH like before but this time the chemical shifts will correspond to the side chain carbons (Figure 3.46), the CB, CG and CD for glutamine and CA, CB and CG for asparagine.

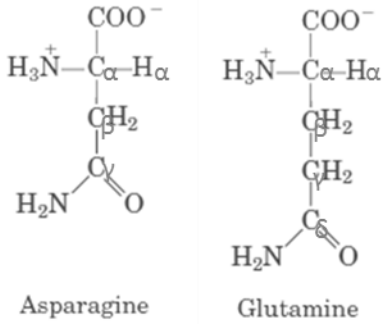

Figure 3.46 – Asparagine and glutamine structure. Adapted from [36]

Both asparagine and glutamine residues have two protons linked to the side chain nitrogen atom, thus there will be two resonances with the same shift in the nitrogen dimension corresponding to the two protons. The 3D CBCA(CO)NH spectrum will allow us to assign these signals to the correct residue, using the side chain carbons we already have assigned. With that we will also be able to assign the CD(*i*) of glutamine and the CG(*i*) of asparagine since the chemical shift in 3D HNCO spectrum will correspond to those carbon atoms.

Due to the resonance of the imidazole ring in histidines (Figure 3.47), their side chain amides will have different chemical shifts when compared to the other amides. We can detect them through 2D ^{1}H-^{15}N HSQC type spectrum with an INEPT constant optimized for ^{2}J detection.[42,43] Figure 3.48 shows the pattern obtained for His84.

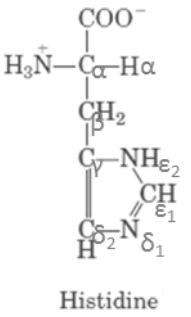

Figure 3.47 – Histidine structure. Adapted from [36]

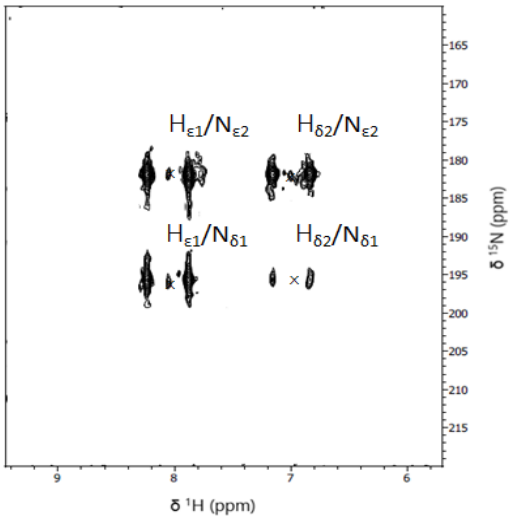

Figure 3.48 – His84 amide pattern

In order to assign the proton and carbon atoms of aromatic side chains, we have used a combination of a 2D ^{1}H-^{13}C HSQC with a 3D ^{13}C-edited NOESY spectra. Unlike all the other spectra used previously the NOESY is a spatial-type spectrum. It relies on the proximity (up to 7 Å) of atoms regardless of its involvement in a chemical bond. The more backbone and side chain atoms assigned of each residue the better we can assign the aromatic atoms resonances.

To assign these resonances, we have to examine the ^{13}C-edited NOESY of each peak and correlate the observed resonances with all the chemical shifts we already have from the aromatic side chain containing residues. If the resonances match it means they are from the same system. Keep in mind that the NOESY spectrum will have resonances other than the ones from that particular residue corresponding to neighbor (but not just sequentially close) residues. Figure 3.49 - A shows almost all the resonances already assign and Figure 3.49 - B is the ^{13}C-edited NOESY spectrum from the HD_2/CD_2 HSQC peak of Tyr88 whose structure is shown in Figure 3.49 - C.

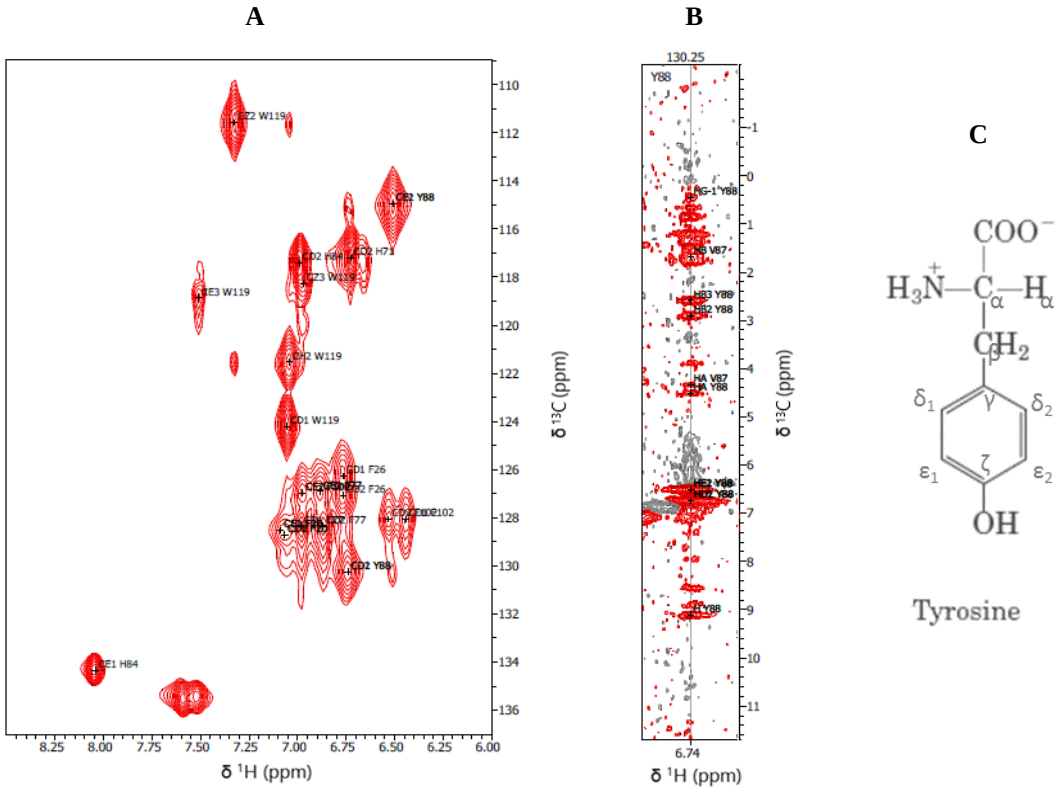

*Figure 3.49 – **A** - 2D ^{1}H-^{13}C HSQC spectra close up in the aromatic region assigned. **B** - ^{13}C-edited NOESY of HD_2/CD_2 resonance from Tyr88. **C** - Tyrosine structure. Adapted from* [36]

At the end of the resonance assignment that was described above, we were able to assign only 101 of the 113 expected backbone NH resonances probably due to exchanges with solution buffer, similar to what occurred for the Apo-form of the homologous protein from *D. gigas.*[6] From the complete sequential assignment it was possible to locate residues that had no NH resonance in four regions of the protein: 4 residues at the N-terminal (Asn10, Leu12, Phe20, Gly21), 5 residues in a loop region that compromises the residues 47-52, two residues Val72-Lys75 and His118. The amide proton resonances of the residues of a loop region were poorly defined, which is located at the same region in other homologous proteins[6,8]. The assignment was made for 4 of the 5 existent prolines and the first residue yielding an assignment percentage of 97.4 % in the terms of residues (116 of 119) and 94% in terms of ^{1}H, ^{15}N and ^{13}C detectable in NMR. For the remaining residues, G50, A51 and P74, no atom could be assigned. The ^{1}H, ^{13}C and ^{15}N chemical shifts have been deposit in the BioMagResBank (http://www.bmrm.wisc.edu) under BMRB accession number 26618. Just as a note, this assignment was further improved and is the one that is deposited in the BMRB. This data will be published in Biomolecular NMR Assignments with the title: *Resonance assignment of DVU2108 that is part of the Orange Protein complex in Desulfovibrio vulgaris Hildenborough* (SI 7.7) (the manuscript is *in Press* since 15 Sept 2015)

Table 3.9 shows the percentage of assignment achieved for each atom and for protein backbone and side chain resonances. A list of all the atoms from all residues assigned is shown in Supporting Information 7.6.

Table 3.9 – Assignment Report for Apo-DVU2108

Shifts / Atom	Total	Found	Missing	%
H	635	600	35	94.5
C	502	483	19	96.2
N	132	114	18	86.4
Backbone	601	558	43	92.9
Side Chain	664	639	25	96.2

3.6 Conformation of Pro, Cys and His in Apo-DVU2108

With the chemical shifts for histidine, proline and cysteine residues we can have structural information of their side chains such as conformational, protonation and redox state.

The pattern of histidine side chain amide resonances allow us to determine their tautomeric state depending on the protonation of the amides in the imidazole ring the resonance will create a different pattern (Figure 3.50). The pattern obtained for His84 (Figure 3.48) shows that the four resonances are between 150 and 200 ppm which is indicative of a fully protonated histidine at the sample pH, which was neutral (7.0).[43]

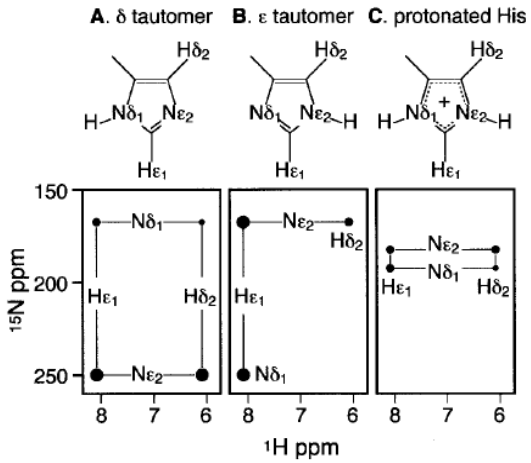

Figure 3.50 – Schematic diagram adapted from[42] showing the expected ^{1}H-^{15}N-HMQC spectrum of the imidazole-ring for each of the three possible protonation states of a histidine residue. Reprinted from [43]

Proline has a distinctive structure since its nitrogen atom involved in the peptide bond is part of a five-side ring. This characteristic creates a cis-trans isomerization in proline according to the positioning of its alpha-carbon relative to the alpha-carbon on the previous residue (Figure 3.51). In native proteins these amino acids exist either in their cis or trans conformation, and as an equilibrium in unfolded proteins.[44]

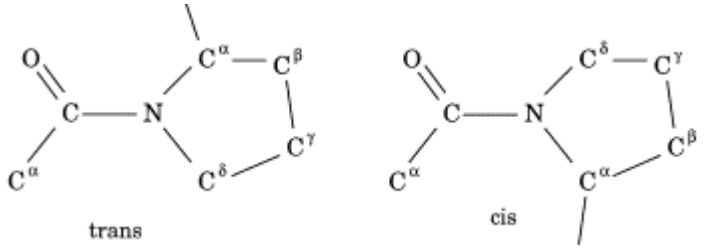

Figure 3.51 – Proline fragments in trans and cis conformations. Reprinted from [44]

The conformation of proline residues affects the chemical shift of the carbon atoms present in the ring since the environment around them will be different. In the *cis* conformation the gama-carbon will be at a higher field and the beta-carbon at a lower field when compared to the *trans* isomer (SI 7.8). Therefore, the chemical shift difference between those two carbon atoms (ΔCβγ) in a proline residue can be an indication of its isomer. The *trans* isomer has an average ΔCβγ of 4.5 ppm, whereas 9.6 ppm is the average value for *cis* conformation.[45]

In Apo-DVU2108 the assignment was made for four of the five existent prolines: Pro15, Pro18, Pro95, Pro109. Table 3.10 shows the ΔCβγ obtained for each one of these residues.

Table 3.10 – Chemical shift different between the CB and CG of assigned proline residues

	Pro15	Pro18	Pro95	Pro109
ΔCβγ (ppm)	3.499	6.402	4.416	4.626

As you can see prolines 15, 95 and 109 are clearly *trans* isomers whereas Pro18 arises some uncertainties. However, since this is a folded protein, the existing residues should be in the same conformation and thus Pro18 is most probably also in a *trans* conformation, but might be in a more dynamic region (backbone could experience different conformations).

In the cysteine residues, the reduced state (free) and the oxidized one (disulfide bridge) can be distinguished after its assignment using the beta-carbon chemical shift, which are very sensitive to the redox state of the sulfur (SI 7.9). Thus, we can point out their involvement or not in a disulfide bridge before we have a full structural model. From 24 to 32 ppm the cysteine is reduced (-SH) and from 34 to 52 ppm it is involved in a disulfide bridge (-S-S-), whereas between those areas it could be either.[46] DVU2108 is the only protein in the Orange Protein family to have two cysteine residues in its sequence, as point out in the Introduction. These two cysteine residues where assigned and their beta carbon shifts was found to be 27.53 ppm and 28.66 ppm for C29 and C90, respectively. As we can see, by comparing these values the two residues are clearly

not involved in a disulfide bridge since the beta-carbon chemical shift is below 32 ppm, which might also be expected as DTT was present in the protein NMR sample buffer.

3.7 Determination of the secondary structure of Apo-DVU2108

The chemical shifts values are dependent of the type of residue and their surrounding environment and, with the help of bioinformatic tools we can predict the protein secondary structure. The program Torsion Angle Likeliness Obtained from Shift and Sequence Similarity, or simply TALOS+, is a program that relates ^{1}H, ^{15}N and ^{13}C chemical shifts with backbone torsion angles φ and ψ. It also takes into account the residue type and its neighbors in the sequence.[47] Figures 3.52 and 3.53 show the results obtained for Apo-DVU2108 using this program. In figure 3.52 the size of each bar represents the probability of a residue to be in a beta-sheet (blue bars) or an alpha helix (red bars). It also shows the random coil index along with secondary chemical

shifts (RCI-S^2). The first corresponds to residue flexibility and the second is the difference between the observed chemical shifts of a particular residue and its random coil value.[48]

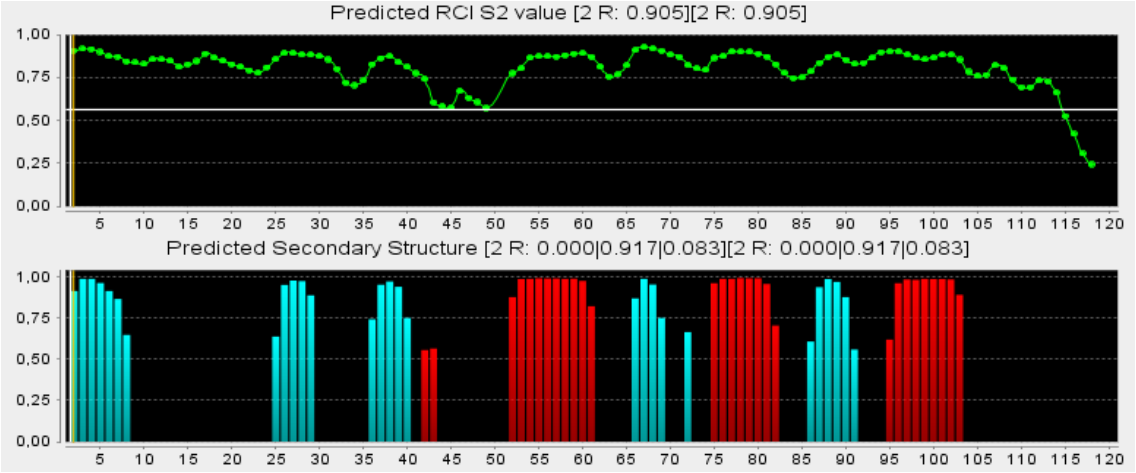

Figure 3.52 – RCI-S^2 value for Apo-DVU2108 is shown as a function of residue number on top and the predicted secondary structure (red, α-helix; blue, β-sheet) on the bottom.

In Figure 3.53 the green represents a consistent prediction, yellow an ambiguous one and in blue the dynamic residues. In gray are the ones that have no classification. The program also generates Ramachandran maps for each residue.

M1	R2	I3	V4	I5	T6	A7	Q8	G9	N10
T11	L12	D13	S14	P15	L16	D17	P18	R19	F20
G21	R22	A23	R24	S25	F26	I27	V28	C29	D30
T31	E32	T33	G34	A35	T36	H37	A38	V39	D40
V41	S42	A43	N44	M45	N46	L47	A48	Q49	G50
A51	G52	I53	Q54	A55	A56	Q57	M58	A59	A60
D61	A62	G63	A64	E65	A66	V67	I68	T69	G70
H71	V72	G73	P74	K75	A76	F77	T78	A79	L80
N81	R82	G83	H84	I85	A86	V87	Y88	L89	C90
D91	L92	A93	T94	P95	R96	E97	A98	L99	A100
A101	F102	I103	E104	G105	K106	L107	R108	P109	A110
D111	T112	A113	D114	R115	E116	G117	H118	W119	

Figure 3.53 – Sequence of Apo-DVU2108 with residues for which no prediction is obtained marked in light grey, consistent predictions in green, ambiguous predictions in yellow, and dynamic residues in blue.

The C-terminal region (R115-H118) in Apo-DVU2108 as shown to be dynamic which is usually a relatively dynamic region in most proteins. There seems to be another dynamic region in this protein with three residues displaying that behavior, N44, M45 and Q49. This region could be interacting with the metal cluster in the native protein has shown in the studies with *D. gigas* ORP

where possible regions for interaction with the metal cluster were determined[32]. Nonetheless, these are mere hypothesis and further investigation is necessary.

Besides TALOS+ there are other programs that can predict a protein secondary structure with chemical shift information. Collaborative Computing Project for NMR, or CCPN, is a program for NMR data analysis and deposition. Among other aspects this program can create a secondary structure prediction using the Chemical Shift Index (CSI). CSI method takes in account the chemical shift difference between the acquired shift data and random coil shifts, which depending on the atom, is directly related to secondary structure. For example, when the HA chemical shift is upfield shifted regarding the random coil shift it is in a helical conformation and in a beta-sheet conformation when downfield shifted. Figure 3.54 shows the CSI for some of the assigned atoms and the corresponding secondary structure.[49]

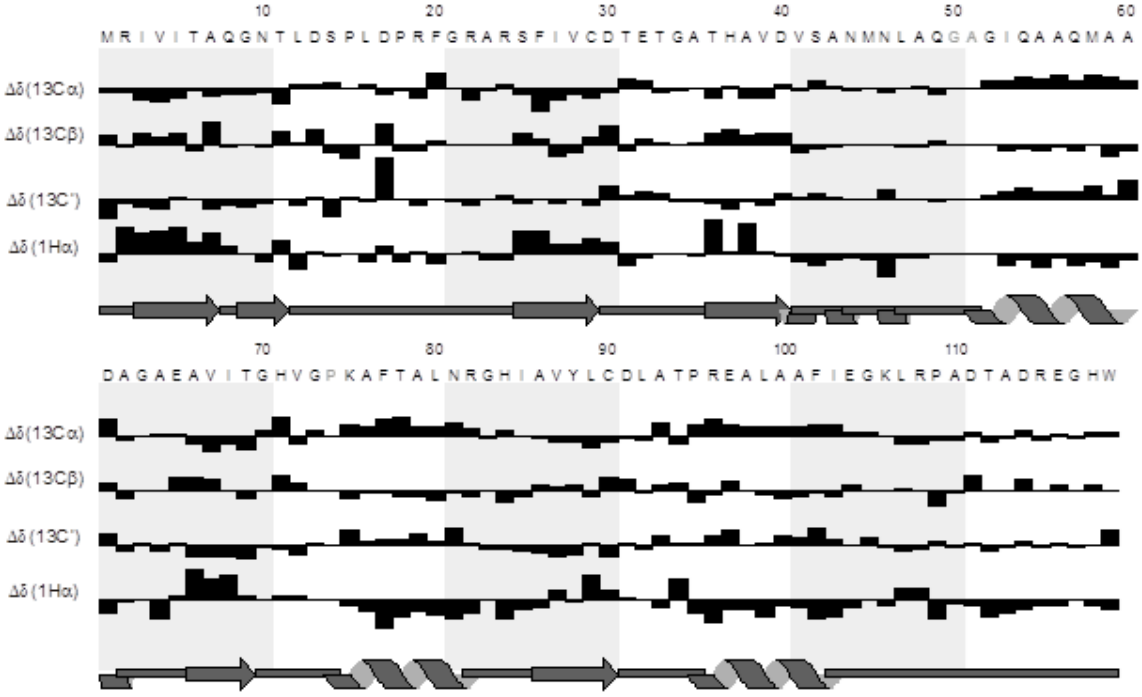

Figure 3.54 – Prediction of secondary structure of Apo-DVU2108 based on chemical shifts of ^{13}CA, ^{13}CB, ^{13}CO and ^{1}HA. Figure was prepared in CCPN Analysis.

This prediction is in correlation with the previous one made by TALOS+ where it shows three main alpha helix regions and five beta-sheet areas.

3.8 Structural Model of Apo-DVU2108

A structural model of Apo-DVU2108 (Figure 3.55) was created simply by using its primary sequence and the deposited structure of *D. gigas* Apo-ORP, that has 48% homology with DVU2108, obtained by X-ray crystallography[10] (PDB code: 2WFB), in the Swiss Model tool via Expasy web server.

Figure 3.55 – Structural model for Apo-DVU2108 from D. vulgaris Hildenborough. Picture generated by Chimera™

The obtained model is very similar to *D*. *gigas* ORP, as expected, and shows five beta-sheet and three main alpha-helixes and a smaller one, which supports the secondary structure predictions made in the previous chapter. Analysis of this structure also shows that the two cysteine residues are not in close proximity to make a disulfide bond with each other (data not shown), confirming the previous NMR results concerning this two residues.

4. Discussion

StrepDVU2108 expression and purification in a Desulfovibrio vulgaris Hildenborough strain. The expression and purification of StrepDVU2108 under anaerobic conditions in *D. vulgaris* Hildenborough, with and without metal supplementation (Cu and Mo), showed similar and interesting results that raised new questions concerning the native form of DVU2108.

The concentration of metals used in metal supplementation of POSTGATE C medium had a minor influence in the strain growth since the optic densities throughout the process and the cellular mass obtained were slightly lower in comparison to the expression in POSTGATE C. Nevertheless, protein concentration was quite similar in both growth conditions even though it was not possible to obtain a pure sample of DVU2108 (Table 4.1).

Table 4.1 – Results comparison from the two growths

Medium	Growth Volume (L)	Final $OD_{600\ nm}$	Cellular Mass (g)	[Protein] (mg/mL)
POSTGATE C	12	0.727	20	0.30
POSTGATE C + Cu and Mo	12	0.599	17	0.39

The spectra obtained for protein sample in both growths were very similar and showed distinct absorption bands that could be from a Fe-S protein[33] (Figure 4.1) with no evidence of contribution from a MoCu-DVU2108 (SI 7.11). In addition, iron concentrations in both protein samples were higher than the ones of molybdenum and copper (Table 4.2) which is in agreement with the acquired spectra.

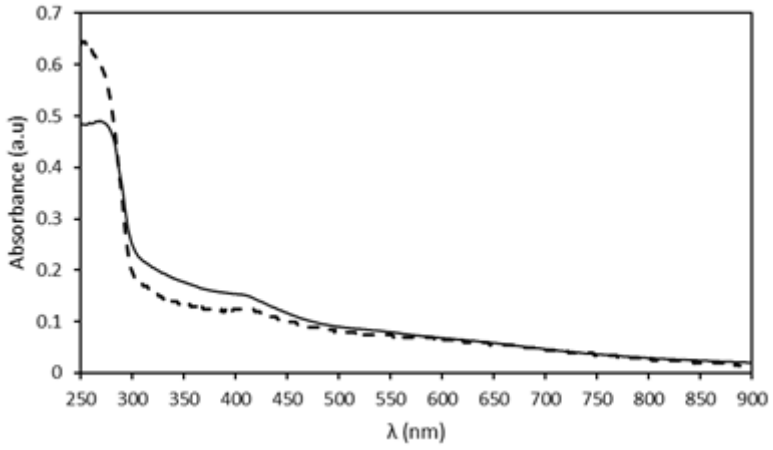

Figure 4.1 – UV-visible spectrum of protein sample after concentration from both growths. Full line – POSTGATE C medium supplemented with Mo and Cu; Dashed line – POSTGATE C medium.

Table 4.2 – Metal concentrations and ratios (normalized to Mo concentration) obtained for both growths.

Medium	[Mo] (μM)	[Cu] (μM)	[Fe] (μM)	Metal ratios
POSTGATE C	0.5	13.50	41.4	1Mo:30Cu:92Fe
POSTGATE C + Cu and Mo	0.7	1.96	14.9	1Mo:2.7Cu:20Fe

The protein sample of the POSTGATE C metal supplementation approach seems to have a higher Fe-S cluster-to-protein ratio with a more intense absorption bands at 400 and 550 nm. Due to the number of contaminants observed in SDS-PAGE analysis of both growths we are unable to determine if it is DVU2108 that has an iron-sulfur cluster or one of the contaminant proteins. Pulldown experiments in *D. vulgaris* Hildenborough with a strep tagged DVU2108 had similar purification results, where they identified the contaminants (SI 7.10).[16] A raw identification of the contaminating proteins obtained for both growths accordingly to their molecular mass is shown in Figure 4.2.

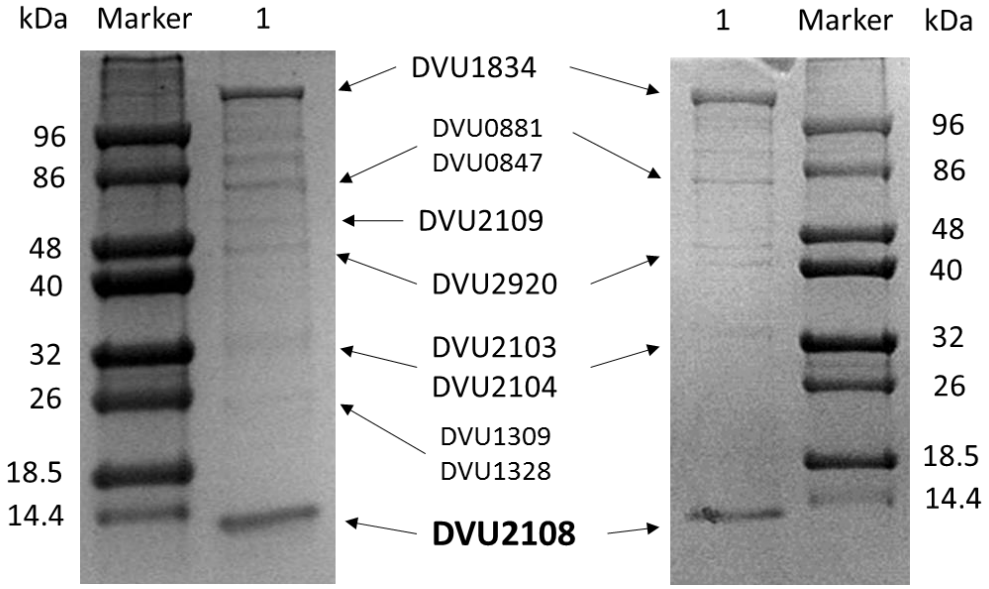

Figure 4.2 – Contaminating proteins from both growths. SDS-PAGE of POSTGATE C growth on the left; SDS-PAGE of POTGATE C supplemented with Cu and Mo on the right.

Proteins DVU2103, DVU2104 and DVU2109 belong to the ORP multiprotein complex in *D. vulgaris* Hildenborough and it has been proposed that DVU2108 interacts directly with DVU2103 and DVU2104[16], which caused them to be co-purified with DVU2108. Even though DVU2103 seems to have two 4Fe-4S clusters in its structure (Carreira and Pauleta, unpublished data), its concentration is negligible in comparison to DVU2108 and other contaminating proteins and thus its contribution to the UV-visible spectrum cannot be taking into account. On the other hand, DVU0847 is an adenyl sulfate reductase with 12 non-hemic irons and a strong absorption band at 372 nm[50], which could explain the higher iron concentration on both samples but not the spectrum. On the top of both lanes 1 is the larger contaminant, DVU1834 which is a pyruvate carboxylase. This kind of proteins have a biotin site[51] that binds strongly to streptavidin, and this interaction is the foundation of strep tagged purification principle, which explains the high concentration of this protein after the purification process. Even though this protein has a single

manganese ion in each subunit it does not contribute to the absorption spectra of the protein sample.[52]

A future suggestion to improve this first purification step will be the use of a higher amount of NaCl, which might disrupt a protein complex that is formed with DVU2108. As for the contamination with pyruvate carboxylase, the only way to separate might be to use a second chromatographic step using a gel filtration chromatography.

This analyze increase the possibility that DVU2108 could be a Fe-S cluster containing protein since the concentrations of molybdenum and copper are very low. However, some of the contaminants could be wrongly identified and there are others which were not identified at all and may be contributing to the UV-visible spectrum and metal content of the protein sample.

Unfortunately, the supplementation of Mo and Cu to the POSTGATE C medium in order to increase the metal uptake by DVU2108 was unsuccessful and surprisingly reveled an even lower amount of both metals. We do not know the reasons for these results, the cluster may have not been formed, DVU2108 may have lost it in the purification process or did not even incorporate it in the first place and so further investigation is necessary.

Apo-DVU2108 metal-cluster reconstitution and Metal Titration The native orange protein from *D. gigas* was successfully isolated and exhibited a Cu-Mo cluster with a metal ratio of two molybdenum atoms per copper atom.[5] This same metal ratio was found to be the preferred ratio for metal incorporation in Apo-DVU2108 as shown by titration with TTMo and TTW. In addition, protein-assisted reconstitution revealed that the Apo-form of DVU2108 is capable of cluster incorporation (1Cu-2Mo and 1Cu-2W) in similar conditions as used in recent studies for reconstitution of Apo-ORP from *D. gigas*[32]. In both reconstitution and titration experiments the metal cluster incorporation by Apo-DVU2108 was performed in the presence of oxygen, at room temperature and, for titration only, in a stirring solution indicating a great stability of the reconstituted DVU2108 which was later confirmed in DSC experiments. Reconstituted DVU2108 is even more stable than the Apo-DVU2108 and can endure temperatures up to 72ºC before total loss of conformation. This results suggest that in StrepDVU2108 expression and purification the hypothesis of cluster loss or failed incorporation are less likely to have happened, if the native cluster is in fact a Mo-Cu as it was suggested in prior studies on *D. gigas* ORP.

As stated before in Introduction, DVU2108 belongs to COG1433 which has sequence similarity with NifB and NifX, two proteins that may have a role in the biosynthesis of a MoFe cofactor in nitrogen fixation. NifB is an iron-sulfur containing protein and has been demonstrated to be a precursor in the cofactor biosynthesis, whereas NifX is thought to be involved in the transfer of MoFe cofactor to dinitrogenase reductase or to other proteins that complete the cofactor synthesis.[15] This suggests that DVU2108 may be involved in a similar process where it is a precursor for the synthesis of a MoCu cofactor, since it is able to bind to such cluster, or a FeS cofactor since neighbor proteins are proposed to have a FeS cluster (Carreira and Pauleta, unpublished data) and there is the possibility that DVU2108 is an iron-sulfur protein as well or might be involved in the transfer of a FeS cluster.

NMR assignment and Structure of Desulfovibrio vulgaris Hildenborough Apo-DVU2108. The assignment yield of 94 % of the NMR detectable atoms in Apo-DVU2108 allowed us to elucidate structural aspects unknown until now. Secondary structure predictions revealed the existence of 5 beta-sheets (β1-5) and 3 main alpha-helixes (α1-3) and a smaller helix with 3 residues long (Figure 4.3).

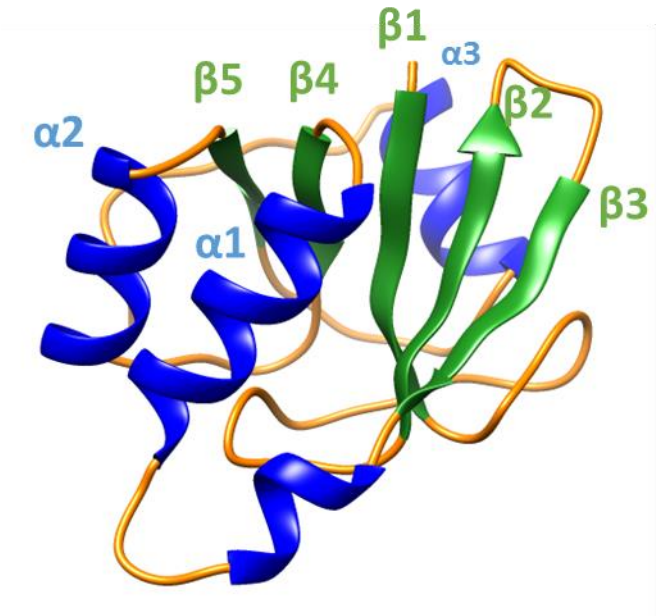

Figure 4.3 – Structural model for Apo-DVU2108 from D. vulgaris Hildenborough. Blue – alpha helix; Green – beta sheets. Picture generated by Chimera™

Secondary structures were numbered according to their type and position in the protein sequence: **β1** (1-8), **β2** (25-30), **β3** (35-40), **α1** (52-61), **β4** (65-70), **α2** (74-82), **β5** (84-90), α3 (95-102). Despite some differences in length, the regions of these secondary structures are consistent in both TALOS+ and CCPN predictions, as well as, in the model obtained from the *D. gigas* Apo-ORP structure. Figure 4.4 shows the sequence alignment of DVU2108 homologous for which the structure has been deposited.

```
D. vulgaris Hildenborough  1   M-RIVITAQGNTLDSPLDPRFGRARSFIVCDTETGATHAVDVSANMNLA-QG-AGIQAAQMAADAGAEAVITGHVGPKAF  77
D. gigas                   1   MQRIAVTAEGPGLDGLVDPRFGRAAGFVVVDAATMAAEYVDNGASQTLS-HG-AGINAAQVLAKSGAGVLLTGYVGPKAF  78
M. thermoautotrophicum     1   M-KIAIASSGTDLGSEVSRFFGRAPYFMIVEMKKGNIESSEVIENPSASASGGAGIRTAQIIANNGVKAVIASSPGPNAF  79
T. maritima                1   MARVAIPSVGKDLSSMVSDRFARAEYFIIYDTESGNV---EVVENTIADAHG-TGPKVVQSLVSKGVEYLIASNVGRNAF  76

D. vulgaris Hildenborough  78  TALNRGHIAVYLC-DLATPREALAAFIEGKLRPADTADR-EGHW--  119
D. gigas                   79  QALQAAGIKVGQDLEGLTVRQAVQRFLDGQVPMAAGPNK-------  117
M. thermoautotrophicum     80  EVLNELGIKIYRA-TGTSVEENLKLFTEGNLEEIRSPGSGRGRRRR  124
T. maritima                77  ETLKAAGVKVYRF-EGGTVQEAIDAFSEGRLEELTTFTR-EG----  116
```

Figure 4.4 – Sequence alignment of DVU2108 with homologues with known and deposited structure.

All four structures show the same type of conformation: five or six β-sheets between the three α-helixes, one on the opposite side of the other two (Figure 4.5), with the arrangement of secondary structure elements being β1-β2-β3-α1-β4-α2-β5-α3. MTH1175 has the same ordering but with the extra β-sheet located after its α3 helix.

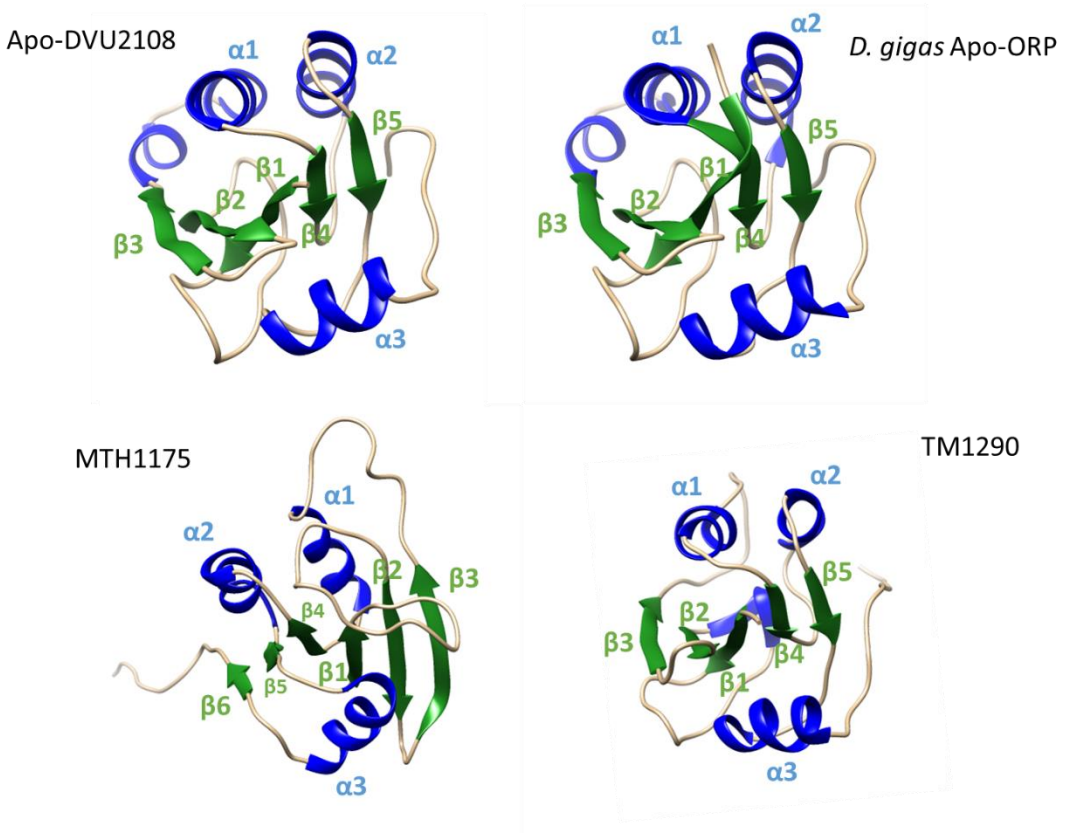

Figure 4.5 – Comparison of DVU2108 to similar structures from the ORP family. Secondary structures are labeled by their number in their respective structure and by color in all proteins (Blue – alpha helix; Green – beta sheets). This figure was generated with CHIMERA with PDB codes for each protein except DVU2108: D. gigas ORP - 2WFD, MTH1175 - 1EO1 and TM1290 – 1RDU.

The five β-sheets in Apo-DVU2108, *D. gigas* Apo-ORP and TM1290 are very similar in terms of their orientation wherein the first three sheets are antiparallel with β1 and β3 oppositely oriented in relation to β2. The remaining sheets, β4 and β5, are parallel with each other and oriented in the same direction as β1 and β3. In MTH1175 the first five sheets display the same arrangement as the other proteins but it possesses an extra β-sheet, β6, antiparallel to β5 which is atypical in β-α-β motifs that usually have parallel oriented sheets[9], just like it happens for the β4-α2-β5 motifs found in all four proteins. TM1290 has a distinctive 3_{10}-helical turn between β1 and β2 sheets and a well-structured loop connecting it to β1. Helices α1 and α2 are placed nearer one side of the β-sheets, specifically β1, β4 and β5, which leaves β2 and β3 (and β6 in MTH1175) sheets more exposed to the solvent.

The first residues of the large loop connecting β3 to α1 in Apo-DVU2108 and *D. gigas* Apo-ORP form a small α-helix which is not present in the other structures. Later in that same loop there is a flexible region that is poorly defined in *D. gigas* Apo-ORP (residues 52-55), MTH1175 (46-52) and TM1290 (45-51) structures. This same region was classified as flexible in Apo-DVU2108 by TALOS+ secondary structure prediction, namely residues N45, M46 and Q49 and residues G50 and A51 did not had a detectable HN resonance in the ^{1}H-^{15}N HSQC. This may be an interaction site between these proteins and their corresponding physiological binding partners,[8] where in DVU2108 could be with DVU2103 or DVU2104 which are the two proteins in the multi protein

complex for which DVU2108 has a direct interaction.[16] Comparative NMR studies of *D. gigas* Apo-ORP with its reconstituted form has shown that this flexible site includes the residues most affected by metal cluster incorporation[32] (Figure 4.6).

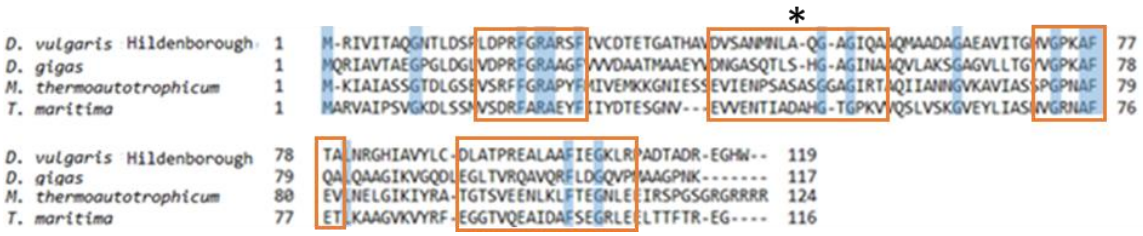

Figure 4.6 – Sequence alignment of the DVU2108 homologues. The orange squares indicate the affected residues by metal cluster incorporation in D. gigas ORP, the asterisk signals the region where the flexible site is present.

This regions in *D. gigas* Apo-ORP structure are in direct contact with the solvent, indicating that the metal cluster binding site is located on the protein surface. This results raise the possibility that the flexible region in Apo-DVU2108 might also be involved in the binding of its metal cluster instead of interacting with partner proteins (Figure 4.7), thus future experiments are needed to establish its role.

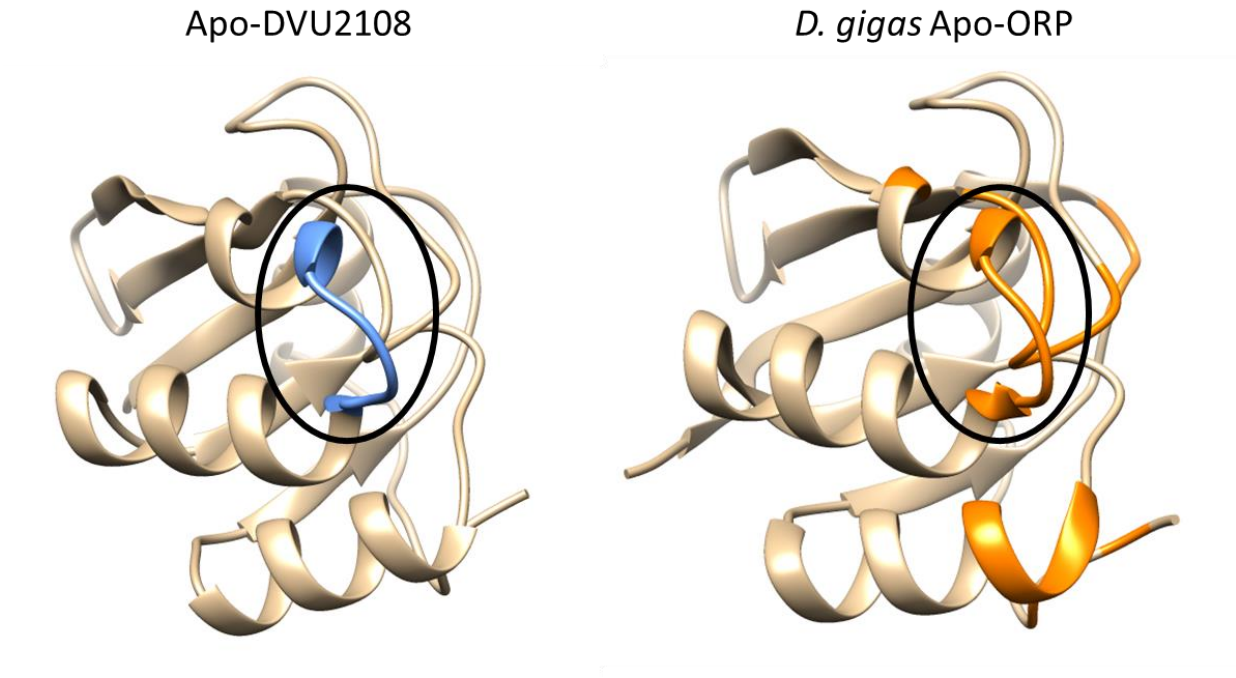

Figure 4.7 – Structure comparison between Apo-DVU2108 and D. gigas ORP with residues affected by the binding of the metal cluster in orange and the flexible site in Apo-DVU2108 in blue.

The electrostatic surface potential of ORP from *D. gigas* and DvH is similar (Figure 4.8) with a more negative region on one side of the structure and a more positive region on the opposite side. The proposed binding site for the metal cluster is located on the positive side in *D. gigas*

ORP, which is as expected due to the negative charge of the cluster.[32] Interestingly, the flexible region identified in DVU2108 also corresponds to a positive surface.

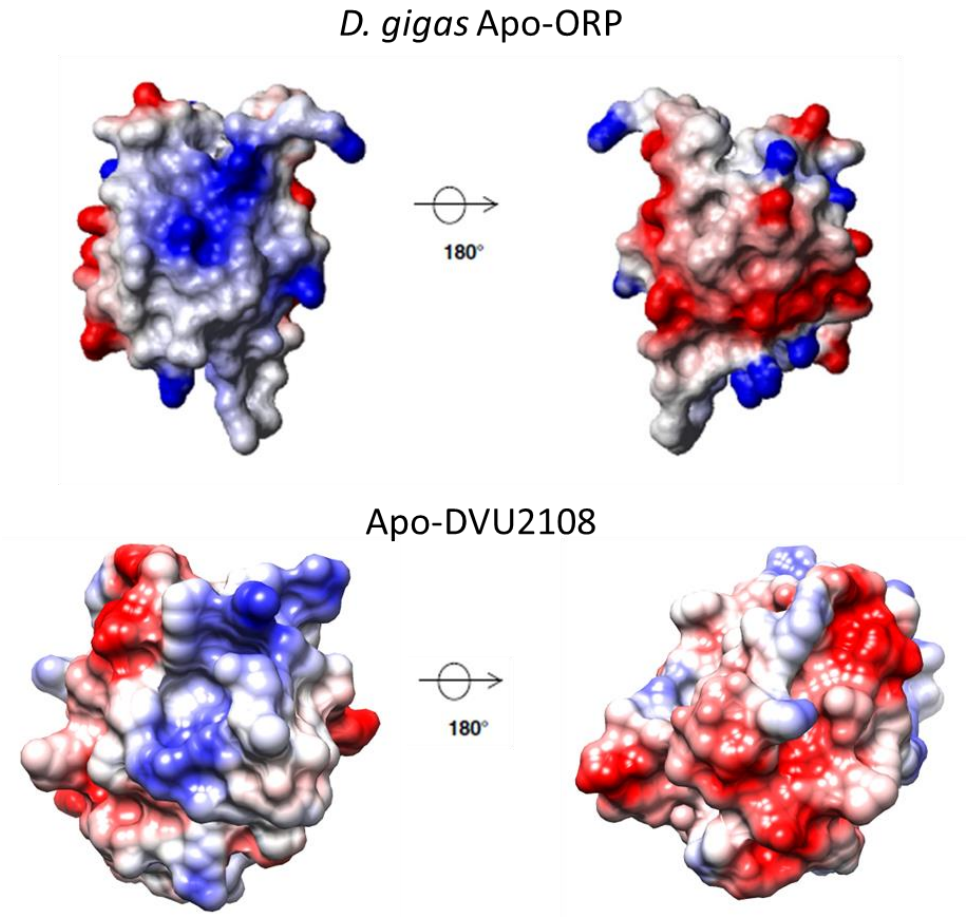

Figure 4.8 – Protein surfaces colored according to the electrostatic potential (blue positive and red negative). Apo-DVU2108 was prepared in Chimera™ and D. gigas Apo-ORP was reprinted from [32]

Since both structures are very similar (with a relatively high sequence homology), by comparing them it is possible to propose that the metal cluster binding site in DVU2108 is on a region of the protein with more positive residues that create a positive surface. Further investigation for the reconstituted DVU2108 is required to confirm this hypothesis for the location of the metal cluster binding site.

5. Conclusion

The Apo form of *Desulfovibrio vulgaris* Hildenborough ORP homologue, DVU2108, is capable of incorporating the unique molybdenum-copper cluster found in native ORP from *D. gigas* in a protein-assisted reconstitution with the same metal to metal ratio: two molybdenum atoms for each copper atom and cluster to protein ratio (1:1). This metal cluster incorporation was shown to improve the protein stability since the reconstituted form endures higher temperatures before

loss of conformation than its Apo form according to DSC results. Despite this ability, and even though it was not possible to fully isolate the native DVU2108, the results obtained in this study raise the possibility of DVU2108 being an iron-sulfur protein or being involved in the transfer of an FeS cluster. Nevertheless, the metal cluster binding site is proposed to be the same as for *D. gigas* ORP since the amide resonances that could not be identified in the ^{1}H-^{15}N HSQC and dynamic residues are located in a similar region as in *D. gigas*, region that was then identified as the one affected by the incorporation of the metal cluster. Secondary structure predictions and the structural model for Apo-DVU2108 revealed a similar structure to other proteins in the ORP family for which structure is known: MTH1175, TM1290 and particularly *D. gigas* ORP. The resonance assignment will enable us to perform protein titrations to identify the interaction surface with the other ORP complex proteins.

6. Bibliography

1. Rabus R, Hansen TA, Widdel F. *Dissimilatory Sulfate-and Sulfur-Reducing Prokaryotes*.; 2006.

2. Muyzer G, Stams AJM. The ecology and biotechnology of sulphate-reducing bacteria. *Nat. Rev. Microbiol.* 2008;6(6):441-54. doi:10.1038/nrmicro1892.

3. Hansen TA. Metabolism of sulfate-reducing prokaryotes. 1994;3(Iv):165-185.

4. Heidelberg JF, Seshadri R, Haveman S a, et al. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat. Biotechnol.* 2004;22(5):554-9. doi:10.1038/nbt959.

5. Bursakov S a, Gavel OY, Di Rocco G, et al. Antagonists Mo and Cu in a heterometallic cluster present on a novel protein (orange protein) isolated from *Desulfovibrio gigas. J. Inorg. Biochem.* 2004;98(5):833-40. doi:10.1016/j.jinorgbio.2003.12.002.

6. Pauleta SR, Duarte AG, Carepo MS, et al. NMR assignment of the apo-form of a *Desulfovibrio gigas* protein containing a novel Mo-Cu cluster. *Biomol. NMR Assign.* 2007;1(1):81-3. doi:10.1007/s12104-007-9022-3.

7. George GN, Pickering IJ, Yu EY, et al. A Novel Protein-Bound Copper - Molybdenum Cluster. 2000:8321-8322.

8. Etezady-Esfarjani T, Herrmann T, Peti W, Klock HE, Lesley S a., Wüthrich K. NMR structure determination of the hypothetical protein TM1290 from *Thermotoga maritima* using automated NOESY analysis [5]. *J. Biomol. NMR* 2004;29(3):403-406. doi:10.1023/B:JNMR.0000032615.51536.1a.

9. Cort JR, Yee a, Edwards a M, Arrowsmith CH, Kennedy M a. NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from *Methanobacterium thermoautotrophicum*. *J. Struct. Funct. Genomics* 2000;1(1):15-25.

10. Najmudin S, Bonifácio C, Duarte AG, et al. Crystallization and crystallographic analysis of the apo form of the orange protein (ORP) from *Desulfovibrio gigas*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 2009;65(7):730-732. doi:10.1107/S1744309109023392.

11. Altschup SF, Science C, Pennsylvania T, University S, Park U. Basic Local Alignment Search Tool. 1990:403-410. doi:10.1016/S0022-2836(05)80360-2.

12. Elias D a, Mukhopadhyay A, Joachimiak MP, et al. Expression profiling of hypothetical genes in *Desulfovibrio vulgaris* leads to improved functional annotation. *Nucleic Acids Res.* 2009;37(9):2926-39. doi:10.1093/nar/gkp164.

13. Tatusov RL, Koonin E V, Lipman DJ. A genomic perspective on protein families. *Science* 1997;278(5338):631-637. doi:10.1126/science.278.5338.631.

14. Dermoun Z, Foulon A, Miller MD, et al. TM0486 from the Hyperthermophilic Anaerobe *Thermotoga maritima* is a Thiamin-binding Protein Involved in Response of the Cell to Oxidative Conditions. *J. Mol. Biol.* 2010;400(3):463-476. doi:10.1016/j.jmb.2010.05.014.

15. Shah VK, Rangaraj P, Chatterjee R, et al. Requirement of NifX and other nif proteins for in vitro biosynthesis of the iron-molybdenum cofactor of nitrogenase. *J. Bacteriol.* 1999;181(9):2797-2801.

16. Fiévet A, My L, Cascales E, et al. The anaerobe-specific orange protein complex of *Desulfovibrio vulgaris* Hildenborough is encoded by two divergent operons coregulated by σ54 and a cognate transcriptional regulator. *J. Bacteriol.* 2011;193(13):3207-19. doi:10.1128/JB.00044-11.

17. Bertoni G, Fujita N, Ishihama A, De Lorenzo V. Active recruitment of σ54-RNA polymerase to the Pu promoter of Pseudomonas putida: Role of IHF and αCTD. *EMBO J.* 1998;17(17):5120-5128. doi:10.1093/emboj/17.17.5120.

18. De Lorenzo V, Herrero M, Metzke M, Timmis KN. An upstream XylR- and IHF-induced nucleoprotein complex regulates the sigma 54-dependent Pu promoter of TOL plasmid. *EMBO J.* 1991;10(5):1159-1167.

19. Swinger KK, Rice P a. IHF and HU: Flexible architects of bent DNA. *Curr. Opin. Struct. Biol.* 2004;14:28-35. doi:10.1016/j.sbi.2003.12.003.

20. Fiévet A, Cascales E, Valette O, Dolla A, Aubert C. IHF is required for the transcriptional regulation of the *Desulfovibrio vulgaris* Hildenborough *orp* operons. *PLoS One* 2014;9(1):e86507. doi:10.1371/journal.pone.0086507.

21. Scholten JC, Culley DE, Brockman FJ, Wu G, Zhang W. Evolution of the syntrophic interaction between *Desulfovibrio vulgaris* and *Methanosarcina barkeri*: Involvement of an ancient horizontal gene transfer. *Biochem. Biophys. Res. Commun.* 2007;352(1):48-54. doi:10.1016/j.bbrc.2006.10.164.

22. Hammes GG. Spectroscopy for the Biological Sciences. In: 1st ed. Hoboken, New Jersey: Wiley-Interscience; 2005:103-120.

23. Clayden J, Greeves N, Warren S, Wothers P. Organic Chemistry. *Structure* 2000;(2008):1536. doi:10.1016/j.snb.2009.03.030.

24. Cavanagh J, Fairbrother WJ, Palmer AG. *Protein NMR Spectroscopy: Principles and Practice*. 2nd ed.; 2007.

25. Xie XQ, Moring J, Makriyannis a. Differential scanning calorimetry and small angle x-ray diffraction study of the interaction of (R)-PAF, (R)-ET-18-OME and (R)-Lyso-PAF with model membranes. *Life Sci.* 1997;61(9):909-923.

26. Gill P, Moghadam TT, Ranjbar B. Differential scanning calorimetry techniques: applications in biology and nanoscience. *J. Biomol. Tech.* 2010;21(4):167-193.

27. Coleman NJ, Craig DQM. Modulated temperature differential scanning calorimetry: A novel approach to pharmaceutical thermal analysis. *Int. J. Pharm.* 1996;135(1-2):13-29. doi:10.1016/0378-5173(95)04463-9.

28. Chiu MH, Prenner EJ. Differential scanning calorimetry: An invaluable tool for a detailed thermodynamic characterization of macromolecules and their interactions. *J. Pharm. bioallied Sci.* 2011;3(1):39-59. doi:10.4103/0975-7406.76463.

29. Holde KE Van, Johnson WC, Ho PS. *Principles of Physical Biochemistry Principles of Physical Biochemistry*.; 2006. doi:10.1021/ed076p474.1.

30. Lowry OH, Rosebrough NJ, Farr AL, Randall RJ. Protein Measurement with the Folin Phenol Reagent. *J Biol Chem* 1951;193(1):265-275. doi:10.1016/0304-3894(92)87011-4.

31. Wagner G, Ferentz AE. NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q Rev Biophys* 2000;01(33):29-65.

32. Carepo MSP, Pauleta SR, Wedd AG, Moura JJG, Moura I. Mo-Cu metal cluster formation and binding in an orange protein isolated from *Desulfovibrio gigas*. *J. Biol. Inorg. Chem.* 2014;19(4-5):605-14. doi:10.1007/s00775-014-1107-8.

33. Liu J, Chakraborty S, Hosseinzadeh P, et al. *Metalloproteins Containing Cytochrome , Iron − Sulfur , or Copper Redox Centers*.; 2015.

34. Chang IS, Groh JL, Ramsey MM, Ballard JD, Krumholz LR. Differential Expression of *Desulfovibrio vulgaris* Genes in Response to Cu (II) and Hg (II) Toxicity Differential Expression of *Desulfovibrio vulgaris* Genes in Response to Cu (II) and Hg (II) Toxicity. *Appl. Environ. Microbiol.* 2004;70(3):1847-1851. doi:10.1128/AEM.70.3.1847.

35. Da Silva SM, Pimentel C, Valente FM a, Rodrigues-Pousada C, Pereira I a C. Tungsten and molybdenum regulation of formate dehydrogenase expression in *Desulfovibrio vulgaris* Hildenborough. *J. Bacteriol.* 2011;193(12):2909-2916. doi:10.1128/JB.00042-11.

36. L. Nelson D, M. Cox M. *Lehninger Principles of Biochemistry*. 4th ed.; 2004.

37. El-Gallad TT, Mills CF, Bremner I, Summers R. Thiomolybdates in rumen contents and rumen cultures. *J. Inorg. Biochem.* 1983;18(4):323-334. doi:10.1016/0162-0134(83)85047-8.

38. Zhou Y, Hall CK, Karplus M. The calorimetric criterion for a two-state process revisited. *Protein Sci.* 1999;8(5):1064-1074. doi:10.1110/ps.8.5.1064.

39. Sattlera M, Schleucherb J, Griesingerc C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* 1999;34(34):93–158.

40. Keller R. *The Computer Aided Resonance Assignment Tutorial*.; 2004.

41. Griesinger C, Sattler M. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. 1999;34:93-158.

42. Glc I. Tautomeric states of the active-site histidines of. 1993:543-558.

43. Simplaceanu V, Lukin J a, Fang TY, Zou M, Ho NT, Ho C. Chain-selective isotopic labeling for NMR studies of large multimeric proteins: application to hemoglobin. *Biophys. J.* 2000;79(2):1146-1154. doi:10.1016/S0006-3495(00)76368-5.

44. Wedemeyer WJ, Welker E, Scheraga H a. Current Topics Proline Cis - Trans Isomerization and Protein Folding †. *Biochemistry* 2002;41:14637-14644. doi:10.1021/bi020574b.

45. Schubert M, Labudde D, Oschkinat H, Schmieder P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J. Biomol. NMR* 2002;24(2):149-154. doi:10.1023/A:1020997118364.

46. Sharma D, Rajarathnam K. 13C NMR chemical shifts can predict disulfide bond formation. *J. Biomol. NMR* 2000;18(2):165-171. doi:10.1023/A:1008398416292.

47. Shen Y, Delaglio F, Cornilescu G, Bax A. TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* 2009;44(4):213-223. doi:10.1007/s10858-009-9333-z.

48. Berjanskii M V., Wishart DS. The RCI server: Rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res.* 2007;35(SUPPL.2):531-537. doi:10.1093/nar/gkm328.

49. Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 1992;31(6):1647-1651. doi:10.1021/bi00121a010.

50. Bramlett RN, Peck HD. Some physical and kinetic properties of adenylyl sulfate reductase from *Desulfovibrio vulgaris*. *J. Biol. Chem.* 1975;250(8):2979-2986.

51. Attwood P V. The structure and the mechanism of action of pyruvate carboxylase. *Int. J. Biochem. Cell Biol.* 1995;27(3):231-249.

52. Frey WH, Utter MF. Binding of acetyl-CoA to chicken liver pyruvate carboxylase. *J. Biol. Chem.* 1977;252(1):51-56.

7. Supplementary Information

7.1 POSTGATE C Medium

Table 7.1 – POSTGATE C Medium Components

Brand	Components	Chemical formula	Quantity (per L)
SIGMA	Potassium phosphate monobasic	KH_2PO_4	0.5 g
SCHARLAU	Ammonium chloride	NH_4Cl	1 g
PANREAC	Sodium sulfate	Na_2SO_4	4.5 g
LABCHEM	Magnesium sulfate heptahydrate	$MgSO_4 \cdot 7H_2O$	0.06 g
SCHARLAU	Calcium chloride dihydrate	$CaCl_2 \cdot 2H_2O$	0.04 g
SIGMA	Sodium lactate – 60%	Na-Lactate 60%	5.3 ml
SCHARLAU	Yeast extract	-	1 g
SIGMA	Iron(II) sulfate heptahydrate	$FeSO_4 \cdot 7H_2O$	0.004 g
RIEDEL-DE HAEN	tri-Sodium citrate dihydrate	$C_6H_5Na_3O_7 \cdot 2H_2O$	0.3 g
ALFA AESER	Ascorbic acid	$C_6H_8O_6$	0.1 g
SIGMA	Thioglycolic acid	$C_2H_4O_2S$	0.1 g

7.2 MoCu Reconstituted ORP from *D. gigas*

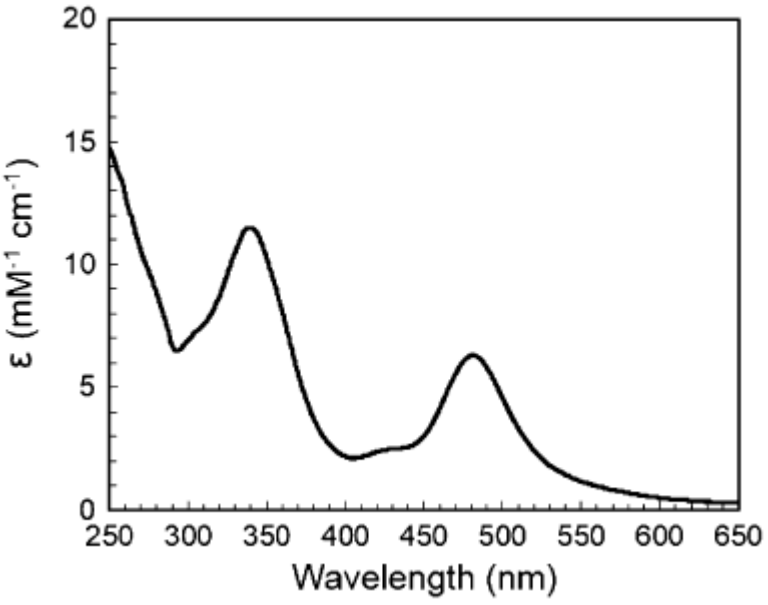

Figure 7.1 – UV-visible spectrum of MoCu reconstituted ORP from D. gigas. Reprinted from [32]

7.3 WCu Reconstituted ORP from *D. gigas*

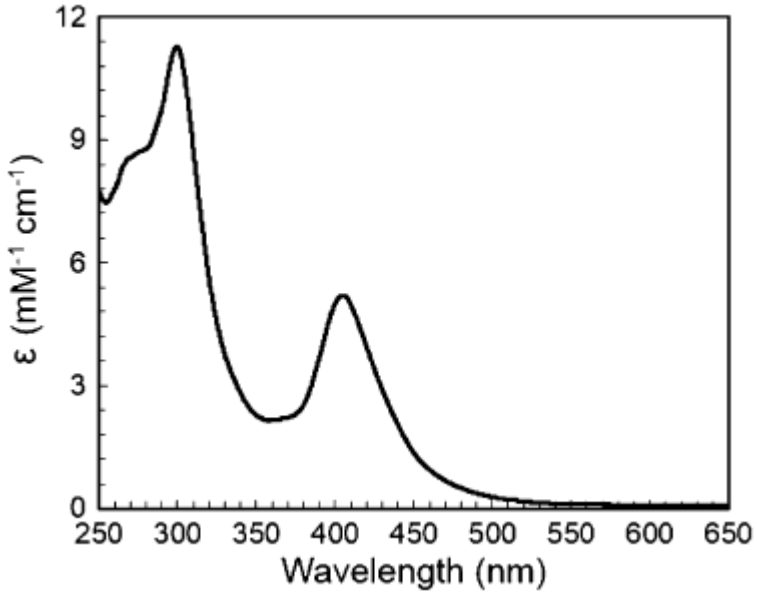

Figure 7.2 – UV-visible spectrum of WCu reconstituted ORP from D. gigas. Reprinted from [32]

7.4 UV-visible spectra for TTMo titration to Apo-DVU2108

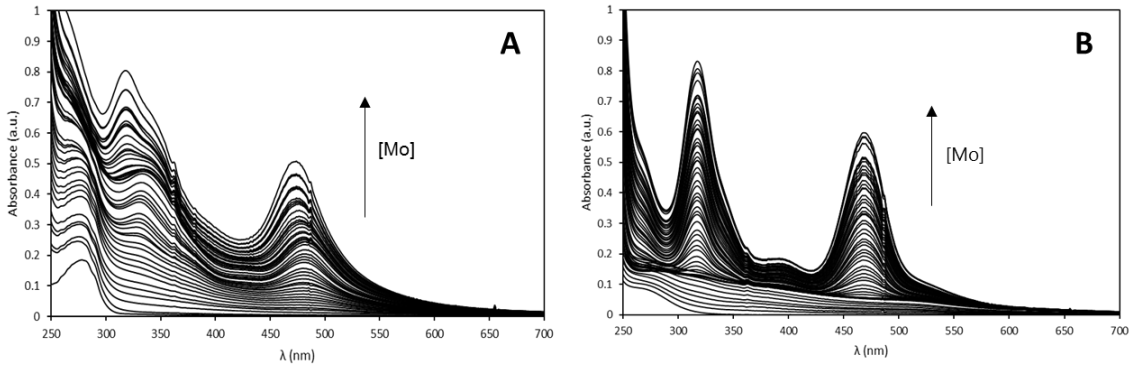

Figure 7.3 – Titration spectra with TTMo in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF. Spectra acquired after each addition. ***A*** *- Titration of 30 µM Apo-DVU2108, in the presence of 30 µM $CuCl_2$, with TTMo.* ***B*** *- Titration of 30 µM $CuCl_2$ with TTMo.*

7.5 UV-visible spectra for TTW titration to Apo-DVU108

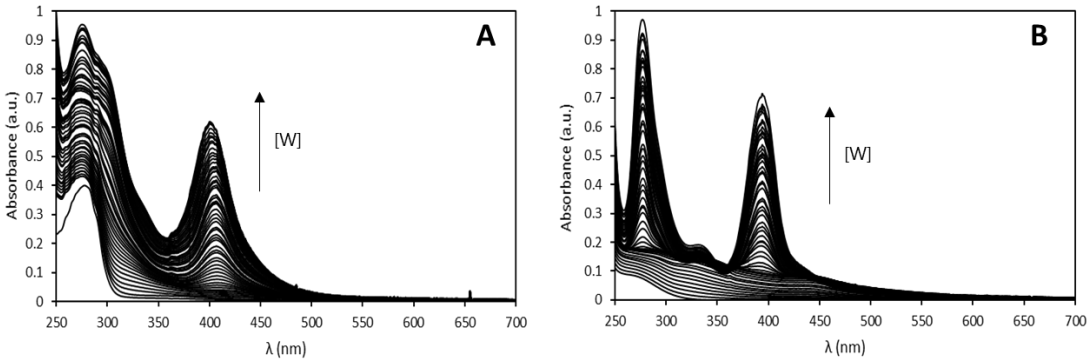

Figure 7.4 – Titration spectra with TTW in 50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 20% DMF. Spectra acquired after each addition. ***A*** *- Titration of 30 µM Apo-DVU2108, in the presence of 30 µM $CuCl_2$, with TTW.* ***B*** *- Titration of 30 µM $CuCl_2$ with TTW.*

7.6 Chemical Shift list of all the atoms from all residues assigned

Table 7.2 - Chemical shift list of all assigned atoms created by CCPN analysis (residues 1-30).

	H	N	H_α	H_β	C	C_α	C_β	
1 Met			4.21	1.91 2.08	169.23	52.16	32.24	$H^{\gamma a}$ 2.39,$H^{\gamma b}$ 2.39,$H^{\varepsilon *}$ 1.92,C^{γ} 28.06, C^{ε} 14.36
2 Arg	9.30	126.43	5.28	1.85 1.69	172.42	52.86	27.87	N^{ε} 83.94,$H^{\gamma a}$ 1.21,$H^{\gamma b}$ 1.21,$H^{\delta a}$ 3.07, $H^{\delta b}$ 3.07,H^{ε} 7.62,C^{γ} 25.18,C^{δ} 40.33
3 Ile	9.04	123.33	4.91	1.21	171.64	56.24	38.86	$H^{\gamma 1a}$ 0.57,$H^{\gamma 1b}$ 0.92,$H^{\gamma 2*}$ 0.28,$H^{\delta 1*}$ -0.04, $C^{\gamma 1}$ 23.77,$C^{\gamma 2}$ 15.11,$C^{\delta 1}$ 10.78
4 Val	8.17	120.22	4.94	1.41	171.15	56.71	31.69	$H^{\gamma a*}$ 0.13,$H^{\gamma b*}$ 0.51,$C^{\gamma a}$ 18.03,$C^{\gamma b}$ 19.33
5 Ile	8.85	124.69	5.17	1.39	174.39	56.97	38.85	$H^{\gamma 1a}$ 0.68,$H^{\gamma 1b}$ 0.68,$H^{\gamma 2*}$ 0.78,$H^{\delta 1*}$ 0.66, $C^{\gamma 1}$ 24.77,$C^{\gamma 2}$ 14.70,$C^{\delta 1}$ 12.32
6 Thr	8.10	116.33	4.74	4.42	171.94	58.90	65.84	$H^{\gamma 2*}$ 1.45,$C^{\gamma 2}$ 21.05
7 Ala	8.92	125.80	5.08	1.30	172.90	48.65	21.76	
8 Gln	8.84	118.42	4.62	2.30 1.69	172.57	52.35	26.24	$N^{\varepsilon 2}$ 111.12,$H^{\gamma a}$ 2.08,$H^{\gamma b}$ 2.08,$H^{\varepsilon 21}$ 7.27, $H^{\varepsilon 22}$ 6.97,C^{γ} 32.04,C^{δ} 176.67
9 Gly	7.11	105.70	4.08 4.05		169.30	41.78		
10 Asn			4.52	2.70 2.70	172.07	49.95	35.90	$N^{\delta 2}$ 112.53,$H^{\delta 21}$ 7.51,$H^{\delta 22}$ 6.85, C^{γ} 174.19
11 Thr	7.52	106.78	4.80	4.48	172.34	56.05	70.36	$H^{\gamma 2*}$ 1.04,$C^{\gamma 2}$ 18.57
12 Leu			3.80	1.16 1.61	173.16	53.95	40.19	$H^{\delta a*}$ 0.63,$H^{\delta b*}$ 0.76,C^{γ} 24.19,$C^{\delta a}$ 21.27, $C^{\delta b}$ 22.89
13 Asp	7.37	112.84	4.86	2.19 2.91	172.64	51.21	39.24	
14 Ser	7.22	119.51	4.46	3.76 3.63	165.59	55.82	59.63	
15 Pro			4.51	1.54 2.31	173.86	60.48	28.99	$H^{\gamma a}$ 1.93,$H^{\gamma b}$ 2.03,$H^{\delta a}$ 3.36,$H^{\delta b}$ 3.84, C^{γ} 25.49,C^{δ} 47.07
16 Leu	8.30	127.19	4.06	1.52 0.76	173.86	53.75	39.61	H^{γ} 0.50,$H^{\delta a*}$ 0.44,$H^{\delta b*}$ 0.44,C^{γ} 21.81, $C^{\delta a}$ 24.21,$C^{\delta b}$ 24.21
17 Asp	8.06	128.40	5.08	2.77 2.00	175.58	47.33	40.38	
18 Pro			4.20	1.83 2.44	173.23	60.20	30.78	$H^{\gamma a}$ 1.90,$H^{\gamma b}$ 1.95,$H^{\delta a}$ 3.90,$H^{\delta b}$ 3.90, C^{γ} 24.38,C^{δ} 48.31
19 Arg	8.06	115.90	4.28	1.80 1.80	172.01	51.74	26.85	$H^{\gamma a}$ 1.44,$H^{\gamma b}$ 1.49,$H^{\delta a}$ 2.81,$H^{\delta b}$ 2.81, C^{γ} 22.77,C^{δ} 40.49
20 Phe			4.24	3.17 3.55		59.27	37.58	$H^{\delta *}$ 7.07,$H^{\varepsilon *}$ 7.09,$C^{\delta *}$ 128.71,$C^{\varepsilon *}$ 128.52
21 Gly			3.39 3.64		170.65	42.95		
22 Arg	7.16	114.54	4.37	1.89 1.33	172.11	51.03	28.03	$H^{\gamma a}$ 1.40,$H^{\gamma b}$ 1.40,$H^{\delta a}$ 2.97,$H^{\delta b}$ 2.97, C^{γ} 24.16,C^{δ} 40.16
23 Ala	7.01	122.52	4.10	1.27	174.79	49.74	16.55	
24 Arg	8.47	118.99	4.08	1.85 1.94	174.49	54.77	28.01	$H^{\gamma a}$ 1.68,$H^{\gamma b}$ 1.68,$H^{\delta a}$ 3.13,$H^{\delta b}$ 3.13, C^{γ} 24.09,C^{δ} 40.43
25 Ser	7.91	112.39	5.22	3.64 3.45	170.51	54.41	63.88	
26 Phe	8.86	120.32	5.41	2.93 2.69	172.71	49.89	38.33	$H^{\delta *}$ 6.76,$C^{\delta *}$ 126.27
27 Ile	9.31	121.57	4.43	1.77	172.58	56.21	33.43	$H^{\gamma 1a}$ 0.90,$H^{\gamma 1b}$ 1.31,$H^{\gamma 2*}$ 0.59,$H^{\delta 1*}$ 0.39, $C^{\gamma 1}$ 24.73,$C^{\gamma 2}$ 15.02,$C^{\delta 1}$ 6.11
28 Val	9.23	129.69	4.46	1.14	173.23	58.45	28.12	$H^{\gamma a*}$ 0.65,$H^{\gamma b*}$ 0.65,$C^{\gamma a}$ 18.04,$C^{\gamma b}$ 18.04
29 Cys	9.12	124.63	5.05	2.46 2.55	170.52	53.73	27.52	
30 Asp	8.38	123.90	5.23	2.32 3.13	175.69	49.82	40.23	

Continuation of Table 7.2 (residues 31-60)

31 Thr	9.15	116.20	3.95	4.41	173.07	61.88	65.64	$H^{\gamma 2*}$ 1.17,$C^{\gamma 2}$ 18.93
32 Glu	8.72	120.57	4.30	2.06 2.06	175.39	55.32	27.55	$H^{\gamma a}$ 2.10,$H^{\gamma b}$ 2.21,C^{γ} 33.94
33 Thr	7.56	106.57	4.36	4.33	174.03	58.70	67.81	$H^{\gamma 2*}$ 1.12,$C^{\gamma 2}$ 18.56
34 Gly	8.17	111.16	3.60 4.14		170.83	42.59		
35 Ala	7.72	123.84	4.26	1.34	174.79	50.22	16.78	
36 Thr	8.17	113.35	5.52	3.81	171.21	57.30	70.18	$H^{\gamma 2*}$ 1.03,$C^{\gamma 2}$ 19.05
37 His	8.62	119.60	4.76	3.09 3.12	169.06	53.53	29.83	
38 Ala	8.52	123.17	5.44	1.25	174.42	47.89	18.82	
39 Val	9.06	121.05	4.21	1.83	171.94	58.10	32.88	$H^{\gamma a*}$ 0.80,$H^{\gamma b*}$ 0.91,$C^{\gamma a}$ 18.35,$C^{\gamma b}$ 18.35
40 Asp	8.57	126.26	4.70	2.65 2.48	173.43	51.36	38.20	
41 Val	7.91	124.01	3.84	1.59	174.06	59.75	28.23	$H^{\gamma a*}$ -0.15,$H^{\gamma b*}$ 0.22,$C^{\gamma a}$ 16.61,$C^{\gamma b}$ 18.13
42 Ser	8.03	116.84	4.05	3.83 3.83	173.21	57.94	60.26	
43 Ala	8.09	124.04	4.08	1.23	175.45	50.85	16.04	
44 Asn	7.85	115.36	4.59	2.54 2.81	172.67	50.58	36.29	$N^{\delta 2}$ 112.30,$H^{\delta 21}$ 7.49,$H^{\delta 22}$ 6.73, C^{γ} 174.03
45 Met	7.68	119.45	4.23	1.92 1.92	173.21	53.37	30.00	$H^{\gamma a}$ 2.39,$H^{\gamma b}$ 2.43,$H^{\varepsilon *}$ 2.05,C^{γ} 29.16, C^{ε} 14.29
46 Asn	8.47	119.26	3.86	2.54 2.54	174.62	49.77	36.35	$N^{\delta 2}$ 110.86,$H^{\delta 21}$ 7.32,$H^{\delta 22}$ 6.75, C^{γ} 174.26
47 Leu			4.19	1.52 1.56		52.90	39.21	H^{γ} 1.55,$H^{\delta a*}$ 0.75,$H^{\delta b*}$ 0.81,C^{γ} 24.11, $C^{\delta a}$ 20.31,$C^{\delta b}$ 22.28
48 Ala			4.13	1.34	175.31	50.67	15.95	
49 Gln	7.79	116.00	4.32	1.87 2.05	173.50	52.67	27.44	$N^{\varepsilon 2}$ 111.66,$H^{\gamma a}$ 2.25,$H^{\gamma b}$ 2.25,$H^{\varepsilon 21}$ 6.78, $H^{\varepsilon 22}$ 7.40,C^{γ} 31.18,C^{δ} 177.50
52 Gly			3.72 4.11		172.29	44.60		
53 Ile	7.76	120.02	3.73	1.83	175.55	60.94	34.58	$H^{\gamma 1a}$ 1.10,$H^{\gamma 1b}$ 1.42,$H^{\gamma 2*}$ 0.83,$H^{\delta 1*}$ 0.75, $C^{\gamma 1}$ 25.17,$C^{\gamma 2}$ 14.90,$C^{\delta 1}$ 9.90
54 Gln	7.94	121.15	4.08	1.98 2.06	175.78	56.40	25.73	$N^{\varepsilon 2}$ 112.44,$H^{\gamma a}$ 2.26,$H^{\gamma b}$ 2.33,$H^{\varepsilon 21}$ 7.13, $H^{\varepsilon 22}$ 6.98,C^{γ} 31.93,C^{δ} 176.97
55 Ala	8.39	122.95	3.81	0.88	176.46	52.54	15.24	
56 Ala	8.21	119.99	4.16	1.49	176.80	53.64	15.75	
57 Gln	7.82	119.23	3.89	2.18 2.16	175.38	55.49	25.25	$N^{\varepsilon 2}$ 115.48,$H^{\gamma a}$ 2.32,$H^{\gamma b}$ 2.32,$H^{\varepsilon 21}$ 7.76, $H^{\varepsilon 22}$ 6.83,C^{γ} 30.69,C^{δ} 177.60
58 Met	8.32	118.52	4.16	2.14 1.96	176.64	56.20	30.23	$H^{\gamma a}$ 2.51,$H^{\gamma b}$ 2.68,$H^{\varepsilon *}$ 1.88,C^{γ} 30.11, C^{ε} 14.99
59 Ala	8.41	122.59	3.77	1.20	175.47	52.72	13.71	
60 Ala	7.72	120.96	4.06	1.44	179.75	52.55	14.92	

Continuation of Table 7.2 (residues 61-90)

61 Asp	8.96	121.72	4.31	2.63 2.75	175.12	54.28	37.55	
62 Ala	7.58	119.93	4.30	1.41	174.47	49.48	14.87	
63 Gly	8.02	106.48	3.69 3.89		171.84	42.48		
64 Ala	7.55	119.07	3.67	1.05	173.99	50.35	16.63	
65 Glu	8.84	117.78	4.47	2.27 1.99	173.75	53.57	29.27	$H^{\gamma a}$ 2.29,$H^{\gamma b}$ 2.39,C^{γ} 35.35
66 Ala	7.68	119.90	5.39	1.22	172.35	47.96	19.59	
67 Val	8.86	120.17	4.88	1.48	170.91	56.34	32.80	$H^{\gamma a*}$ 0.52,$H^{\gamma b*}$ 0.52,$C^{\gamma a}$ 18.13,$C^{\gamma b}$ 18.54
68 Ile	8.95	125.58	5.16	1.54	171.41	57.32	36.09	$H^{\gamma 1a}$ 1.11,$H^{\gamma 1b}$ 1.11,$H^{\gamma 2*}$ 0.70,$H^{\delta 1*}$ 0.24, $C^{\gamma 1}$ 23.96,$C^{\gamma 2}$ 16.01,$C^{\delta 1}$ 11.09
69 Thr	8.31	120.58	4.55	4.07	169.40	56.21	65.61	$H^{\gamma 2*}$ 0.69,$C^{\gamma 2}$ 14.81
70 Gly	8.36	109.22			171.05	43.97		??[675] 3.83
71 His	8.13	120.01	4.95	2.80 2.86	170.88	57.07	30.30	$H^{\delta 2}$ 6.73,$C^{\delta 2}$ 117.21
72 Val			4.34	1.58	171.35	58.10	31.90	$H^{\gamma a*}$ 0.52,$H^{\gamma b*}$ 0.60,$C^{\gamma a}$ 19.00,$C^{\gamma b}$ 19.00
73 Gly	8.87	114.00	4.26 3.86		169.46	41.77		
75 Lys			4.05	1.66 1.87	177.19	56.73	28.85	$H^{\gamma a}$ 1.35,$H^{\gamma b}$ 1.51,$H^{\delta a}$ 1.61,$H^{\delta b}$ 1.61, $H^{\varepsilon a}$ 2.91,$H^{\varepsilon b}$ 2.91,C^{γ} 23.00,C^{δ} 26.36, C^{ε} 39.35
76 Ala	8.23	123.21	3.81	1.43	175.64	52.29	15.98	
77 Phe	8.68	118.35	3.62	3.01 2.87	174.88	59.54	36.61	$H^{\delta*}$ 6.87,$H^{\varepsilon*}$ 6.88,$C^{\delta*}$ 128.35,$C^{\varepsilon*}$ 126.90
78 Thr	8.12	114.82	3.72	4.12	173.13	64.05	66.14	$H^{\gamma 2*}$ 1.17,$C^{\gamma 2}$ 18.67
79 Ala	7.35	123.45	3.90	1.34	177.86	52.47	15.35	
80 Leu	8.21	120.11	3.79	1.11 1.46	175.84	55.27	37.73	H^{γ} 0.64,$H^{\delta a*}$ 0.05,$H^{\delta b*}$ 0.05,C^{γ} 20.67, $C^{\delta a}$ 23.01,$C^{\delta b}$ 23.01
81 Asn	8.39	117.94	4.06	2.35 2.35	177.20	54.06	36.17	$N^{\delta 2}$ 111.35,$H^{\delta 21}$ 7.32,$H^{\delta 22}$ 6.29, C^{γ} 173.03
82 Arg	8.30	121.00	3.85	1.77 1.77	174.55	55.89	26.80	$H^{\gamma a}$ 1.56,$H^{\gamma b}$ 1.69,$H^{\delta a}$ 3.04,$H^{\delta b}$ 3.04, C^{γ} 24.63,C^{δ} 40.57
83 Gly	7.24	105.00	3.36 4.10		170.30	42.04		
84 His	7.63	114.12	4.08	3.29 3.35	171.15	53.81	23.75	$N^{\delta 1}$ 195.54,$N^{\varepsilon 2}$ 181.77,$H^{\delta 2}$ 6.99, $H^{\varepsilon 1}$ 8.04,$C^{\delta 2}$ 117.41,$C^{\varepsilon 1}$ 134.34
85 Ile	7.88	121.10	3.82	1.40	172.07	58.61	35.06	$H^{\gamma 1a}$ 0.76,$H^{\gamma 1b}$ 0.76,$H^{\gamma 2*}$ 0.36,$H^{\delta 1*}$ 0.60, $C^{\gamma 1}$ 24.30,$C^{\gamma 2}$ 14.56,$C^{\delta 1}$ 10.77
86 Ala	7.41	129.95	4.06	1.38	172.81	49.39	18.07	
87 Val	8.43	120.89	4.34	1.68	170.51	58.60	31.05	$H^{\gamma a*}$ 0.44,$H^{\gamma b*}$ 0.44,$C^{\gamma a}$ 18.91,$C^{\gamma b}$ 19.53
88 Tyr	9.11	126.68	4.51	2.58 2.91	171.05	54.16	37.68	$H^{\delta*}$ 6.74,$H^{\varepsilon*}$ 6.51,$C^{\delta*}$ 130.25,$C^{\varepsilon*}$ 114.95
89 Leu	8.56	121.27	5.18	1.30 1.67	175.25	49.97	38.46	H^{γ} 0.81,$H^{\delta a*}$ 0.40,$H^{\delta b*}$ 0.40,C^{γ} 22.35, $C^{\delta a}$ 19.65,$C^{\delta b}$ 19.65
90 Cys	8.92	119.11	4.79	2.78	168.97	54.50	28.66	

Continuation of Table 7.2 (residues 91-119)

91 Asp	8.36	122.52	4.83	2.57 2.57	173.03	49.89	38.44	
92 Leu	6.95	119.33	4.36	1.39 1.39	173.93	51.85	39.33	H^{γ} 0.73,$H^{\delta a*}$ 0.85,$H^{\delta b*}$ 0.85,C^{γ} 23.43, $C^{\delta a}$ 19.32,$C^{\delta b}$ 19.32
93 Ala	8.36	117.84	4.11	1.54	177.37	53.08	17.81	
94 Thr	7.04	103.61	5.28	4.67	170.65	55.55	69.46	$H^{\gamma 2*}$ 1.04,$C^{\gamma 2}$ 19.02
95 Pro			4.13	2.23 1.91	173.73	63.03	29.39	$H^{\gamma a}$ 1.93,$H^{\gamma b}$ 2.02,$H^{\delta a}$ 3.62,$H^{\delta b}$ 3.62, C^{γ} 24.97,C^{δ} 46.90
96 Arg	8.54	116.40	3.53	1.60 1.70	176.04	57.94	27.53	N^{ε} 82.60,$H^{\gamma a}$ 1.46,$H^{\gamma b}$ 1.46,$H^{\delta a}$ 3.29, $H^{\delta b}$ 3.29,H^{ε} 8.86,C^{γ} 24.63,C^{δ} 39.77
97 Glu	7.71	118.01	3.98	2.14 2.14	176.62	56.16	28.46	$H^{\gamma a}$ 2.29,$H^{\gamma b}$ 2.29,C^{γ} 34.48
98 Ala	8.05	123.16	3.98	1.31	175.07	52.35	16.50	
99 Leu	8.87	119.11	3.74	1.28 1.80	175.16	55.09	39.12	H^{γ} 1.54,$H^{\delta a*}$ 0.72,$H^{\delta b*}$ 0.75,C^{γ} 24.14, $C^{\delta a}$ 20.25,$C^{\delta b}$ 22.74
100 Ala	7.38	119.27	4.00	1.37	176.90	52.38	14.95	
101 Ala	7.73	119.06	3.93	1.31	176.53	52.33	15.23	
102 Phe	8.47	118.98	3.93	3.13 2.86	177.43	58.20	36.45	$H^{\delta*}$ 6.53,$H^{\varepsilon*}$ 6.97,$C^{\delta*}$ 128.07,$C^{\varepsilon*}$ 126.97
103 Ile	8.62	122.32	3.49	1.96	175.31	61.72	35.10	$H^{\gamma 1a}$ 1.31,$H^{\gamma 1b}$ 1.75,$H^{\gamma 2*}$ 0.89,$H^{\delta 1*}$ 0.83, $C^{\gamma 1}$ 26.25,$C^{\gamma 2}$ 14.34,$C^{\delta 1}$ 10.12
104 Glu	7.85	117.51	4.08	1.95 1.95	173.99	54.18	27.57	$H^{\gamma a}$ 2.20,$H^{\gamma b}$ 2.41,C^{γ} 34.14
105 Gly	7.57	107.69	3.90 3.78		173.36	43.60		
106 Lys	8.05	116.56	4.13	1.96 1.55	173.99	53.99	30.84	$H^{\gamma a}$ 1.28,$H^{\gamma b}$ 1.28,$H^{\delta a}$ 1.56,$H^{\delta b}$ 1.56, $H^{\varepsilon a}$ 2.86,$H^{\varepsilon b}$ 2.86,C^{γ} 22.93,C^{δ} 26.19, C^{ε} 39.54
107 Leu	7.95	117.71	4.75	1.70 1.77	173.13	50.67	40.43	H^{γ} 0.87,$H^{\delta a*}$ 0.84,$H^{\delta b*}$ 0.84,C^{γ} 23.76, $C^{\delta a}$ 19.32,$C^{\delta b}$ 19.32
108 Arg	8.55	121.97	4.89	1.77 1.67	170.62	50.05	28.51	$H^{\gamma a}$ 1.60,$H^{\gamma b}$ 1.60,$H^{\delta a}$ 3.13,$H^{\delta b}$ 3.13, C^{γ} 24.09,C^{δ} 40.43
109 Pro			3.86	1.47 1.62	174.12	59.25	28.60	$H^{\gamma a}$ 1.80,$H^{\gamma b}$ 1.92,$H^{\delta a}$ 3.57,$H^{\delta b}$ 3.75, C^{γ} 23.98,C^{δ} 47.58
110 Ala	8.90	125.62	4.21	1.23	174.55	49.36	16.23	
111 Asp	8.49	120.23	4.63	2.59 2.59	173.29	51.07	39.23	
112 Thr	7.59	112.43	3.76	3.95	170.49	58.14	67.28	$H^{\gamma 2*}$ 0.94,$C^{\gamma 2}$ 18.18
113 Ala	7.82	124.08	3.85	1.11	174.55	49.64	16.45	
114 Asp	8.08	118.47	4.46	2.54 2.54	173.59	51.57	38.61	
115 Arg	8.04	120.59	4.18	1.69 1.69	173.52	52.98	28.26	$H^{\gamma a}$ 1.33,$H^{\gamma b}$ 1.39,$H^{\delta a}$ 2.86,$H^{\delta b}$ 2.91, C^{γ} 24.19,C^{δ} 40.22
116 Glu	8.28	120.12	4.16	1.95 1.84	174.09	53.83	27.53	$H^{\gamma a}$ 1.96,$H^{\gamma b}$ 2.16,C^{γ} 33.69
117 Gly	8.14	108.88	3.65 3.65		170.78	42.46		
118 His			4.48	2.92 2.79	171.48	53.37	28.11	
119 Trp	7.49	126.52	4.36	3.08 3.19	177.83	55.86	27.40	$N^{\varepsilon 1}$ 128.93,$H^{\delta 1}$ 7.06,$H^{\varepsilon 1}$ 10.04,$H^{\varepsilon 3}$ 7.51, $H^{\zeta 2}$ 7.33,$H^{\zeta 3}$ 6.97,$H^{\eta 2}$ 7.04,$C^{\delta 1}$ 124.20, $C^{\varepsilon 3}$ 118.87,$C^{\zeta 2}$ 111.56,$C^{\zeta 3}$ 118.26, $C^{\eta 2}$ 121.53

7.7 Published Paper

Biomol NMR Assign
DOI 10.1007/s12104-015-9648-5

ARTICLE

Resonance assignment of DVU2108 that is part of the Orange Protein complex in *Desulfovibrio vulgaris* Hildenborough

António J. Neca[1] · Rui Soares[1] · Marta S. P. Carepo[2] · Sofia R. Pauleta[1]

Received: 30 July 2015 / Accepted: 9 September 2015

Abstract We report the 94 % assignment of DVU2108, a protein belonging to the Orange Protein family, that in *Desulfovibrio vulgaris* Hildenborough forms a protein complex named the Orange Protein complex. This complex has been shown to be implicated in the cell division of this organism. DVU2108 is a conserved protein in anaerobic microorganisms and in *Desulfovibrio gigas* the homologous protein was isolated with a novel Mo–Cu cluster noncovalently attached to the polypeptide chain. However, the heterologously produced DVU2108 did not contain any bound metal. These assignments provide the means to characterize the interaction of DVU2108 with the proteins that form the Orange Protein complex using NMR methods.

Keywords NMR assignment · Cell division · Novel Mo–Cu cluster · *Desulfovibrio* · Anaerobic bacteria

Biological context

Desulfovibrio (D.) vulgaris Hildenborough is considered a model organism for the family of sulfate-reducing bacteria (SRB), and was the first to have its genome sequenced (Heidelberg et al. 2004). SRB are prokaryotes with the ability to use sulfur species as the final electron acceptor in anaerobic respiration producing hydrogen sulfide as an end product (Rabus et al. 2006). Being toxic, odorous and corrosive (Muyzer and Stams 2008), sulfide is an undesirable compound in engineering environments where sulfate and therefore SRB are found, indicating a role in microbial influenced corrosion (Heidelberg et al. 2004). Not only are they important for sulfur and carbon cycles but also have biotechnological interest as they can be used to remove sulfate and metals from waste water and sulfur dioxide from flue gas (Hansen 1994; Muyzer and Stams 2008).

Therefore, it is important to better understand this bacteria's lifestyle. Recent studies revealed a gene cluster encoding conserved hypothetical proteins that have been implicated in cellular division since its inactivation affects bacterial cell morphology (Fievet et al. 2011). One of these proteins, encoded by DVU2108, is a homologue of *D. gigas* Orange Protein (ORP) (with which shares 48 % sequence identity) (Bursakov et al. 2004; Carepo et al. 2014; Pauleta et al. 2007), that has a novel heterometalic sulfide cluster ($S_2MoS_2CuS_2MoS_2$) noncovalently bond to the polypeptide chain (George et al. 2000). The *D. vulgaris* Hildenborough ORP (DVU2108) has 119 residues and forms a protein complex with other proteins encoded by two divergent operons, *orp1* (*DVU2107-DVU2108-DVU2109*) and *orp2* (*DVU2103-DVU2104-DVU2105*), that are under the transcription regulation of DVU2106 (Fievet et al. 2011). This transcription regulator is a σ^{54}-dependent transcription factor that is down regulated in the presence of oxygen (Fievet et al. 2011).

A phylogenetic analysis shows that homologues of *DVU2108* genes tend to be clustered with *DVU2103* and *DVU2104* homologues in most archaea and strict anaerobic bacteria genomes (Scholten et al. 2007). The proteins that

✉ Sofia R. Pauleta
srp@fct.unl.pt

[1] UCIBIO, REQUIMTE, Departamento de Química, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa, 2829-516 Caparica, Portugal

[2] Laboratório de Bioinorgânica, Departamento de Química Orgânica e Inorgânica, Universidade Federal do Ceará, Cx. Postal 6021, Fortaleza 60440-900, Brazil

Published online: 15 September 2015

form this complex are metaloproteins containing most probably iron-sulfur clusters that might be oxygen sensitive. The structure of the ORP in the form of apo-protein has been determined for the homologue from *Thermotoga maritima* (TM1290) (Etezady-Esfarjani et al. 2004), *Methanobacterium thermoautotrophicum* (MTH1175) (Cort et al. 2000) and *D. gigas* (PDB ID 2WFB) (Najmudin et al. 2009; Pauleta et al. 2007), though insights into its function in these organisms has not yet been obtained.

Here, we report the ^{1}H, ^{15}N and ^{13}C assignment of the apo-form of DVU2108 from *D. vulgaris* Hildenborough. This data will be used to determine the metal-cluster binding site in the reconstituted protein and also to identify the DVU2108 surface that interacts with the other proteins that together form the ORP complex.

Methods and experiments

Protein expression and purification

The PCR isolated DVU2108 gene was cloned into the *NdeI*/*XhoI* restricted pET 21-c expression vector (Novagen) and the apo-DVU2108 was heterologously expressed in the *E. coli* strain, BL21(DE3) (Invitrogen). Expression of uniformly ^{15}N or $^{13}C/^{15}N$-labelled apo-DVU2108 was carried out by growing the cells in M9 minimum medium containing either 1 g/L $^{15}NH_4Cl$ or 1 g/L $^{15}NH_4Cl$ and 4 g/L [$^{13}C_6$]-glucose, respectively, as the sole nitrogen and carbon sources. The conditions for protein production were similar to the ones reported for apo-Orange Protein from *D. gigas* (Carepo et al. 2014; Pauleta et al. 2007), with the protein production being induced at an OD_{600nm} of 0.6 with 0.5 mM IPTG, and carried for 16 h at room temperature. The cells were harvested by centrifugation and the pellet was resuspended in 50 mM Tris–HCl at pH 7.6, containing protease inhibitors (Complete EDTA-free, Roche). The cell-free soluble extract was obtained by breaking the cells with a French-Press, the cell debris and membrane fraction were removed by low-speed centrifugation followed by an ultracentrifugation at 138,000*g*.

The purification of the DVU2108 was carried out in two-steps. The soluble cellular extract was diluted in cold water prior to being loaded onto a DEAE-FF (GE HealthCare, ∅ 26 × 100 mm), equilibrated with 10 mM Tris–HCl, pH 7.6. The unbound proteins were removed with the equilibration buffer and DVU2108 was eluted with a gradient between 0 and 500 mM NaCl in 10 mM Tris–HCl, pH 7.6. The fractions containing DVU2108, as judged by SDS-PAGE (12.5 % acrylamide Tris-Tricine buffer system), were concentrated using a Vivacell (Sartorius Stedim Biotech) apparatus over a YM5 exclusion membrane, at 4 °C. This concentrated fraction was then loaded onto a size-exclusion column Superdex 75 (GE HealthCare, ∅ 16 × 600 mm), equilibrated with 50 mM Tris–HCl, pH 7.6, 150 mM NaCl. The fractions containing the pure DVU2108 were pooled and concentrated as before, and its buffer was exchanged to 20 mM sodium phosphate, pH 7.0, 100 mM NaCl using a PD10 desalting column (GE HealthCare). The purified heterologously produced DVU2108 was stored at −80 °C until further use.

Protein concentration was determined using a modified version of the Lowry Method (Lowry et al. 1951) with BSA (Bovine Serum Albumin) as the standard protein. The yield of the heterologous expression was 50 mg/L of minimum medium, and the purified protein did not present any bound metal (Zn, Mo, Cu) as determined by ICP (REQUIMTE lab Analysis) and also by the inspection of its UV–visible spectrum.

NMR spectroscopy

For the NMR experiments, three protein samples were prepared, the unlabelled, the uniformly ^{15}N-labelled, and the $^{13}C/^{15}N$-labelled apo-DVU2108 following the procedure described above. Samples were 1.0 mM in protein concentration in 20 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM sodium azide, 1 mM DTT and 10 % $^{2}H_2O$.

NMR experiments were carried out at 298 K on Bruker AvanceIII 600 MHz spectrometer equipped with a TCI cryoprobe. The sequence assignments of the protein backbone resonances were obtained using 2D [^{15}N, ^{1}H]–HSQC, 3D HNHA, 3D HNCO, 3D HN(CA)CO, 3D CBCACONH, 3D CBCANH, 3D HBHA(CBCACO)NH, 3D (H)C(CCO)NH and 3D H(CCCO)NH spectra (Ferentz and Wagner 2000). For the side chain assignment 2D [^{13}C, ^{1}H]–HSQC and 3D (H)CCH-TOCSY experiments were performed (Ferentz and Wagner 2000). The ^{1}H spin systems of the aromatic rings of His residues were identified using a 2D ^{1}H-^{15}N HSQC type spectrum with a INEPT constant optimized for ^{2}J detection. The assignment of the aromatic side chains was performed with the analysis of the [^{13}C, ^{1}H]–HSQC together with 3D ^{13}C-edited NOESY (Ferentz and Wagner 2000).

Assignments and data deposition

The polypeptide chain of DVU2108 is composed of 119 residues with 5 being prolines. The sequence also contains 2 cysteine residues that are reduced due to the presence of DTT. The redox state of the side-chain was confirmed by the chemical shift of the Cβ that is in the expected range for a cysteine with -SH sidechain (Sharma and Rajarathnam

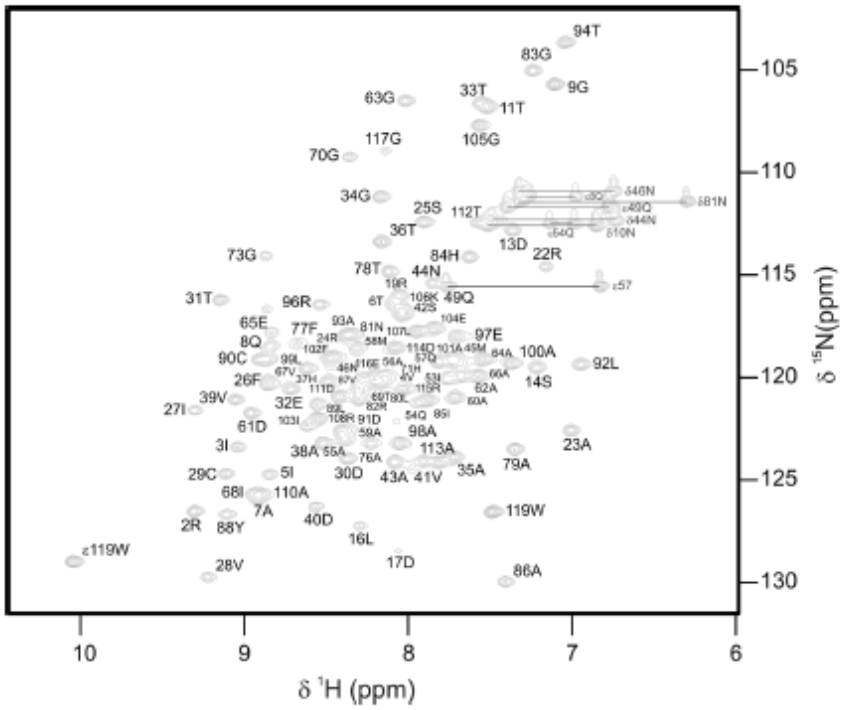

Fig. 1 [^{1}H, ^{15}N] HSQC spectrum of the 1 mM apo-DVU2108. The spectrum was acquired with a 600 MHz spectrometer equipped with a TCI cryoprobe, at 298 K. Figure was prepared in CCPN analysis

Fig. 2 Prediction of secondary structure of apo-DVU2108 based on chemical shifts of $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$ and $^{1}H_{\alpha}$. Figure was prepared in CCPN analysis

2000). These cysteine residues are not expected to form a disulfide bridge by the bioinformatic analysis of DVU2108 primary sequence and the residues are not even in close proximity in a structural model obtained using the coordinates of *D. gigas* apo-ORP deposited in the RCSB data bank with the PDB ID 2WFB.

Only 101 from the expected 113 amide proton resonances were observed in the $^{1}H–^{15}N$ HSQC (Fig. 1), probably due to exchange with the bulk solution, similarly to what was observed for the apo-form of the homologous protein from *D. gigas* (Carepo et al. 2014; Pauleta et al. 2007). After the complete sequential assignment of these NH resonances it was possible to locate the missing amide proton resonances in four regions of the protein: 4 residues at the N-terminal (Asn10, Leu12, Phe20-Gly21), 5 residues in a loop region that compromises the residues 47–52, two residues Val72-Lys75 and His118. Similarly, in other homologous proteins, the amide proton resonances of the residues of a loop located in this same region were poorly defined (Etezady-Esfarjani et al. 2004; Pauleta et al. 2007). The assignment was established for all residues, including the prolines and the first residue. The total extent of the assignment for the ^{1}H, ^{13}C and ^{15}N is 96.4, 98.0 and 86.4 %, respectively. The ^{1}H, ^{13}C and ^{15}N chemical shifts have been deposit in the BioMagResBank (http://www.bmrm.wisc.edu) under BMRB accession number 26618.

The secondary structure of DVU2108 was predicted based on the chemical shifts of $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$ and $^{1}H_{\alpha}$ (Fig. 2) using Talos+ (Shen et al. 2009) and CSI (Wishart and Sykes 1994), indicating that it is similar to the one of other ORP homologues (Cort et al. 2000; Etezady-Esfarjani et al. 2004; Pauleta et al. 2007).

Acknowledgments This work was supported by a research project from the Fundação para a Ciência e Tecnologia (FCT-ANR/BBB-MET/0023/2012) to SRP, which also supported the NMR spectrometer (RECI/BBB-BQB/0230/2012), which is part of the National NMR Network. MSPC would like to thank CAPES-BJT programme for financial support.

References

Bursakov SA et al (2004) Antagonists Mo and Cu in a heterometallic cluster present on a novel protein (orange protein) isolated from *Desulfovibrio gigas*. J Inorg Biochem 98:833–840. doi:10.1016/j.jinorgbio.2003.12.002

Carepo MS, Pauleta SR, Wedd AG, Moura JJ, Moura I (2014) Mo–Cu metal cluster formation and binding in an orange protein isolated from *Desulfovibrio gigas*. J Biol Inorg Chem 19:605–614. doi:10.1007/s00775-014-1107-8

Cort JR, Yee A, Edwards AM, Arrowsmith CH, Kennedy MA (2000) NMR structure determination and structure-based functional characterization of conserved hypothetical protein MTH1175 from *Methanobacterium thermoautotrophicum*. J Struct Funct Genomics 1:15–25

Etezady-Esfarjani T, Herrmann T, Peti W, Klock HE, Lesley SA, Wuthrich K (2004) NMR structure determination of the hypothetical protein TM1290 from *Thermotoga maritima* using automated NOESY analysis. J Biomol NMR 29:403–406. doi:10.1023/B:JNMR.0000032615.51536.1a5274592

Ferentz AE, Wagner G (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. Q Rev Biophys 33:29–65

Fievet A et al (2011) The anaerobe-specific orange protein complex of *Desulfovibrio vulgaris* hildenborough is encoded by two divergent operons coregulated by sigma54 and a cognate transcriptional regulator. J Bacteriol 193:3207–3219. doi:10.1128/JB.00044-11JB.00044-11

George GN et al (2000) A novel protein-bound copper–molybdenum cluster. J Am Chem Soc 122:8321–8322. doi:10.1021/Ja000955h

Hansen TA (1994) Metabolism of sulfate-reducing prokaryotes. Antonie Van Leeuwenhoek 66:165–185

Heidelberg JF et al (2004) The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. Nat Biotechnol 22:554–559. doi:10.1038/nbt959nbt959

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193:265–275

Muyzer G, Stams AJ (2008) The ecology and biotechnology of sulphate-reducing bacteria. Nat Rev Microbiol 6:441–454. doi:10.1038/nrmicro1892nrmicro1892

Najmudin S, Bonifacio C, Duarte AG, Pauleta SR, Moura I, Moura JJ, Romao MJ (2009) Crystallization and crystallographic analysis of the apo form of the orange protein (ORP) from *Desulfovibrio gigas*. Acta Crystallogr Sect F Struct Biol Cryst Commun 65:730–732. doi:10.1107/S1744309109023392S1744309109023392

Pauleta SR, Duarte AG, Carepo MS, Pereira AS, Tavares P, Moura I, Moura JJ (2007) NMR assignment of the apo-form of a *Desulfovibrio gigas* protein containing a novel Mo–Cu cluster. Biomol NMR Assign 1:81–83. doi:10.1007/s12104-007-9022-3

Rabus R, Hansen T, Widdel F (2006) Dissimilatory sulfate- and sulfur-reducing prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds) The prokaryotes. Springer, New York, pp 659–768. doi:10.1007/0-387-30742-7_22

Scholten JC, Culley DE, Brockman FJ, Wu G, Zhang W (2007) Evolution of the syntrophic interaction between *Desulfovibrio vulgaris* and *Methanosarcina barkeri*: involvement of an ancient horizontal gene transfer. Biochem Biophys Res Commun 352:48–54. doi:10.1016/j.bbrc.2006.10.164

Sharma D, Rajarathnam K (2000) 13C NMR chemical shifts can predict disulfide bond formation. J Biomol NMR 18:165–171

Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+ : a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223. doi:10.1007/s10858-009-9333-z

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

7.8 Proline conformations as a function of its 13Cβ and 13Cγ chemical shifts

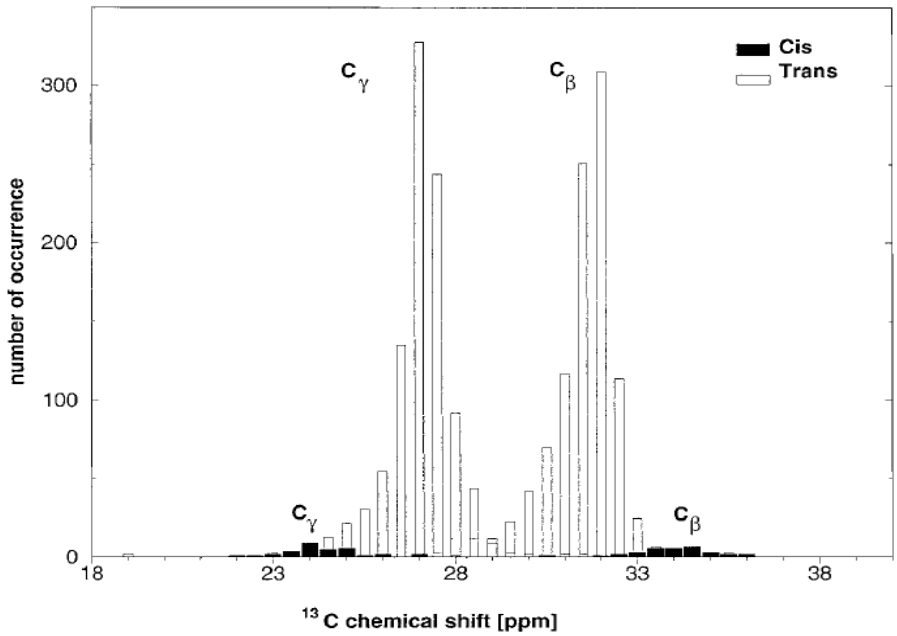

Figure 7.5 – Number of occurrence of Proline conformations as a function of its ^{13}Cβ and ^{13}Cγ chemical shifts, plotted in 0.5 ppm intervals. Reprinted from [45]

7.9 Distribution of cysteine Cβ chemical shifts according to their redox state

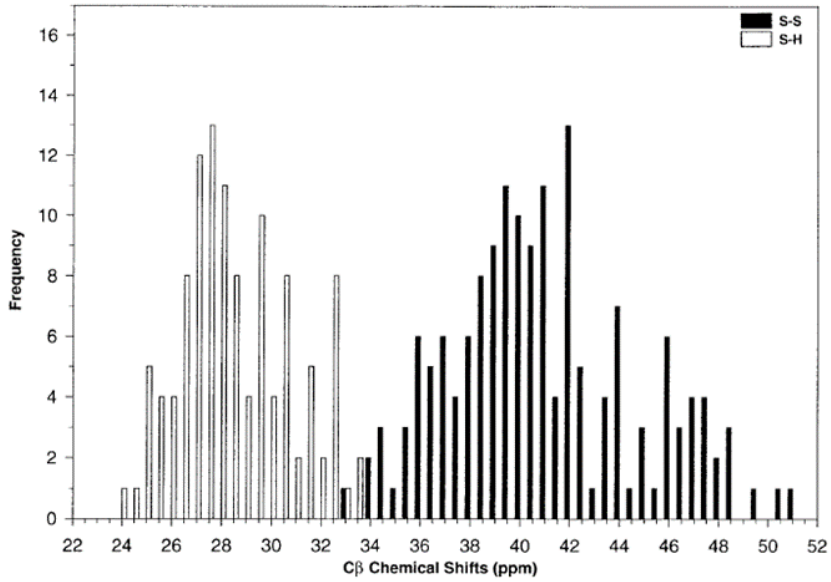

Figure 7.6 – Distribution of cysteine Cβ chemical shifts as a function of redox state. Reprinted from [46]

7.10 StrepDVU2108 purification contaminants

ORF number	Name
DVU0811	Dna K
DVU0847	Adenylsulfate reductase
DVU0881	Elongation factor G
DVU1834	Pyruvate carboxylase
DVU1976	GroEL
DVU2920	Elongation factor TU
DVU0958, DVU1303, DVU1306, DVU1309, DVU1320, DVU1328, DVU2518, DVU3150	Ribosomal proteins

Figure 7.7 – Recurrent copurified proteins defined as "contaminants" in streptrap purification. In black rectangles are the possible contaminating proteins found in this works purification of StrepDVU2108. Adapted from [16]

7.11 Protein sample spectra superimposed with MoCu-DVU2108 spectra

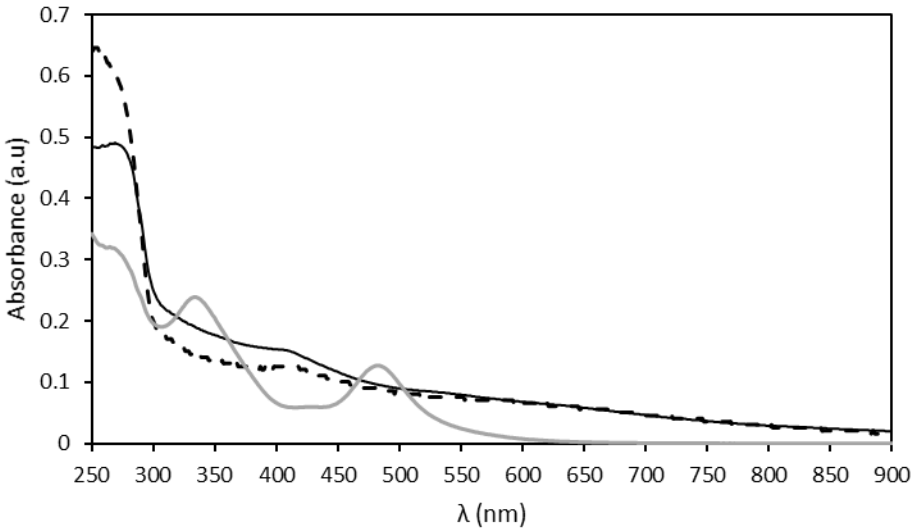

Figure 7.8 – UV-visible spectrum of protein sample after concentration from both growths superimposed with MoCu-DVU2108. Black full line – POSTGATE C medium supplemented with Mo and Cu (0.39 mg/ml); Black dashed line – POSTGATE C medium (030 mg/ml); Grey full line – MoCu-DVU2108 (± 0.15 mg/ml)