Cecilia de Agrela Pinto

The two-component system of a novel copper resistant operon of *Marinobacter hydrocarbonoclasticus*

Dissertation for the Master's Degree in
Structural and Functional Biochemistry

Supervisor: Dr. Sofia Rocha Pauleta
Assistant Researcher at Faculdade de Ciências e Tecnologia of Universidade Nova de Lisboa

Jury:

President: Prof. Dr. Ricardo Franco Tavares
Arguente: Prof. Dr. Manolis Matzapetakis
Vogal: Dr. Sofia Rocha Pauleta

March 2012

FACULDADE DE CIÊNCIAS E TECNOLO
UNIVERSIDADE NOVA DE L

Cecilia de Agrela Pinto

The two-component system of a novel copper resistant operon of *Marinobacter hydrocarbonoclasticus*

Supervisor: Dr. Sofia Rocha Pauleta
Assistant Researcher at Faculdade de Ciências e Tecnologia of
Universidade Nova de Lisboa

February 2012

The two-component system of a novel copper resistant operon of *Marinobacter hydrocarbonoclasticus*

ACKNOWLEDGEMENTS

The list of those who in some way influenced the final product, that is this thesis, is quite long. There are however five people without whom this work would have been impossible: Dr. Sofia Pauleta, Christina Pinto, my Parents and Vânia Reis. Each of you played a role far more important than I can even begin to explain.

Dr. Sofia Pauleta, your guidance throughout was invaluable, as were the words of encouragement. For your help in planning the experiments, letting me try a few of my own, and all the knowledge that was gained in between, thank you. You have my never ending gratitude for all that you have done, not only for the completion of this thesis, but also for me.

To my Sister, Christina, and parents, Daniel and Elsa, the last page, the symbolic "The End" is written for, and because of you. You have given me the greatest gifts, the ability to take care of myself, an education, a future. How does one say thank you for that?

To all my collegues, Luís, Nathalia, Irís, Jacopo, Catarina, among so many others who never failed to encourage when failure upon failure almost lead to insanity. The much needed laughter, the "sage" advice or sympathetic shoulders, and who could forget the endless slices of cake and cookies, are irreplaceable, as all of you played an essential role in helping me become a better scientist.

To Dr. Manolis Matzapetakis, without whom the CD experiments would have been impossible and who makes things seem simpler than they were while imparting an *nth* of his knowledge.

To my friends, your presence in my life has shaped and moulded me into the person I am today while your encouragement throughout these months has kept me going.

To all the above, and those I forgot, I would like to end with a simple: Thank you.

"Success is not measured by what you accomplish, but by the opposition you have encountered, and the courage with which you have maintained the struggle against overwhelming odds."

- Orison Swett Marden

ABSTRACT

The majority of bacterial heavy metal resistance systems are regulated by two-component signal transduction systems. Stimuli from the environment interact with the histidine kinase, which in turn activates the response regulator by phosphorylation. The effector domain of the response regulator then binds to DNA, eliciting the specific response.

Analysis of the *Marinobacter hydrocarbonoclasticus* genome revealed the presence of genes, *copXAB*, that code for proteins associated with copper response. The biochemical characterization of the two-component signal transduction system, *copSR*, is of interest due to the vital role it plays in the regulation of expression of the *copXAB* operon.

The genes that encode for the CopR and CopS_C (cytosolic sensor domain of CopS) proteins were heterologously expressed in *E.coli* and expression was optimized for the production of soluble protein using LB medium. Due to solubility problems, the genes that code for these proteins were cloned as hexahistidine or glutathione S-transferase fusion proteins. CopR and its domains were optimally expressed at 16°C for 16 and 3 h after induction, respectively, whilst CopS_C was expressed at 37°C during 3 h after induction. Proteins were purified using different chromatographic strategies, most of them using affinity chromatography. The yields of pure protein per liter of growth culture obtained after complete purification from the soluble cellular extract were: 0.14 to 0.23 mg/L for CopR; 0.42 mg/L, CopR_NHis$_6$; for the CopR_CHis$_6$ it was 0.16 mg/L and 4.2 mg/L of CopS_C.

The molecular mass of each protein was determined by gel filtration, 31 kDa for CopR, 17.5 kDa for CopR_NHis$_6$, 15.1 kDa for CopR_CHis$_6$ and 38.2 kDa for CopS_C. In the case of CopS_C there is the possibility that a dimer is formed, which should be evaluated. From the evaluation of disulfide bonds, using SDS PAGE and PAGE gels, all proteins or protein domains appeared to be monomers when in the presence of β-MEtOH. Circular dichroism evaluated the state of folding of the CopS_C and CopR proteins, which were shown to be folded in which the α-helix structures predominate. A model structure for CopR was also determined which agrees with this analysis. However, in the case of the CopR domains, the data obtained merely indicate folding, due to the low concentrations of the proteins.

Phosphorylation and electrophoresis mobility shift assays of the CopR protein were, for the most part, inconclusive. However, in the absence of BSA, formation of the CopR:DNA complex in a gel filtration column is observed, though requires additional evaluation.

Keywords

Copper tolerance, Response Regulator, Histidine Kinase, Two-component system, *Marinobacter*

RESUMO

A maioria dos sistemas de resistência a metais pesados são regulados por sistemas de transdução de sinal de dois componentes. Estímulos do meio ambiente interagem com a histidina cinase, que por sua vez activa, por fosforilação, o regulador de resposta. O domínio effector do regulador de resposta liga-se ao DNA, levando à resposta desejada.

Análise do genoma de *Marinobacter hydrocarbonoclasticus* revelou a presença de genes, *copXAB,* que codificam proteínas associadas à resistência ao cobre. A caracterização bioquímica do sistema de transdução de sinal de dois componentes, *copSR*, é de interesse devido ao papel vital que desempenha na regulação da expressão do operão *copXAB*.

Os genes que codificam para as proteínas CopR e CopS_C (domínio citosólico do sensor, CopS) foram heterologamente expressas em *E.coli* e expressão foi optimizado para a produção de proteína solúvel em meio LB. Devido aos problemas de solubilidade, os genes que codificam estas proteínas foram clonadas como proteínas de fusão hexa-histidina ou glutathiona S-transferase. CopR e os seus domínios foram optimamente expressas a 16°C durante 16 e 3 h após indução, respectivamente, enquanto CopS_C foi expresso a 37°C após 3 h de indução. As proteínas foram purificadas usando estratégias cromatográficas diferentes, a maioria dos quais usando cromatografia de afinidade. Os rendimentos obtidos de proteína pura por litro de cultura obtidos após purificação do extracto celular solúvel foram: 0.14 a 0.23 mg/L, CopR; 0.42 mg/L, CopR_NHis$_6$, para CopR_CHis$_6$ foi 0.16 mg/L e 4.2 mg/L de CopS_C.

A massa molecular das proteínas foram determinadas por filtração em gel, 31 kDa, CopR, 17.5 kDa CopR_NHis$_6$, 15.1 kDa CopR_CHis$_6$ e 38.2 kDa para CopS_C. No caso da CopS_C existe a possibilidade da formação de um dimero, que deve ser avaliada. Da avaliação de ligações dissulfureto, usando SDS-PAGE e PAGE, as proteínas ou domínios parecem como monómeros quando na presença de DTT. A análise do espectro de dicroismo circular indicou que a estrutura secundaria que predomina é α-hélice. Um modelo da estrutura da CopR foi determinada que esta de acordo com esta análise. No entanto, no caso dos domínios de CopR, os dados recolhido meramente indicam que as proteínas estão *folded*, devido a baixa concentração das proteínas.

Fosforilação e ensaios de mobilidade da proteína CopR foram, na sua maioria, inconclusivas. Contudo, na ausência de BSA, a formação do complexo CopR:DNA numa coluna de filtracao em gel é observado embora necessite de avaliação adicional.

CONTENTS

FIGURE INDEX

TABLE INDEX

ABBREVIATIONS

Abs.	Absorbance
ATP/ADP	Adenosine diphosphate/triphosphate
AU	Absorbance unit
BCA	Bicinchoninic Acid
bp, kb	Base pair, kilo base
BSA	Bovine Serum Albumin
c.v	Column volume
CA domain	Catalytic ATP-binding domain
CD	Circular Dichroism
cDNA	Complementary DNA
Da, kDa	Dalton, kiloDalton
DHp domain	Dimerization and Histidine phosphotransfer domain
DNA, dsDNA	Deoxyribonucleic Acid, double stranded Deoxyribonucleic Acid
DTT	DL-Dithiothreitol
EDTA	Ethylenediaminetetracetate
EMSA	Electrophoresis mobility shift assay
GST-tag	Glutathione S-transferase tag
h, min.	hour, minute
His_6-tag	Hexahistidine tag
HK	Histidine Kinase
HPt domain	Histidine phosphotransfer domain
IPTG	Isopropyl β-D-thiogalactopyranoside
Kav	Gel phase distribution coefficient
L	liter
LB	Luria-Bertani broth
LIC	Ligation Independent cloning
LMW	Low molecular weight
NMR	Nuclear Magnetic Resonance
Nr	Non-redundant
Ø	Diameter
OD	Optical density
ORF	Open Reading Frame
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction

pI	Isoelectric point
pro	Putative Promoter for *copSRXAB*
RNA	Ribonucleic Acid
Rpm	rotation per minute
RR	Response regulator
RT	Room temperature
SDS	Sodium dodecyl sulphate
TAE	Tris-acetate-EDTA buffer
TB	Tris borate buffer
TBE	Tris borate EDTA buffer
TCS	Two-component system
Tm	melting temperature
U	Unit
UV	Ultraviolet
Vo	Void volume
wHTH	winged Helix-Turn-Helix
X-P	Phosphorylated version of X (X – HK, HPt, CopS, RR, CopR, Asp)
β-MEtOH	β-mercaptoethanol
σ^{70}	Sigma 70
αCTD	C-terminal domain of the α-subunit

Bioinformatic programs

BLAST	Basic Local Alignment Search Tool
CDD	Conserved Domain Database
CELLO	Subcellular Localization Predictor
ExPASy	Expert Protein Analysis System
Gneg-mPLoc	Predicting subcellular localization of Gram-negative bacterial proteins
K_2D_2	Estimate of Protein Secondary Structure from Circular Dichroism Spectra
MONSTER	Minnesota protein Sequence annotation server
NCBI	National Center for Biotechnology information.
NetSurfP	Protein Surface Accessibility and Secondary Structure Predictions
P2CS	Protein two-component systems database
Pfam	Protein Families Database
Porter	Protein Secondary Structure Prediction
PROCHECK	Program to Check the Stereochemical Quality of Protein Structures
ProSA	Protein Structure Analysis
PSIPRED	Protein Structure Prediction Server

PSSpred	Protein Secondary Structure Prediction
SMART	Simple Modular Architecture Research Tool

Microorganisms

E.coli	*Escherichia coli*
M.	*Marinobacter*
M. hydro	*Marinobacter hydrocarbonoclasticus*
T. maritima	*Thermotoga maritima*

Amino acids

Ala, A	Alanine
Arg, R	Arginine
Asn, N	Asparagine
Asp, D	Aspartic acid
Cys, C	Cysteine
Gln, Q	Glutamine
Glu, E	Glutamic acid
Gly, G	Glycine
His, H	Histidine
Ile, I	Isoleucine
Leu, L	Leucine
Lys, K	Lysine
Met, M	Methionine
Phe, F	Phenylalanine
Pro, P	Proline
Ser, S	Serine
Thr, T	Threonine
Trp, W	Tryptophan
Tyr, Y	Tyrosine
Val, V	Valine

CHAPTER 1 – INTRODUCTION

1. INTRODUCTION

Bacteria inhabit a variety of habitats, from the intestinal gut of mammals to hydrothermal vents hundreds of meters below the ocean surface. It is therefore logical that these organisms encounter constant changes in their living conditions to which they must adapt, enabling them to survive. One such alteration is the overabundance of metal ions, which in many cases is fatal for the cell. The observation that high concentrations of metal ions are potentially lethal for bacteria has since been used to human advantage, for example in hospitals and in the food industry in disinfectants, as well as in agriculture as a component of bactericides and fungicides [1]. However, there are bacteria which survive in such conditions.

The intracellular metal content, or metallome, of bacteria reflects the metabolic requirement of metals in the cell. An analysis performed by Barton *et al.* [2], which confirmed an earlier elemental analysis done by Outten and O'Halloran [3], compared the molar abundance of trace metals in bacteria, yielding the following stoichiometric formula: $Fe_1Mn_{0.3}Zn_{0.26}Cu_{0.03}Co_{0.03}Mo_{0.03}$. The authors observed a degree of consistency across species, which reflects a specific requirement of trace metals in the metabolism and suggest that even with differences in cellular activities, due to variations in environmental settings, there are only minimal changes in trace metal content. These trace metals, in contrast to the alkali metals and alkaline earth metals are not free ions inside the cell but are bound in specific sites of proteins, enzymes and related compounds [2]. Not only does this demonstrate the relative use of these metals by the microorganisms, for common physiological processes, but also implies that the intracellular metallome is highly regulated.

The toxicity of free metal ions stems from the fact that heavy metal ions and metalloids can be very toxic to cells in a variety of ways, *e.g.* through binding to essential respiratory chain proteins, oxidative damage via the production of reactive oxygen species, which in turn causes damage to DNA, oxidation of polyunsaturated fatty acids in lipids, oxidation of amino acids in proteins and inactivation of specific enzymes by oxidation of their co-factors [4].

Due to the toxicity of trace metals at high levels, versus their inherent utility in cells, it is of paramount importance that metal homeostasis be maintained. To this end, microorganisms use a number of mechanisms to maintain the correct equilibrium, as well as, resistance mechanisms which include the uptake, chelation and extrusion of metals (for reviews see [5] and [6]). Such systems have already been identified in bacteria and involve metallothioneins, P-type ATPases, cation/proton antiporters and redox enzymes, which are in many cases regulated by two-component systems (TCS).

1.1 Copper in Biological Systems

Of all the metal ions, those that have an incomplete *d* orbital can be employed in the active site of enzymes. The trace metal content of most cells consists of Fe, Mn, Co, Ni, Cu, Mo, W, V (the redox active metals) and Zn (a nonredox active metal). Copper, for example, is used in most bacteria, yeast and humans, and has been identified as a component of the active site in over 30 human enzymes [7], such as lysyl oxidase, superoxide dismutase, tyrosinase, cytochrome *c* oxidase, superamine β-hydroxylase. Another class of copper proteins is the one belonging to the plastocyanin or azurin family, which act as electron carriers in bacteria. Another remark is that in bacteria most copper containing proteins are periplasmic or extracellular.

It is proposed that the incorporation of copper in biological systems began upon release of oxygen into the atmosphere some 2.7 billion years ago, causing the oxidation of the water insoluble Cu (I) oxide to soluble Cu (II) [8]. Consequently, copper-requiring proteins are now widely distributed among aerobic organisms. Copper (I), the cuprous oxide form, and Copper (II), the oxidized cupric form, are the most effective monovalent and divalent ions, respectively, for binding to organic molecules and possess almost equal tendencies to form complexes with many organic ligands [9].

Though Cu (I) is the preferred form of copper for handling inside the cell, both forms of copper usually bind to proteins via cysteine (Cys), methionine (Met), or histidine (His) amino acid side chains [10]. Cu (I) is a closed shell $3d^{10}$ transition metal ion and hence diamagnetic. Cu (I) prefers coordination numbers 2, 3 or 4 (tetrahedral geometry). As a soft Lewis acid, it is stabilized by soft ligands and tends to bind to soft bases, such as thiols, hydrides, alkyl groups, cyanide, and phosphines [11]. Cu (II) has a $3d^9$ configuration and is paramagnetic. As an intermediate Lewis acid, it forms complexes with additional ligands to Cu (I), including sulphate and nitrate [11]. The stable Cu (II)-N bonds are inert while the bonds with oxygen donor ligands are more labile.

The high redox potential of the Cu (I)/Cu (II) couple makes it particularly useful and adaptable for enzymes, and most copper enzymes have redox potentials in the 0.2 - 0.8 V range, depending on the coordination sphere of the metal, enabling them to directly oxidize substrates, such as superoxide, ascorbate, catechol, and phenolates [11]. These chemical properties of copper allow this metal to undergo redox cycling between these two forms, so in addition to being a part of redox-active enzymes, it can also act as an electron donor/acceptor in the electron transport chain. As well as being essential for metalloproteins, copper can also be incorporated in proteins by providing the appropriate coordination chemistry necessary to maintain higher order structure [7].

Copper proteins are involved in a variety of biological processes and as such deficiency in these enzymes, or alteration of their activities, often cause disease states or pathophysiological conditions. In humans, severe copper deregulation can lead to Wilson's

disease or Menkes syndrome [12]. Wilson's disease leads to an accumulation of copper mainly in the liver, brain and characteristically the cornea, while Menkes syndrome is characterized by a severe copper deficiency. Disturbed copper homeostasis has also been implicated in neurological diseases, such as the Alzheimer, prion and Parkinson diseases [13].

When copper is unbound within the cell, redox cycling means that when in the Cu (I) form it is extremely toxic, largely due to its ability to catalyze Fenton chemistry [1], causing production of highly reactive hydroxyl radicals that damage biomolecules, such as DNA, proteins, and lipids. Moreover it may bind to adventitious sites in these biomolecules causing the displacement of the native metal ions, as well as, alterations to their structure and/or function. Furthermore, copper appears to be much more toxic under anaerobic conditions, possibly due to the reduction of Cu (II) to Cu (I), which can diffuse through the cytoplasmatic membrane causing increased copper accumulation [14].

In addition to exposure from naturally occurring ores and during biogeochemical cycling of metals, anthropogenic release of metals has increased bacterial exposure to high levels of metals in some environments, for example from copper smelting, the use of copper sulphate as a fungicide in horticulture and in animal feed to promote weight gain. Such uses of copper provide suitable conditions in which copper resistance determinants can be selected and spread [5]. Bacteria must therefore achieve precise intracellular copper concentration to avoid copper-mediated toxicity whilst maintaining the supply for its metabolic necessities. To achieve this, copper-sensing transcriptional regulators must distinguish copper from other metal ions and in response to levels above or below a threshold value, trigger an appropriate physiological response, such as copper import, export, or detoxification. Cells may also restrict access to copper by channelling copper away from adventitious sites, for example using copper-chaperones to deliver copper to its targets [15].

1.2 Two-Component Systems

TCS regulation affect processes at several levels upon change in environmental conditions, primarily through transcriptional, post-transcriptional and post-translational regulation of gene expression, but also through a variety of protein–protein interactions. Essentially TCS regulate a wide variety of behaviours, including fundamental bacterial processes, such as metabolism and motility, in addition to specialized processes such as biofilm formation, virulence and development [16].

These systems are widespread and exist not only in nearly all prokaryotes and many archaea, but also in eukaryotes, such as plants, fungi and yeasts [17]. Figure 1.1 illustrates a prototypical TCS, which consists of a phosphotransfer reaction between two conserved components, a histidine kinase (HK) and a response regulator (RR).

Typically, extracellular stimuli are sensed by, and serve to modulate the activity of the HK. The HK then transfers a phosphoryl group from a conserved His residue in its kinase domain to the conserved Asp residue on the RR, a reaction which is catalyzed by the RR. Phosphotransfer to the RR results in activation of a downstream effector domain, which elicits the necessary physiological response [18]. For the most part, structural and functional features intrinsic to individual domains of HKs and RRs are conserved [19].

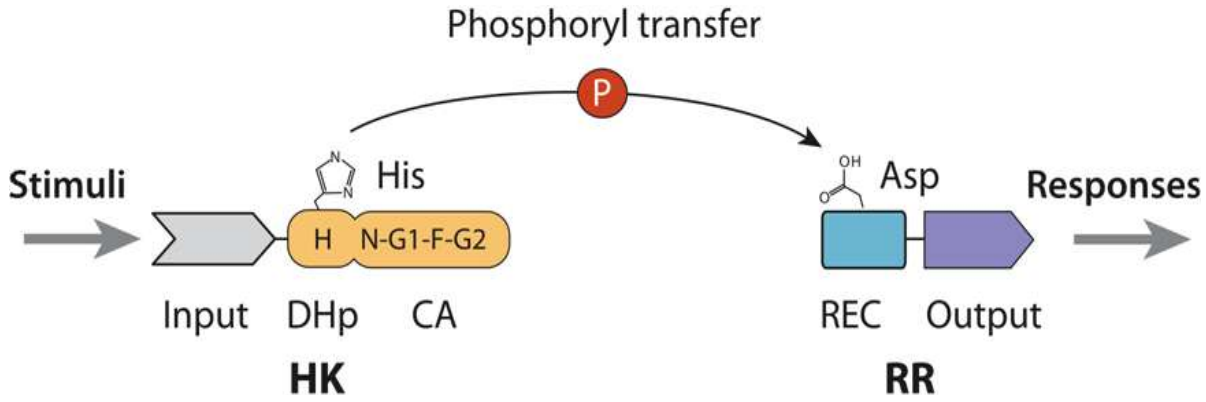


Figure 1.1 A prototypical two-component system transduction pathway which features a conserved phosphoryl transfer between the highly conserved kinase core (Dimerization and Histidine phosphotransfer - DHp and Catalytic ATP-binding - CA) and receiver (REC) domains to couple various input stimuli and output responses. The kinase core contains several homology boxes, H, N, G1, F and G2, which are conserved across histidine kinases (HKs). RR, response regulator. Figure from R. Gao, and A. M. Stock. 2009. "Biological insights from structures of two-component proteins." Annual review of microbiology 63: 133-154.

Auxiliary proteins, which regulate the activities of the HK or that influence the stability of RR phosphorylation, can augment this basic scheme, and many two-component pathways consist of more than two proteins [19]. Additionally, the phosphotransfer pathway can be expanded into phosphorelay pathways, with two or more phosphotransfers between multiple His and Asp containing proteins. However, it is more common for bacteria to exploit archetypal two-component signal transduction pathways, which couple environmental stimuli to adaptive responses, while with other organisms phosphotransfer schemes are more common [20].

For every general feature of HKs and RRs identified, exceptions have been found, which include even the conserved His and Asp residues that constitute the definitive His-Asp phosphotransfer reaction, and it is expected that the list of variations will undoubtedly increase as more systems are characterized. With approximately 80,000 TCS proteins identified from genome sequences [21] and approximately 300 structures of two-component proteins determined [22], it is unrealistic to identify and describe a single representative system. TCS proteins are therefore modular and versatile, likely reflecting their widespread abundance in bacterial signaling.

The understanding of these fundamental mechanisms, however, is greatly derived from studies performed on a subset of the identified proteins and the obtained knowledge is therefore restrictive [19]. Despite the quantity of structures solved, the ways in which domains interact with each other and the regulatory mechanisms resulting from these domain arrangements often differ.

1.2.1 Histidine Kinases

Members of this family vary in size from < 40 kDa to > 200 kDa; the larger of which consist of five or six functionally and structurally unique domains [23]. The extreme diversity observed in the HK protein family is generated from the simple combination of sensing, catalytic, and auxiliary domains.

Despite this diversity, HKs can be roughly divided into two classes: orthodox and hybrid kinases [23]. For the sake of brevity, and due to the fact that the system of focus contains an orthodox HK (see section 1.4), only orthodox kinases are described in this introduction. It is important, however, to note that hybrid kinases are more elaborate, containing multiple phosphodonors and phosphoacceptor sites. In addition, instead of promoting a single phosphoryl transfer, hybrid kinases make use of multistep phosphorelay schemes, the overall complexity of which allows different checks and inputs to be integrated into the signalling pathway [23].

Orthodox HKs are characterized by three functional domains [17]: a variable N-terminal sensing domain, a His-containing phosphotransfer and dimerization domain (DHp), and a C-terminal adenosine triphosphate-binding (ATP-binding) kinase domain (CA), as shown in Figure 1.1. The modular nature of these proteins enables the structural architecture of individual HKs to be adapted to the specific needs of the signaling system, coupling input signals to the appropriate output responses.

Stimuli are sensed by the N-terminal sensing domains, either directly or indirectly, through protein-protein interactions with auxiliary signal transduction proteins. The prototypical HK is a homodimeric integral membrane protein, in which the sensor domain is formed by an extracellular loop enclosed between two membrane-spanning segments, and the kinase core, follows the last transmembrane segment and is localized in the cytoplasm (Figure 1.2 a).

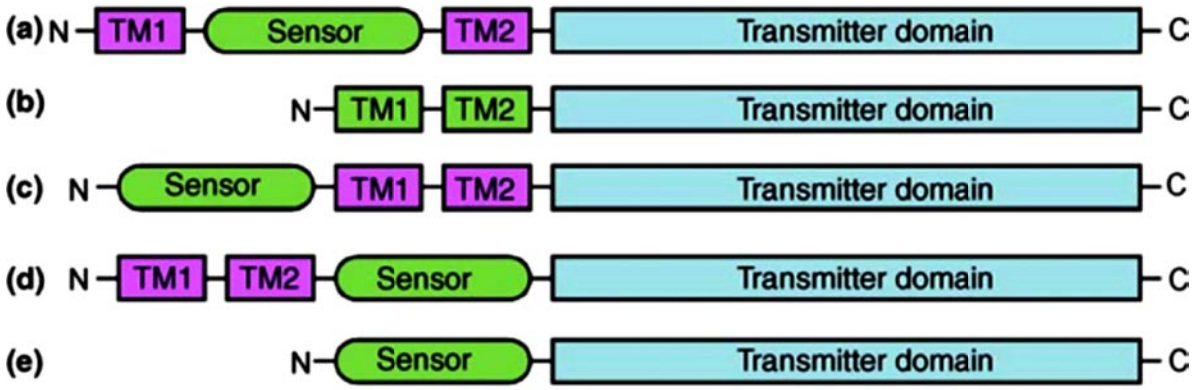


Figure 1.2 Histidine kinase domain organization. A schematic showing some basic examples of sensor domain organization in context with the full length histidine kinase receptor: **(a)** The sensor domain is usually formed as a folded extracellular loop between two transmembrane segments in a membrane-spanning histidine kinase. **(b)** The sensor domain may be embedded within the membrane, composed from transmembrane helices as if having a truncated extracellular loop truncated to a stub. **(c)** The cytoplasmic sensor domain may be located at the N-terminal of two or more transmembrane segments, in a membrane-anchored histidine kinase. **(d)** The cytoplasmic sensor domain may be located at the C-terminal of two or more transmembrane segments in a membrane-anchored histidine kinase. **(e)** The cytoplasmic sensor domain may reside N-terminal to the transmitter domain in a soluble histidine kinase that lacks transmembrane segments. Transmembrane segments designated TM1 and TM2 may be composed of more than one transmembrane segment, but always an odd number. From Cheung, J., & Hendrickson, W. A. (2010). Sensor Domains of Two-Component Regulatory Systems. Current Opinion in Microbiology, 13(2), 116-23.

As shown in Figure 1.2, HKs can also be classified based on their membrane topologies [24]. The largest group of HKs is represented by the classic HKs, in which the extracytoplasmic sensory domain includes various domain families, and perception of extracellular stimuli is transduced across the membrane to regulate the kinase/phosphatase activities (Figure 1.2 a, c). The second group has multiple (2 to 20) membrane-spanning helices but does not contain an apparent extracellular domain (Figure 1.2 b, d). The stimuli sensed by these HKs are believed to be membrane-associated stimuli (*e.g.*, cell envelope integrity) derived from membrane-integral components and ion gradients. The third group of HKs features a cytoplasmic sensory domain responsible for sensing diffused and internal stimuli (Figure 1.2 e). Some HKs may combine multiple features of these topologies, integrating signals from different input domains, extracellular or cytoplasmic, for a concerted response to complex environments [19].

The diverse range of sensory domains, in terms of size and amino acid sequence, enables HKs to sense a wide variety of stimuli, such as small molecules, ions, light, turgor pressure, cell envelope stress, redox potential and electrochemical gradients. Sensory domains usually share little primary sequence similarity, thus supporting the idea that they have been designed for specific ligand/stimulus interactions [25]. Even within metal ion sensing domains, it is essential that the protein be able to differentiate between structurally similar metal ions [26], such as Mn^{2+}, Fe^{2+}, Co^{2+}, Ni^{2+}, Cu^{2+} and Zn^{2+}, which have ionic radii between 61 (low spin – Fe^{2+}) and 83 (high spin – Mn^{2+}) pm, in addition to all carrying a double positive charge [27].

Within the sub-family of classic HK, with an extracellular sensor domain, these can be further divided into three structural classes [24]; mixed α/β folds (the PDC sensors), all α-folds and the predicted sensor domains that show similar fold to periplasmic binding proteins, exemplified by a single representative HK29 [28] (Figure 1.3).

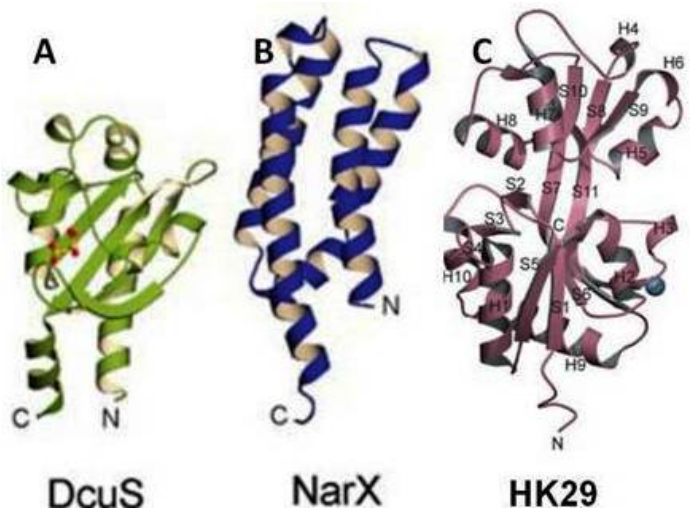


Figure 1.3 Ribbon diagrams of **(A)** PDC sensor DcuS, **(B)** all-α sensor NarX and **(C)** periplasmic binding protein fold represented by sensor HK29. Non-protein moieties are shown in ball-and-stick representation with carbon in yellow, oxygen in red, and calcium in magenta. Figures A and B adapted from Cheung, Jonah, and Wayne A. Hendrickson (2010). "Sensor domains of two-component regulatory systems." Current Opinion in Microbiology 13(2): 116-123. Figure C adapted from Cheung, J., Le-Khac, M., & Hendrickson, W. A. (2009). "Crystal structure of a histidine kinase sensor domain with similarity to periplasmic binding proteins." Proteins, 77(1), 235-41.

PDC sensors, which appear to be the most prevalent, are distinguished by a central five-stranded anti-parallel β-sheet scaffold that is flanked by α-helices on either side, beginning with a long N-terminal helix and often ending with a short C-terminal helix [24].

Despite great advances in structural characterization of signal perception mechanisms in several HKs, the structures of sensory domains and the identity of exact stimuli still remain unknown for most HKs. For others, structural information such as that obtained by Cheung, J. *et al*. for the sensing domain of PhoQ from *E.coli* [29], may help to define extracellular receptor/ligand interactions.

The unifying structural feature of the HK family is the characteristic kinase core composed of a dimerization and Histidine Phosphotransfer (DHp) domain, referred to as the His kinase A domain clan in Pfam [22] (Figure 1.4) and an ATP/ADP-binding phosphotransfer or catalytic domain (CA), also known as HATPase c in the Pfam database [22].

The kinase core is approximately 350 amino acids in length and is responsible for binding ATP and directing kinase transphosphorylation. The highly conserved kinase core has a unique fold, distinct from that of the Ser/Thr/Tyr kinase superfamily. Since, in two-component

signaling systems, a different phosphorylation scheme predominates with the formation of phosphoramidates instead of the phosphoesters seen in Ser/Thr/Tyr protein kinases [30,31].

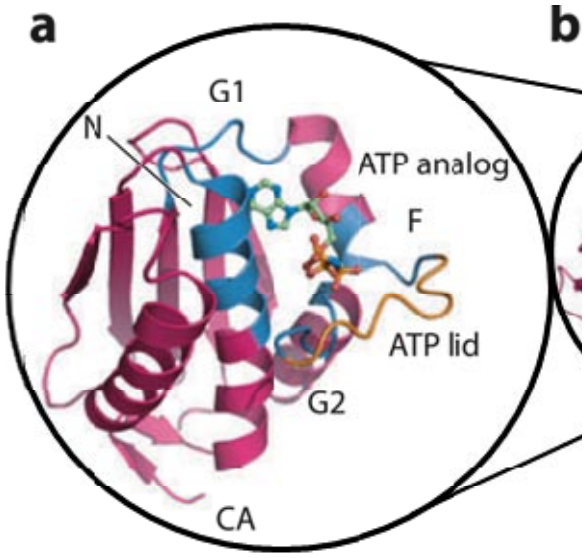


Figure 1.4 Structure of the kinase core. Crystal structure of **(a)** the CA domain of *E. coli* PhoQ and **(b)** the entire kinase core of *T. maritima* HK853 (PDB IDs: 1IDO and 2C2A, respectively). Homology boxes crucial for ATP binding are shown in blue in the PhoQ CA structure. A flexible ATP lid (orange in a) covers the bound ATP analogue, AMPPNP. HK853 is dimeric with one monomer shown in orange and pink and the other in grey. Figure adapted from R. Gao, and A. M. Stock (2009) "Biological insights from structures of two-component proteins." Annual Review of Microbiology 63: 133-154.

There are five conserved amino acid motifs present in both eukaryotic and prokaryotic HKs [32,33]. The conserved His substrate is the central feature in the H box, whereas the N, G1, F, and G2 boxes define the nucleotide binding cleft (Figure 1.4 a). In most HKs, the H box is part of the dimerization domain. However, for some proteins, the conserved His is located at the far N-terminus of the protein, in a separate HPt domain. The N, G1, F and G2 boxes are usually contiguous, but the spacing between these motifs is somewhat varied. The ATP-binding site of the HKs consists of conserved residues from the N, G1, F, and G2 boxes [34]. In both the CheA and EnvZ structures, this binding pocket is a highly flexible region of the protein, which may reflect conformational changes that accompany ATP binding [23].

HKs function as dimers, and in the HKs characterized to date a transphosphorylation mechanism has been demonstrated in which the CA domain from one subunit of the dimer phosphorylates the specific His residue on the DHp domain from the other subunit. However, the diversity of domain arrangements that are postulated to exist in different HKs might enable *cis*-phosphorylation in some HKs [35,36].

The high-energy N-P bond which is formed between the phosphoryl group and the imidazole ring of His is relatively unstable, making phosphoHis residues more suitable as phosphotransfer intermediates than as stoichiometrically phosphorylated sites for protein

recognition. The chemical instability of phosphoHis hinders the detection of HK phosphorylation by conventional phosphoamino acid analysis at acidic pH because phosphoHis is labile to acid, though resistant to base treatment [19].

Given the intracellular ATP/ADP ratio, only a small percentage of the HK population exists in a phosphorylated state. Thus, it is the flux of phosphoryl groups, rather than stoichiometric phosphorylation, that is relevant to the function of HKs [25]. Phosphorylation of HKs usually occurs spontaneously in kinase core domains, and the phosphoryl group can be transferred to the cognate RR, if this protein is present [19].

In addition to kinase activity, many HKs possess a phosphatase activity toward their cognate phosphorylated RRs. The phosphatase activity resides in the DHp domain, and interaction between the DHp and ATP binding domains greatly affects the activity. The capability of decreasing the level of RR phosphorylation through rapid dephosphorylation provides a swift mechanism to shut down the signaling pathway and is suggested to be important for suppressing cross-talk from nonspecific phosphorylation [37,38]. It seems that this bifunctionality is a general feature of most prototypical HKs.

1.2.2 Response Regulators

RRs share the common N-terminal phosphoacceptor domain (receiver domains are designated REC in the SMART protein domain database [39] or Response_Reg in the Pfam protein family database [40]) but differ in their C-terminal effector (output) domains. Structural studies have revealed that, despite common structures and mechanisms of function within individual domains, a range of interactions between receiver and effector domains confer great diversity in regulatory strategies, optimizing individual RRs for the specific regulatory needs of different signaling systems [41].

Conserved features of receiver domain amino acid sequence correlate with structure and hence function. The regulatory domains of RRs have three activities:

i) First, they interact with phosphorylated HKs and catalyze transfer of a phosphoryl group to one of their own Asp residues,

ii) Second, they catalyze autodephosphorylation, and

iii) Third, they regulate the activities of their associated effector domains in a phosphorylation-dependent manner.

The conserved receiver domain can also be found within hybrid HKs or as isolated proteins within phosphorelay pathways. In these contexts, the receiver domains are not physically connected to effector domains and play no direct role in regulating effector domain functions [17].

Receiver domains typically adopt an $(\alpha\beta)_5$ fold consisting of approximately 125 residues. This fold is conserved in at least ten different RR receiver domains for which structures have been determined [25], and consists of a central five-stranded parallel β-sheet flanked on both faces by amphipathic helices (Figure 1.5). The overall structure reflects patterns within the receiver domain amino acid sequences that are sufficiently conserved to be manually recognized. Three runs of consecutive hydrophobic residues correspond to the three central β-strands in the core of the receiver domain. The active site consists of a characteristic quintet of highly conserved residues located at the C-terminal end of these three strands [42]. Three Asp residues bind a divalent metal ion that is essential for all phosphoryl group chemistry, one of which is also the phosphorylation site. The Lys and Thr/Ser residues are critical for signal transduction, as is a more distant Phe/Tyr residue [42] (Figure 1.5).

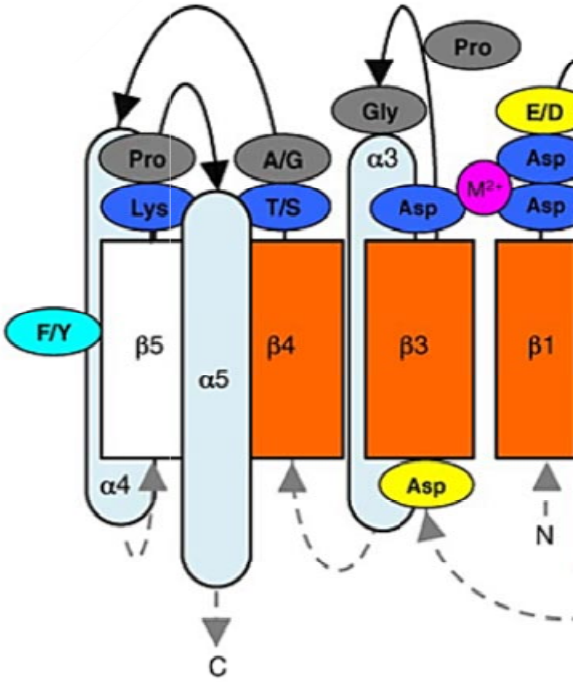


Figure 1.5 Schematic diagram of the relationship between receiver domain amino acid sequence and basic structural elements as described in the text, with the active site viewed from the side. Five α-helices surround a parallel five-stranded β-sheet. Loops connecting strands and helices are shown as black solid lines on the active site side of the domain and grey dashed lines on the opposite side, with arrowheads indicating N-terminal to C-terminal direction. Orange indicates pattern of conserved hydrophobic residues on the central β-strands and faces of three α-helices. The highly conserved residues of the active site quintet are in blue, with the moderately conserved aromatic residue in cyan. Divalent metal ion in magenta. Residues presumed to be strongly conserved for structural reasons are in grey. Frequently conserved acidic residues of unknown function are in yellow. Figure from Bourret, R. B. (2010) "Receiver domain structure and function in response regulator proteins." Current opinion in microbiology 13(2): 142-149.

The mentioned divalent cation (Fig. 1.5), adjacent to the phosphorylation site is necessary to add or remove phosphoryl groups in the receiver domain, whether the reactions are mediated by the receiver alone or also involve a HK, a HPt domain, or a phosphatase [42].

The six coordination positions of the metal ion are occupied by the three conserved Asp residues (one acting through a water molecule), a backbone carbonyl group, and two water molecules [42]. Due to the fact that multiple coordination positions are occupied by water and hence not restricted in space, the binding site can accommodate metal ions of different sizes. For many receivers, the preferred metal ion is Mg^{2+}, which presumably forms a pentavalent phosphorous intermediate through an SN_2 reaction, but other divalent cations, particularly Mn^{2+}, support different reactions to varying extents. During phosphorylation an oxygen atom in the phosphoryl group replaces one of the waters in the metal ion coordination sphere [42]. The determinants responsible for variation between receiver domains in metal ion binding affinity and specificity have not been characterized. However, hydrophobic amino acids in the position immediately C-terminal to the conserved pair of metal-binding acidic residues have been proposed to strengthen nearby electrostatic interactions with the metal ion, whereas hydrophilic residues may weaken binding [43].

Although often referred to as receiver domains, the conserved domains of RRs are not passive partners in phosphotransfer from HKs or HPt domains. Rather, they actively catalyze phosphotransfer, as evidenced by the observation that small molecules, such as acetyl phosphate and phosphoramidate can serve as phosphodonors in vitro reactions [44,45].

Small molecule phosphodonors for RRs fall into two categories: phosphoamidates (*e.g.* phosphoamidate, monophosphoimidazole) and acyl phosphates (*e.g.* acetyl phosphate, carbamoyl phosphate) [42]. The phospho-His of HKs is a phosphoramidate, so chemically similar small molecule phosphoamidates are useful tools for the investigation of RR autophosphorylation, but are not known to be physiologically relevant themselves. In contrast, autophosphorylation with acyl phosphates may be a physiologically relevant means to connect metabolic state to two-component signal transduction. The fundamental reaction chemistry is presumably similar whether a RR catalyses phosphorylation using a small molecule, an HK, or an HPt. Although HKs or HPt domains are not required for catalysis, rates of phosphotransfer from these phosphoHis-containing proteins are much greater than from small molecules [46,47].

RRs also possess autophosphatase activity which limits the lifetime of their active states. The lifetime of phospho-Asp within RRs varies significantly, typically ranging from seconds to hours [48,49]. Autodephosphorylation presumably proceeds through a pentavalent phosphorus intermediate, similar to that proposed for phosphotransfer, with water serving as the nucleophile in the axial position [17]. In a few cases the RR can also stabilize the phospho-Asp, increasing the half-life significantly beyond the half-life of a typical acyl-phosphate (Acyl-phosphates are rapidly hydrolyzed in both acidic and alkaline conditions but have half-lives of several hours in neutral conditions [23]). The lifetimes of different RRs appear well correlated with their physiological functions and other regulatory strategies of the system. However, the short lifetimes of phosphorylated RRs have, until recently, been a hindrance to a full characterization of this state.

A recent breakthrough showed that beryllofluoride (BeF^{3-}), a non-covalent mimic of a phosphoryl group, binds to the active-site Asp residue of many RRs, creating a stable analog of the phosphorylated protein [50], when analyzed by crystallography [51–53], however when analyzed in solution beryllofluoride as an analog for phosphorylation is inappropriate [54].

Structures of four different phosphorylated or otherwise activated RR regulatory domains have been determined [53,55,56]. These structures have confirmed the long-postulated notion that phosphorylation of the active-site Asp is associated with an altered conformation of the receiver domain. Structural differences between the unphosphorylated and phosphorylated receiver domains map to a relatively large surface involving α_3, β_4, α_4, β_5, α_5 and adjacent loops, with backbone displacements ranging from 1–7 Å (Figure 1.6). The spatial extent and magnitude of the conformational changes associated with phosphorylation vary significantly in the different RRs that have been characterized [52,55–58] .

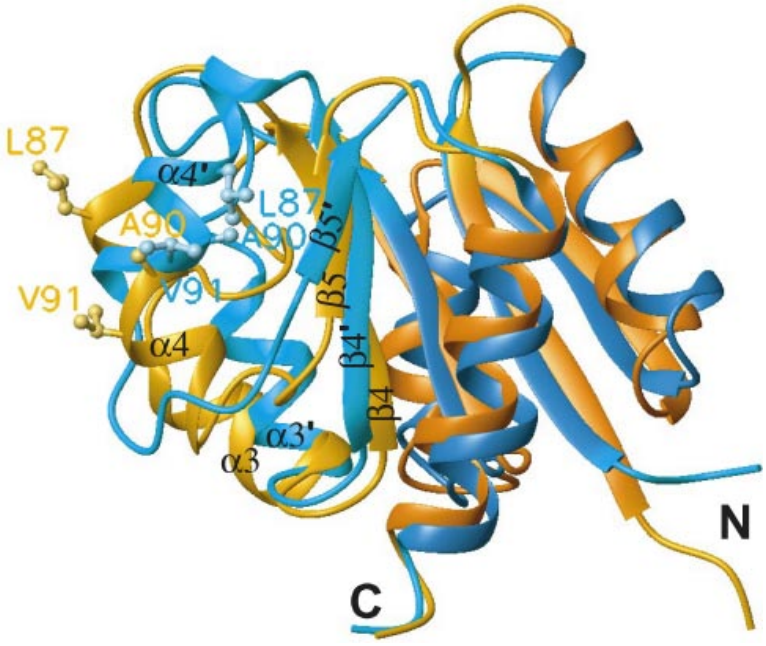


Figure 1.6 Molecular switch upon phosphorylation of D54 in NtrCr. Superimposed ribbon structures of NtrCr-P (yellow/orange) and NtrCr (cyan/blue), generated using the conformer closest to the mean structure for each. The molecule has been rotated by 20° about the vertical axis to emphasize the regions of structural differences between the two forms. Structures were superimposed using residues 4-9, 14-53 and 108-121, which are indicated in darker colours (orange and blue). The regions of greatest difference are highlighted (yellow and cyan, the switch area) and the corresponding secondary structure elements are labelled, with the prime indicating the unphosphorylated form. Upon phosphorylation of D54, β-strands 4 and 5 and α-helices 3 and 4 tilt to the left - away from the active site. In addition, a register shift by about two amino-acid residues from the N to the C terminus and a rotation by about 100° about the helical axis are induced in helix 4 upon phosphorylation. The rotation results in a change in orientation of the hydrophobic side chains in helix 4 from the inside to the outside. Figure from Kern, D., Volkman, B. F., Luginbühl, P., Nohaile, M. J., Kustu, S., & Wemmer, D. E. (1999). “Structure of a Transiently Phosphorylated Switch” in Bacterial Signal Transduction. Nature, 402(6764), 894-8.

Importantly, the surface that undergoes structural alteration upon phosphorylation corresponds to surfaces that have been identified as *loci* for phosphorylation-regulated protein–protein interactions in different RRs [59]. These and other data support the current view that RR regulatory domains function as generic on–off switch modules. The domains can exist in two distinct structural states with phosphorylation modulating the equilibrium between the two conformations. This provides a very simple and adaptable mechanism for regulation of RR activity. Surfaces of the regulatory domain that have altered structures in the two different conformations can be exploited for protein–protein interactions that regulate effector domain function [60].

Receiver domains are remarkably versatile, in that the same basic structure can be paired with and apparently regulate more than 60 different output domains. Effector domains are diverse with respect to both structure and function and, not surprisingly, many different mechanisms are used for regulation of their activities by the receiver domains. The majority of RRs are transcription factors with effector domains that can be divided into three major subfamilies based on the homology of their DNA-binding domains: the OmpR/PhoB winged-helix-turn-helix (wHTH) domains, the NarL/FixJ four-helix domains and the NtrC ATPase-coupled transcription factors.

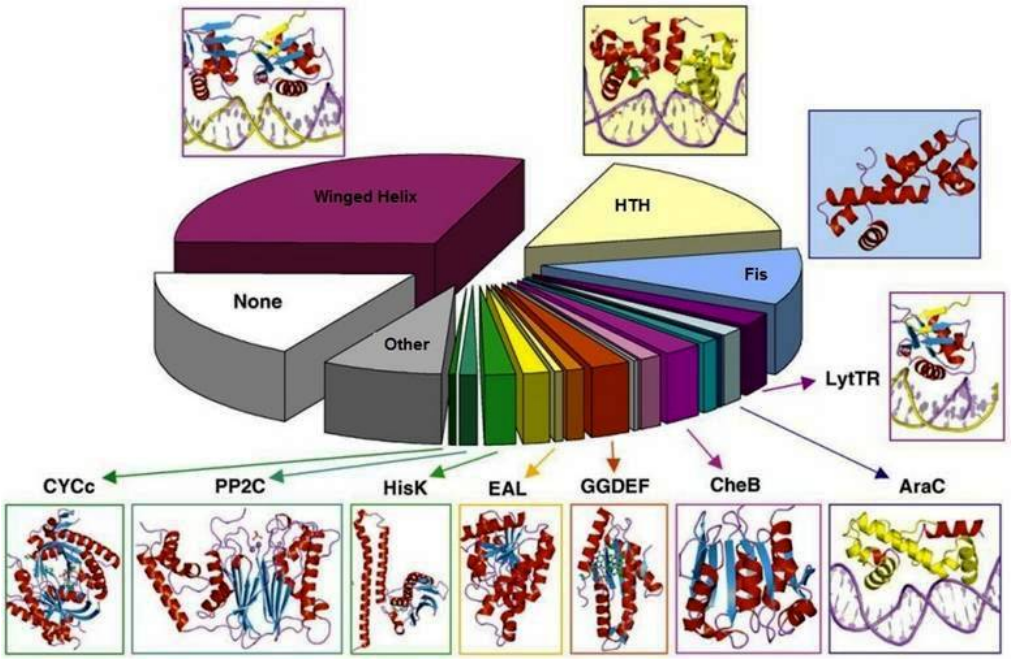


Figure 1.7 Distribution of common output domains of bacterial response regulators. Aggregate data for all bacterial species. The sector for Fis domain includes response regulators of the NtrC and ActR families (no structure of a Fis-DNA complex is currently available). Unlabelled output domains (clockwise from AraC) are as follows: Spo0A (dark blue), ANTAR (cyan), CheW (pink), CheC (white), GGDEF + EAL (orange), and HD-GYP (yellow). The domain structures are as indicated in the tables presented in the article and the structures are taken from Protein Data Bank in Europe. Image from Galperin, M. Y. (2010). "Diversity of structure and function of response regulator output domains." Current Opinion in Microbiology, 13 (2), 150-9.

Structures of effector domains, and multidomain RRs, of representative members of each of these subfamilies have been determined [61–63], Figure 1.7, but direct structural information about how these structural perturbations control the activity of the attached effector domains is limited by the relatively small number of structures of intact multidomain response regulator proteins available.

The OmpR subfamily effector domain fold (Figure 1.8) is characterized by three α-helices flanked on two sides by antiparallel β-sheets, an N-terminal four-stranded β-sheet and a C-terminal hairpin that interacts with a short β-strand connecting helices α1 and α2 to form a three-stranded β-sheet.

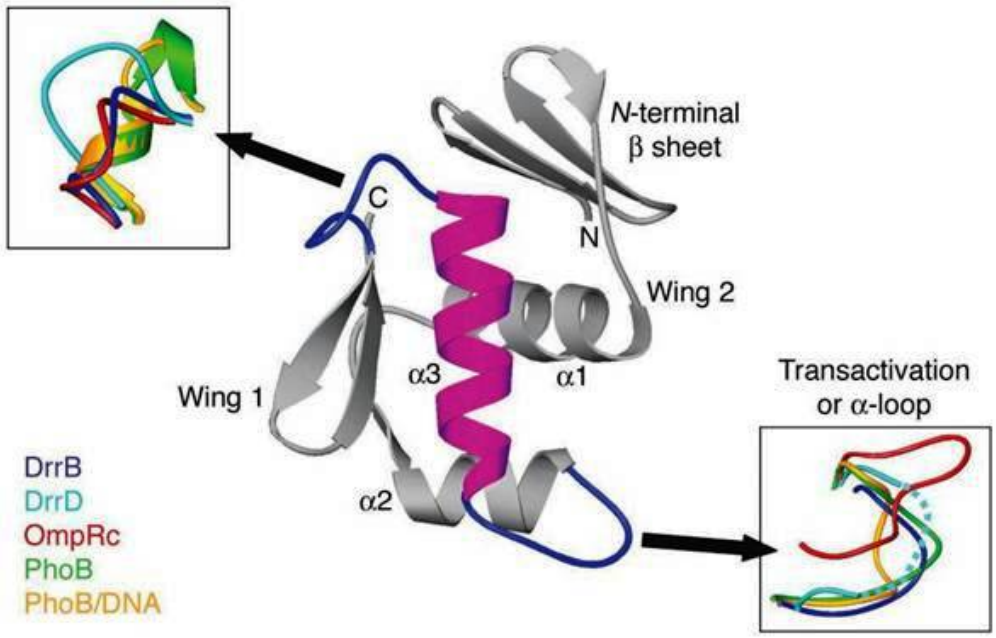


Figure 1.8 OmpR/PhoB subfamily winged-helix domains. Grey, ribbon diagram of a representative winged-helix domain that defines the OmpR/PhoB subfamily of response regulators; magenta, recognition helix that binds to the major groove of DNA. Effector domains of DrrB, DrrD (PDB accession code 1KGS), OmpR (1OPC), PhoB (1GXQ), and PhoB bound to DNA (1GXP) were superimposed based on the core region of the protein, defined as the three α-helices and six β-strands. In all proteins, the two loops flanking the recognition helix exhibit considerable conformational flexibility, as shown in the insets. The α loop of DrrD (dotted line) is disordered. Image from Robinson, V. L., Wu, T., & Stock, A. M. (2003). "Structural Analysis of the Domain Interface in DrrB , a Response Regulator of the OmpR/PhoB Subfamily." Journal of Bacteriology, 185 (14), 4186-94.

These transcription factors bind DNA and interact productively with RNA polymerase to activate transcription. The two functions, DNA-binding and transcriptional activation, have been localized within the 100 amino acid DNA-binding domain that characterizes members of this family. The fold of the DNA-binding domain represents a variation on the winged-helix DNA-binding motif. The α2-loop-α3 region forms a helix-turn-helix motif. Helix α3 corresponds to the recognition helix that interacts with the major groove of DNA and helix α2 corresponds to the

positioning helix. The loop connecting β6 and β7 of the C-terminal hairpin has been shown in other winged-helix proteins to interact with the minor groove of DNA and has been termed recognition wing, W1 [61]. In some wHTH proteins, a second wing, W2, that is positioned adjacent to the recognition helix opposite to wing W1, contacts the minor groove [64].

OmpR-like response regulators have significantly shorter linkers than typical multidomain response regulators, with lengths ranging from 5 to 21 residues. 29% of the OmpR homologs have linkers of just five amino acids and 93% of the OmpR homologs have linkers shorter than 13 amino acids [61].

The presence of the N-terminal β-sheet formed by β1 to β4 and an unusually large loop, designated α-loop, that connects the positioning helix α2 to the recognition helix α3 are conserved and distinguishing features of the OmpR subfamily, not present in other winged-helix proteins [65]. The disorder observed in this region, as implied by the previously solved structures of the OmpR/PhoB subfamily of DNA binding domains, supports the notion that conformational mobility in the α-loop is important in enabling proper interactions with RNA polymerase for transcription initiation [66].

Amino acid residues that form the hydrophobic core of the OmpR subfamily effector domain are conserved as hydrophobic residues, indicating that these domains can be interpreted in terms of a common fold [65]. The most variable regions of the domain coincide with the loop that precedes the recognition helix, the α-loop, and the loop that follows the recognition helix (α3). Amino acids that form part of helices α1 and α2, as well as the hydrophobic core residues, are highly conserved. The most conserved residues form part of helix α3 and recognition wing, W1. The recognition helix is positioned perpendicular to the major groove and contacts particular DNA bases, thus determining the sequence specificity of binding. Therefore, the recognition helix must include amino acids that are critical for determining DNA-binding specificity [67].

The four-stranded antiparallel β-sheet, which directly follows the linker, is an integral part of the OmpR DNA-binding fold. The low conservation might reflect the importance of this region for a specific function in each homolog, such as an interaction with the regulatory domain.

Structures of intact multidomain RRs reveal that RRs use different subsets of the regulatory domain surface for phosphorylation-dependent regulatory interactions (Figure 1.9). Thus, although RRs are fundamentally similar in the design of their receiver domains, there is significant versatility in the way these domains are used to regulate effector activity.

Biochemical studies have indicated many different strategies for regulation of effector domains by regulatory domains in RRs, including inhibition of effector domains by unphosphorylated regulatory domains [63], allosteric activation of effector domains by phosphorylated regulatory domains [68], dimerization [57] or higher-order oligomerization of phosphorylated RRs and interaction of RRs with heterologous partners [53].

Phosphorylation-induced dimerization plays a key role in the regulation of DNA-binding RRs, with all OmpR/PhoB subfamily members forming dimers or higher-order oligomers to dramatically improve their binding to the palindromic or direct-repeat chromosomal binding sites. This mechanism has been demonstrated, Figure 1.9, for the RRs of the OmpR, NarL, LytR, and PrrA families [61,63,69,70].

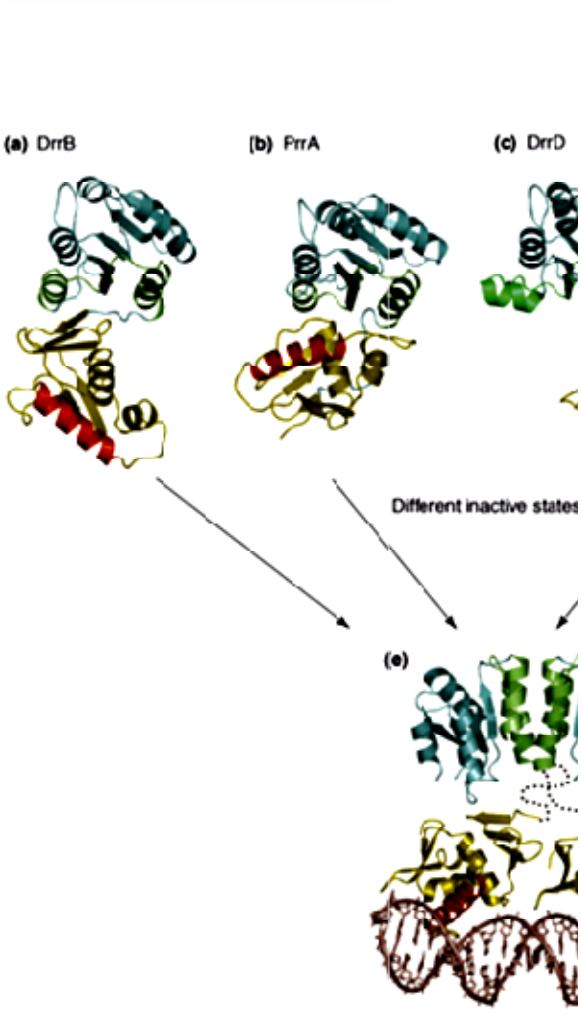


Figure 1.9 Inactive and active domain arrangements in OmpR/PhoB subfamily members. OmpR/PhoB RRs have different domain orientations in the inactive state, yet all assume a common active state. Structures are available for three inactive full-length multidomain RRs: **(a)** *T. maritima* DrrB (PDB code: 1P2F), **(b)** *Mycobacterium tuberculosis* PrrA (PDB code: 1YS6) and **(c)** *T. maritima* DrrD (PDB code: 1KGS). Domain arrangements in a fourth RR, **(d)** *E. coli* PhoB, can be modelled from structures of the isolated receiver domain dimer (PDB code: 1B00) and the isolated DNA-binding domain (PDB code: 1GXQ). The orientation of the DNA-binding domains (bracketed) relative to the receiver domains in PhoB is unknown, but the short linkers that connect the domains (broken lines) restrict placement of the DNA-binding domains to diagonal positions across the receiver domain dimer. Although no structures of active multidomain OmpR/PhoB RRs have been determined, an active state can be readily envisioned **(e)**. The model depicted here is constructed from structures of the isolated active receiver domain dimer of PhoB (PDB code: 1ZES) and the complex of PhoB DNA-binding domains bound to target DNA (PDB code: 1GXP). Receiver domains are shown in teal with α4–β5–α5 faces highlighted in green; DNA-binding domains are shown in gold with recognition helices highlighted in red. Figure from R. Gao, Timothy R. Mack, and A. M. Stock (2007) “Bacterial Response Regulators: Versatile Regulatory Strategies from Common Domains.” Trends in Biochemical Sciences 32 (5): 225-234.

It is expected that phosphorylation-induced activation mechanisms are more complex than simple interdomain repositioning for many, if not most, RR transcription regulators. Phosphorylation is postulated to expose a hydrophobic surface on helix α4 leading to protein-protein interactions and consequently formation of the active dimer [71]. Remarkably, even very similar RRs appear to differ in their activation mechanisms. For example, PhoB has been shown to form dimers upon phosphorylation in solution in the absence of DNA, whereas OmpR fails to appreciably form dimers after phosphorylation unless a DNA binding target is present [72].

All members of the OmpR family for which DNA recognition sites have been determined appear to bind to direct repeat DNA sequences [73]. However, there is variation in the arrangement of sites, both with respect to the number of recognition sites and the spacing between them. When bound to the DNA, each monomer recognizes a half-site of 2-fold symmetric DNA sequences. It has been proposed that OmpR binds to DNA in a cooperative manner, as OmpR homodimers have not been detected in solution, while phosphorylation of the regulatory domain effectively enhances its DNA-binding activity and vice-versa [74]. Site-directed DNA cleavage at the OmpR binding sites led to the proposal that OmpR binds asymmetrically as a tandemly arranged dimer with the recognition helix of each monomer in the major groove[75].

There is significant variation in the degree of sequence similarity among the different DNA-binding domains of OmpR homologs. However, low level of sequence conservation in this family is not entirely unexpected. Each DNA-binding domain must be involved in specific interactions, such as recognition of a particular DNA sequence, interaction with the regulatory domain and with either the α subunit or the σ^{70} subunit of the RNA polymerase.

1.3 Copper Response Regulators

Several promoters of copper (and silver)-inducible genes are preceded by a palindromic sequence, termed the “cop-box”, which are the sites of DNA binding of a cognate response regulator. It should be appreciated that palindromic sequences have the property that they are identical on each strand of DNA, allowing a transcription factor to bind to either strand, which essentially doubles their concentration compared to non-palindromic sequences that need to be recognized by a transcription factor on only one strand of DNA [76]. This palindromic sequence is transcription factor specific, being that each transcription factor should bind preferentially to its own cop box sequence. The specificity of this interaction depends on the RR and is determined by interactions between the recognition helix and the DNA bases.

However, within the same sub family of RRs the sequence must be somewhat conserved as even the variable recognition helix has some conserved regions. In bacteria, copper-responsive regulators can be divided into four main groups, according to their structural features and mechanisms of action.

1. The best studied copper-responsive regulators belong to the MerR family which includes CueR. There response regulators mediate copper-induced transcription in *Escherichia coli* [77,78], *Bacillus subtilis* [79], *Cupriavidus metallidurans* [80] and *Pseudomonas aeruginosa* [81]. CueR regulators are dimeric, two-domain proteins containing a wHTH DNA-binding domain and a copper-binding effector domain separated by a linker element. Their DNA-binding domains are related to those found in regulators of the MerR superfamily, which bind to their target promoter both in the presence and in the absence of their effector, and act on the topology of the promoter to induce transcription in the presence of copper or repress it in its absence [78,81,82].

2. The second type of bacterial copper-responsive regulators includes CopY of *Enterococcus hirae* and CopY (formerly CopR) of *Lactococcus lactis* [83,84]. A common conserved motif of the sequence TACAnnTGTA appears to be the binding site for CopY like copper-responsive repressors [85]. Two other such sequences are the cop-box of copper- and silver-inducible promoters of *E.coli* K 12, AnnTnACAnnAnTGTnATnAnnCnGT, and of *P. syringae,* CnAAGCTTACnGAAATGTAAT. Both are metal-fist-type repressors belonging to the CopY/TcrY family, and bind to DNA in the form of dimers stabilized by zinc. In the presence of copper, zinc is displaced and the repressor is released from the promoter, allowing transcription [84].

3. The third family of copper-responsive repressors, DUF156, is represented by CsoR of *M. tuberculosis* [86], its only characterized member. The Cu-specific regulator CsoR senses high intracellular copper concentrations and activates (or derepresses) transcription of the copper export ATPase CtpV which also might be responsible for exporting excess copper from the cytoplasm [87].

4. Finally, bacteria also possess copper-responsive two-component systems, which include the well-characterized CusRS and PcoRS found in *E. coli* [88,89], and PcoRS and CopRS of *Pseudomonas syringae* [90], which are required for the activation of copper-inducible genes. The CusR cop box, located in the -42 to -67 region upstream of the *cusC* start codon, has a sequence of TTTTTACTGTTTTAACAGTAAAA. The PcoR copper box is AGnTTACAnAAnTGTAATnAnnnnG s identified in *E.coli*.

1.4 *Marinobacter aquaeolei* VT8: The Case Study

Marinobacter (M.) aquaeolei or *M. hydrocarbonoclasticus* [91] is a gram-negative marine bacterium that thrives in oil platforms and can degrade hydrocarbons and some components of crude oil [92]. They are facultative anaerobes, non spore forming, halofilic, mesofilic and with mobility conferred by a single polar flagellum. This gammaproteobacteria, from the alteromonadaceae family, exhibits, in addition to a chromosomal genome (EMBL accession number CP000514) of 4327 kbp, two additional plasmids, pMAQU01 (EMBL accession number CP000515) and pMAQU02 (EMBL accession number CP000516) of 240 and 213 kbp respectively.

A bioinformatic search in the *M. aquaeolei* VT8 genome (NCBI RefSeq accession number: NC_008740) for proteins responsible for resistance to copper revealed the presence of a unique and still uncharacterized, *copSRXAB* determinant (Figure 1.10 a), which consists of two proposed operons, *copSR* and *copXAB*, which are proposed to be regulated by the two-component signal transduction system, CopRS (Figure 1.10 b).

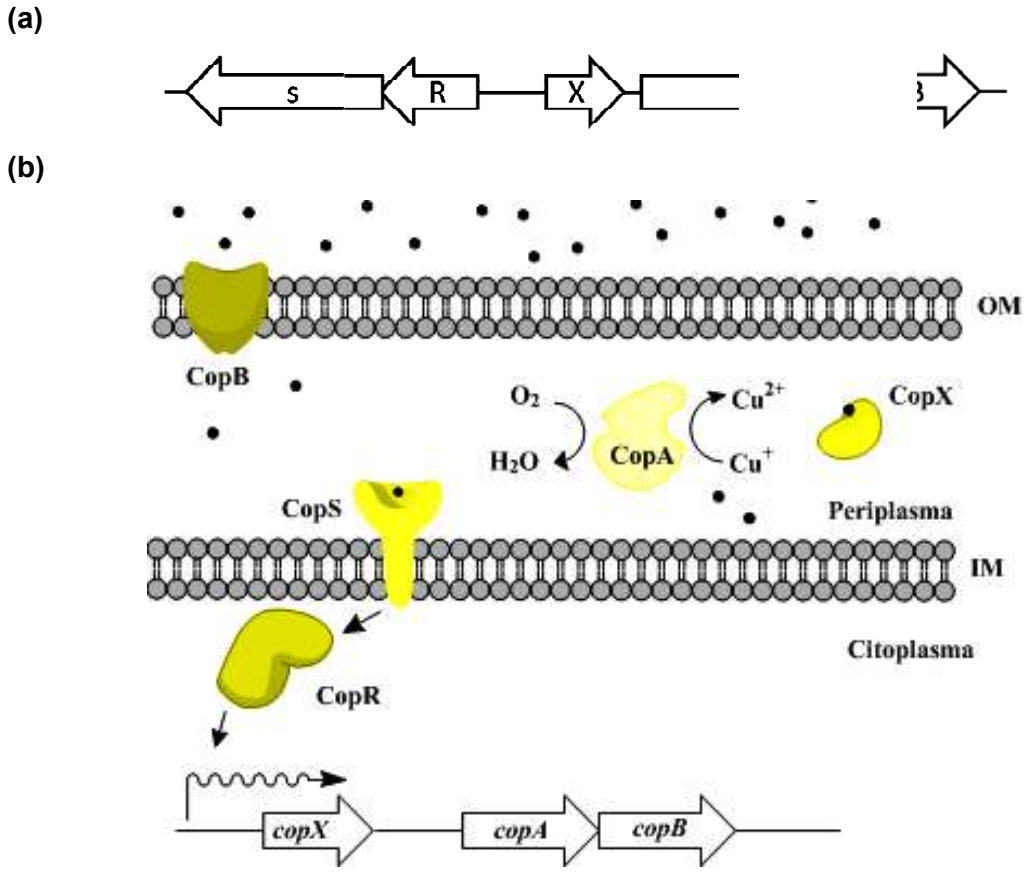


Figure 1.10 (a) Genomic organization of the proposed copper resistance operon *copSRXAB* in *M. aquaeolei* VT8. **(b)** Schematic representation of the copper resistance system CopSRXAB. CopS and CopR represent the proposed orthodox histidine kinase and the response regulator of the OmpR subfamily, respectively. CopA is a proposed multicopper oxidase whilst the role of CopB and CopX in copper resistance are currently under investigation. Figure 1.10.b was drawn using the ChemBioOffice 2010.

The inherent interest in studying the *M.aquaeolei* VT8 *copS/R* system lies not only in the limited knowledge existent for the mechanistic activation and interaction of the proteins, but also in the resistance operon itself, which is novel. Whilst CopA is proposed to be a multicopper oxidase, the functions of CopX and CopB are still unknown and under investigation. Prior work showed that the CopX protein, annotated as coding for a copper-binding protein, is different from other small copper chaperones usually associated with copper resistance systems. In fact, CopX appears to be a type 1 copper binding protein, much larger than expected (Nóbrega, F.L. 2010 Masters Thesis, FCT-UNL).

1.5 Importance of Studying these Systems

TCSs are the most abundant multistep signaling pathways in nature and have been extensively studied since their discovery approximately 20 years ago [19]. Apart from their inherent interest as a fundamental signaling strategy, it is expected that understanding these systems will be useful for the development of antibiotics that disrupt essential systems in pathogenic organisms or might be exploited for the beneficial use of bacteria to enhance agriculture, environmental remediation and nitrogen fixation [41].

Moreover, because two-component proteins are extremely important in microorganisms but appear to be rare or absent in animals, they represent logical targets for the development of novel antimicrobial agents [66]. Identification of small molecule and peptide inhibitors of HKs and RRs suggests that this avenue of pursuit could indeed be successful. However, little progress has been made in developing drugs that inhibit two-component proteins, which, for several different reasons, have proven to be difficult targets [66].

It is hoped that a more focused approach, aimed at specific RR proteins, might be beneficial. The general difficulty in crystallizing multidomain RRs, to date, may suggest that, interdomain flexibility, is the rule rather than the exception for this superfamily [93]. Additional biochemical, biophysical, and structural studies will be required to further advance the understanding of the coupling between phosphorylation and activation for multi-domain RRs.

1.6 Objectives of the Present Work

The main objective of this work was to obtain a pure, soluble protein sample for assays and NMR analysis. In order to achieve this, the following tasks were necessary:

- Construction of expression systems for the hexahistidine tagged response regulator (CopR) and C-terminal domain of the histidine kinase (CopS). This involved gene isolation and genetic engineering to clone the gene into the appropriate expression vectors;

- Heterologous expression in *E.coli* and optimization of the protein expression and purification;
- Basic biochemical characterization of the isolated proteins;
- Protein structure determination.

The second aim is to identify the promoter region of *copXAB* and obtain insights into its regulation. Towards these objectives, the following tasks were necessary:

- Phosphorylation of CopR and DNA-protein interaction assays for the CopR protein, using electrophoretic mobility shift assay (EMSA) as well as optimization of the phosphorylation assay and EMSA assay conditions;
- Protein-protein (Phosphotransfer) CopR/S interaction assays.

CHAPTER 2 – MATERIALS AND METHODS

2. MATERIALS AND METHODS

2.1 Bioinformatic Analysis

2.1.1 Analysis of *M. aquaeolei* VT8 Genome

The presence of TCS' in the complete sequenced chromosomal genome of *M. aquaeolei* VT8 was identified with the graphic interface P2CS [16]. A complete list of the 42 and 54 histidine kinase and response regulator proteins, respectively, identified in this bacteria, is presented in Appendix C.1, Tables 6.3 and 6.4. These nonredundant (nr) protein sequences were used to illustrate the relative abundance of these systems and the protein classes in the genome, Figure 3.1 A and B in Chapter 3 - Results and Discussion.

2.1.2 Analysis of the Primary Sequence and Protein Homology

The ExPASy [94] "Translate" tool was used to obtain the protein sequence of the ORFs maqu_0123 and maqu_0124 (Frames of translation presented in Appendix C.2, Figures 6.5 and 6.6) and the "ProtParam" tool was used to analyze those primary sequences with the default settings (Amino acid distribution presented in Appendix C.3, Table 6.5).

The amino acid sequences of the proposed HK and RR of *M. aquaeolei* VT8, maqu_0123 and maqu_0124, respectively, were aligned with other homologous two-component systems. This study was performed with the online software Basic Local Alignment Search Tool (BLAST) using the graphic interface Protein BLAST [95]. The selected database was the nr protein sequence and the algorithm used was blastp (protein-protein BLAST), with no cut-off or additional limitations.

The resulting sequences, as well as queries, were aligned using the ClustalW (version 2.0) multiple alignment software [96]. Accession numbers for all the protein sequences used in this study are presented in Appendix B4, Table 6.2.

PSORTb [97] (version 3.0.2.), sub**CEL**lular **LO**calization predictor – CELLO [98] (version 2.5.) and Gneg-mPLoc [99] (version 2.0.) were used in conjunction to determine with the highest probability the cellular localization of the CopS and CopR proteins. The resulting predictions, as well as accuracy of each predictor are presented in Appendix C.4, Table 6.6.

The protein sequences of CopS and CopR were additionally subject to analysis by SMART [39], Simple Modular Architecture Research Tool, which identifies conserved domains, or families. These domains were corroborated by the Conserved Domain Database [100] (CDD) and Pfam [40] bioinformatic programs and adjustments of a few amino acids were made in order to average the three predictions, Appendix C.5, Tables 6.7 and 6.8.

2.1.3 Secondary Structure Prediction and Protein Structural Model

Secondary structure predictions were performed using the following bioinformatic servers: Porter prediction server [101], for the predictions presented in Chapter 3 - Results and Discussion; PSIPRED [102], NetSurfP [103] (version 1.1), Protein Secondary Structure PREDiction – PSSpred [104], Minnesota protein sequence annotation server - MONSTER [105], and JPred 3 [106] were used to corroborate/refute this prediction, Appendix C.6, Figures 6.7, 6.8 and 6.9.

In the case of the CopR protein and respective domains, in addition to the secondary structure analysis preformed as indicated, structure modelling was also performed, which is an independent form of analysing the secondary structure prediction when the primary structure is restricted to the template parameters.

In the case of CopR, given the limited set, 6 (PBD IDs: 1KGS, 1P2F, 3R0J, 1YS6, 2GWR, 2OQR), of full-length structures of proteins from the OmpR/PhoB family, a manual retrieval of their amino acid sequences from the PDB database was performed, followed by a multiple sequence alignment using the ClustalW alignment software [96].

The structure of an OmpR/PhoB homolog (DrrD) of *T. maritima* (PDB entry 1KGS, resolution of 1.5 Å, R value = 0.181, Rfree value = 0.210, presented in Appendix C.8, Figure 6.9) was used as a template for homology modelling (36% of sequence identity). The comparative 3-D structure model of *M. hydrocarbonoclasticus* 617 CopR was generated by homology modelling with the online software, SWISS-model [107] from the ExPASy database [94]. The generated 3-D structural model was visualised by Chimera software (version 1.5.3) [108] after utilizing the default setting for energy minimization (100 steps with 0.02 Å increments and 10 update interval).

In the case of the CopR domains, CopR_N and CopR_C, the homology models were generated using the online software, SWISS-model and the automated setting. Visualization of the resulting models was done as for CopR.

The structure verification server, PROCHECK [109], was used for evaluation of the generated models, namely the Ramachandran plot, which provides the details of various aspects of the model. The energy of the generated models was calculated using ProSA [110], comparing it with the potential of the mean force derived from a large set of known 3-D protein structures available in the database.

2.2. Molecular Biology Techniques

2.2.1 Genomic DNA Extraction

For heterologous expression of the *copS* and *copR* genes from *M. hydrocarbonoclasticus* 617, genomic DNA was extracted using the distilled water boiling

method [111] from bacterial cells grown as described by Sambrook J. *et al.*[112]. The resulting DNA suspension was stored at -20°C until further use.

2.2.2 PCR Amplification and Primers

Primers for the full-length *copR* gene, as well as the C-terminal domains of *copR* and *copS, copR_C* and *copS_C* respectively, and the N-terminal domain of *copR, copR_N,* are presented in Table 2.1. These oligonucleotides were designed based on the sequence information available for *M. aquaeolei* VT8 (Accession no. NC_008740) and were produced by Sigma, Genosys. Table 2.1. also lists the primers, *proF* and *proR*, for a fragment proposed to contain the promoter region of the *copXAB* operon.

Table 2.1 Primers used throughout the work for amplification of the required DNA fragments. *pro* – putative fragment which contains the promoter, DNA region between *copR* and *copX* genes. [a] Underlined nucleotides are restriction sites of the enzymes indicated in brackets at the end. In bold is either the start (ATG) or stop (TAA, TGA) codon. R and F indicate Forward and Reverse primers, respectively.

Primer	Sequence (5' → 3') [a]	Expected fragment size	Gene name
copR1F	5'AACTGCAGAACATATGCGTTTATTGCTCGTTGA 3'(NdeI)	666 bp	*copR1*
copR1R	5'ATAAGAATCTCGAGCGGCCGCTTAACGTTGCGCATTGAACA 3'(XhoI)		
copR2F	5'GACGACGACAAGATGCTGGAAGTTCTGTTCCAGGGGCCAATGCGTTTATTGCTCGTTGA 3'	705 bp	*copR2*
copR2R	5'GAGGAGAAGCCCGGTTTAACGTTGGGCATTGAACA 3'		
copR_N1F	5'TAAGTAACATATGCGTTTATTGCTCGTTGA 3'(NdeI)	363 bp	*copR_N1*
copR_N1R	5'CCGCTCGAGTTATTCGCTGCGGCGTA 3'(XhoI)		
copR_N2F	5'TAAGTAACATATGCGTTTATTGCTCGTTGA 3'(NdeI)	363 bp	*copR_N2*
copR_N2R	5'CCGCTCGAGTTCGCTGCGGCGTATG 3'(XhoI)		
copR_C1F	5'TAAGTAACATATGCACTCCCAGGTTAAAGCCG 3'(NdeI)	303 bp	*copR_C1*
copR_C1R	5'CCGCTCGAGTTAACGTTGGGCATTGAACA 3'(XhoI)		
copR_C2F	5'TAAGTAACATATGCACTCCCAGGTTAAAGCCG 3'(NdeI)	297 bp	*copR_C2*
copR_C2R	5'CCGCTCGAGACGTTGGGCATTGAACATG 3'(XhoI)		
copS_CF	5'GACGACGACAAGATGCTGGAAGTTCTGTTCCAGGGGCCAGCGGCGTTAAAACGATTG 3'	756 bp	*copS_C*
copS_CR	5'GAGGAGAAGCCCGGTTTAAAACTCTATGGTTACGCGCA 3'		
proF	5'CATCCGTTTGTCTCCTGTGG 3'	202 bp	*Pro*
proR	5'GTATCCGTCCTTAACGGTGC 3'		

The aforementioned DNA fragments were amplified, by PCR, in a mix containing 1X PCR buffer (10 mM Tris-HCl pH 9.0, 1.5 mM $MgCl_2$ and 50 mM KCl), 0.2 mM dNTP's, 50 pmol of each primer (Forward and Reverse), 1 µL *M. hydrocarbonoclasticus* genomic DNA and 1.25 U HighFidelity DNA Polymerase (Fermentas), in a total reaction volume of 50 µL.

Amplification was performed in a PCR apparatus (Biometra) using the following program: an initial denaturation at 94°C for 5 min, PCR was performed for 30 cycles of denaturation at 95°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 1 min. Once concluded, the mixtures were subjected to a final extension at 72°C for 10 min, thereafter

the reactions were cooled and maintained at 4ºC. These amplicons were purified using the "QIAquick Gel extraction Kit" (Qiagen).

2.2.3. Construction of the Heterologous Expression Vectors

Insertion of purified fragments into the appropriate vectors was achieved by one of two methods: the ligation dependent method using restriction enzymes, into a pET-21 c (+) vector (Novagen) or by ligation independent cloning (LIC), into pET-30 EK/LIC vector (Novagen) or a pET-41 EK/LIC vector (Novagen). All plasmids obtained in this study are listed in Table 2.2.

Table 2.2 Plasmids constructed throughout the process of expression of the CopR, and respective domains and the C-terminal domain of CopS. Amp^R and Kan^R refer to ampicillin and kanamycin resistance respectively. N-His_6 and C-His_6 refer to an N-terminus or C-terminus hexahistidine tag, and N-GST to a glutathione-S-transferase tag. [1] Proteins contain a HRV3C cleavage site to the C-terminus of the tag. LIC – ligation independent cloning, n/a – non applicable. Gene refers to the amplified fragment which was cloned using the primers listed in Table 2.1.

Plasmid	Gene	Vector	Cloning Methodology	Resistance	Tag
pCopR1	*copR1*	pET-21 c (+)	Restriction Enzymes	Amp^R	n/a
pCopR2	*copR2*	pET-30 EK/LIC	LIC	Kan^R	N-His_6 [1]
pCopR_NA	*copR_NA*	pET-21 c (+)	Restriction Enzymes	Amp^R	C-His_6
pCopR_NB	*copR_NB*	pET-21 c (+)	Restriction Enzymes	Amp^R	n/a
pCopR_CA	*copR_CA*	pET-21 c (+)	Restriction Enzymes	Amp^R	C-His_6
pCopR_CB	*copR_CB*	pET-21 c (+)	Restriction Enzymes	Amp^R	n/a
pCopS_C1	*copS_C1*	pET-30 EK/LIC	LIC	Kan^R	N-His_6
pCopS_C2	*copS_C2*	pET-41 EK/LIC	LIC	Kan^R	N-GST, N-His_6 [1]

2.2.3.1 Ligation Dependent Cloning via Restriction Enzymes

A. Double Digestion with Restriction Enzymes

The pET-21 c (+) vector, as well as the previously PCR amplified and purified *copR1, copR_NA/B* and *copR_CA/B* DNA fragments, were separately digested in React buffer 2 (Invitrogen) with 10 Units XhoI and NdeI enzymes (Invitrogen) for 4 h and purified with the "QIAquick Gel extraction Kit" (Qiagen). Thereupon, dephosphorylation of 5 µL of the cleaved purified plasmid was achieved by 30 min incubation at room temperature (RT) with 1 U of alkaline phosphatase (Promega) in 1X Multicore buffer to a total volume of 10 µL. The enzyme was subsequently inactivated upon incubation at 75°C for 20 min.

B. Ligation with T4 Ligase and Transformation

Insertion of the DNA fragments, *copR1, copR_NA/B* and *copR_CA/B*, into the dephosphorylated vector, between the NdeI and XhoI restriction sites, was accomplished by adding 1 U of T4 DNA ligase (Roche) to 1 µL of the dephosphorylated vector, 2 µL of the pure

DNA fragment (diluted 1:10), in 1X ligase buffer, to a total reaction volume of 20 µL. This reaction was performed at RT for 15 min. Consequently, 7 µL of the ligation reaction was used to transform 50 µL of *E.coli* NovaBlue GigaSingles cells (Novagen), following the supplied protocol and plated on LB plates containing 100 µg/mL ampicillin. Plates were inverted and incubated at 37°C for approximately 16 h.

2.2.3.2 Ligation Independent Cloning and Transformation

Purified PCR fragments were treated with T4 DNA polymerase in a reaction mixture containing: 4 µL purified PCR product, T4 DNA polymerase buffer 1X, 2.5 mM dATP, 0.5 mM DTT and 1 U of T4 DNA polymerase. This reaction was carried out at 22°C for 30 min, followed by incubation at 75°C for 20 min to inactivate the enzyme.

Annealing between the LIC vectors and the DNA fragments of choice was achieved by addition of 1 µL of the appropriate EK/LIC vector to 2 µL of the pre-T4 DNA polymerase treated PCR fragment and incubating this reaction for 5 min at 22°C. After which 1 µL of 25 mM EDTA was added and the mixture was once more incubated for 5 min at 22°C. Then, 1 µL of each of the plasmids were used to transform 50 µL NovaBlue GigaSingles strain of *E.coli* cells (Novagen), following the established protocol and plated on LB plates containing 25 µg/mL kanamycin. Plates were inverted and incubated at 37°C, for approximately 16 h.

2.2.3.3 DNA Sequencing

A single cell colony of the putative recombinants was selected and inoculated in 5 mL sterile LB media, supplemented with the adequate antibiotic (100 µg/mL ampicillin or 25 µg/mL kanamycin) and grown for 16 h at 37°C on an orbital shaker at 210 rpm (Infors AG, CH-4103). Plasmid DNA was extracted from these cells with the miniprep extraction kit (Nzytech) and consequently digested for 1 h at 37°C by XhoI and NdeI (Fermentas). Insertion of the fragments was confirmed by electrophoretic agarose gel. The chosen plasmids were then sequenced by automated DNA sequencing (Stabvida, Portugal).

2.2.4 DNA Electrophoresis

DNA electrophoresis was performed at various stages, to check PCR amplification, purity, and for verification of plasmid purity and concentration. Typically, 0.8 or 1% agarose gels were prepared in 1X TAE buffer (40 mM Tris, pH 8.5, 20 mM acetic acid, and 1 mM EDTA), which was also the running buffer.

Samples were loaded using the loading solution (Fermentas, BioPortugal) in a 1:4 ratio and the DNA ladders 100 bp and/or 1 kbp (Fermentas, BioPortugal) were used as molecular markers. Electrophoresis was run at a constant voltage of 100 V, for approximately 20 min.

For visualization of the electrophoresis results, Sybr Safe (Invitrogen) was prepared and used in 1X TAE buffer. The gel was left in the staining solution for 30 min. As the staining agent retains a fluorescence excitation maximum at 280 nm and 502 nm, and an emission maximum at 530 nm, the gels were visualized using a blue-light transiluminator (Invitrogen – S3710). Images were recorded under UV-light with a VWR genosmart gel documentation system.

2.2.5 Heterologous Expression of Recombinant Proteins

2.2.5.1 Transformation into Competent Cells for Heterologous Expression

Following the established transformation protocol, 20 µL competent *E.coli* BL21(DE3) strain cells (Nzytech) were transformed with 1 µL of each of the plasmids (Table 2.2) and 200 µL of SOC medium ((20 g/L tryptone, 5 g/L yeast extract, 0.584 g/L NaCl, 0.186 g/L KCl, 0.952 g/L $MgCl_2$ and 3.603 g/L glucose); this medium was sterilized with a 0.2 µM pore filter (Millipore) and thereafter maintained at -20°C) was added. For strain construction and maintenance, these cells were grown on LB/agar plates supplemented with the appropriate antibiotic (100 µg/mL ampicillin or 25 µg/mL kanamycin) for 16 h at 37°C and then maintained at -4°C for no longer than a month.

2.2.5.2 Growth Conditions

All bacterial strains used in this study were grown in sterile LB medium (composition in Appendix A.1) to which antibiotics were added to the following final concentrations: 100 µg/mL ampicillin, and 25 µg/mL kanamycin.

A. Heterologous Test Expressions

A single colony was selected from the previously plated transformants and inoculated in 5 mL LB media with the appropriate antibiotic. These cells were aerobically incubated for 16 h at 37°C with an orbital agitation of 210 rpm (Infors AG, CH-4103). Of this growth, 1 mL was used to inoculate 50 mL LB medium in 250 mL erlenmeyer's at 37°C for expression assays. The optical density (OD) of the cellular growth was measured at 600 nm, at hourly intervals, on the UV spectrophotometer (Shimadzu, UV-160A) in plastic cuvettes, at RT with a 1 cm path length. Expression of the target proteins, when desired, was induced via addition of sterile isopropyl β-D-1-thiogalactoside (IPTG) to a final concentration of 0.1, 0.5 or 1 mM when the cellular growths presented an OD_{600nm} of approximately 0.6. After addition of IPTG the bacterial growths were

incubated at 16, 37°C or RT, with orbital rotation at 210 rpm for a range of hours, listed in Tables 2.3 and 2.4.

To verify expression of proteins on SDS-PAGE, 1 mL samples were collected and normalized to an OD_{600nm} of 1.2, assuming a loading sample of 20 µL.

Table 2.3 Conditions tested for the induction of expression for the CopR protein and domains, as well as the C-terminal domain of CopS_C. 16 hours correspond roughly to an expression which is grown over night and cells are harvested the following morning.

		pCopR1	
Induction OD_{600nm}	**[IPTG] (mM)**	**Temperature of induction (°C)**	**Length of induction (h)**
0.6	0.5	20	2
0.6	0.5	26	2
0.6	0.5	RT	2
0.6	0.5	37	2
0.6	0.5	RT	16
-	-	RT	2
-	-	26	16
-	-	RT	16
		pCopR2	
Induction OD_{600nm}	**[IPTG] (mM)**	**Temperature of induction (°C)**	**Length of induction (h)**
0.6	0.1	16	16
0.6	0.5	16	16
0.6	1.0	16	16
0.6	0.1	37	16
0.6	0.5	37	16
0.6	1.0	37	16
		pCopR_NA/B	
Induction OD_{600nm}	**[IPTG] (mM)**	**Temperature of induction (°C)**	**Length of induction (h)**
0.6	0.1	16	16
0.6	0.1	37	16
0.6	1.0	37	16
		pCopR_CA/B	
Induction OD_{600nm}	**[IPTG] (mM)**	**Temperature of induction (°C)**	**Length of induction (h)**
0.6	0.1	16	16
0.6	0.1	37	16
0.6	1.0	37	16
		pCopS_C1	
Induction OD_{600nm}	**[IPTG] (mM)**	**Temperature of induction (°C)**	**Length of induction (h)**
0.6	0.1	16	16
0.6	1.0	16	16
0.6	0.1	37	16
0.6	1.0	37	16
		pCopS_C2	
Induction OD_{600nm}	**[IPTG] (mM)**	**Temperature of induction (°C)**	**Length of induction (h)**
0.6	0.1	16	16
0.6	1.0	16	16
0.6	0.1	37	16
0.6	1.0	37	16

B. Solubility Assays

In cases of verified protein expression in the SDS-PAGE, before advancing to large-scale protein production, an additional assay was performed. Using the BugBuster reagent (Novagen) and following the supplied protocol and guidelines. The SDS-PAGE analysis of these fractions indicates the degree of solubility of the protein.

C. Large-scale Production of Heterologously Expressed Protein

Large-scale protein production was carried out in 2 L erlenmeyer's, in 0.5 L LB media upon determination of optimal growth conditions. Single cell colonies were aerobically cultured at 37°C in 20 mL LB supplemented with the appropriate antibiotics for 16 h and 10 mL of this pre-inoculum was used to inoculate the LB media for growth. Cells were grown at 37°C and 210 rpm, with the appropriate antibiotic, to an OD_{600nm} of approximately 0.6 before adding IPTG at the established best concentration. Upon induction of expression the cells were grown in the conditions described in Table 2.4.

Table 2.4 Variables chosen for protein expression; temperature and time of bacterial cell growth. Cells were invariably grown aerobically in LB media with orbital rotation (210 rpm) and induced at OD_{600nm} of approximately 0.6 with IPTG to a final concentration of 0.1 mM.

Plasmid	Temperature (°C)	Length of expression (h)	Cell mass (g/L)
pCopR2	16	16	4.52
pCopR_NB	16	3	6.44
pCopR_CB	16	3	5.46
pCopS_C2	37	3	5.77

2.2.5.4 Cell Fractionation

Cells were harvested in a Beckman centrifuge, model Avanti J-25 (rotor JA-10), at 8000 rpm, during 15 min at 6°C and resuspended in 10 mM Tris-HCl (pH 7.6) at an approximate concentration of 1 g of wet cell weight per 2 mL of buffer. Cells were stored at -20°C until further use.

Prior to cellular disruption, DNase I (Sigma) was added to the cells, as well as a cocktail of protease inhibitors (Roche) to a concentration of 1 tablet per 25 mL of cell extract. Lysis was achieved either by three passages in a French Press (Thermo-FA-080A using the 40K cell) at 700 psi or by high pressure homogenization. Clarification of the cellular extract was done by ultracentrifugation in a Beckman Coulter Optima LE-80K (rotor 45Ti) for 2 h, at 45000 rpm and 6°C. The resulting pellet was then resuspended in 50 mM Tris-HCl (pH 7.6); both fractions were stored at -20°C until further use, or directly loaded onto the chromatographic column for protein purification.

2.2.6 Protein Purification

Prior to injection in the columns, all buffers were filtered and samples were subjected to a 10 min, 14.000 rpm centrifugation (RC5C from Sorvall instruments (rotor GSA)) and kept on ice.

2.2.6.1 Hexahistidine-tag (His_6-tag) Containing Proteins

Proteins containing solely the His_6-tag, CopR2, CopR_NA and CopR_CA, were purified by Ni^{2+}-chelate affinity chromatography, on a 5 mL HisTrap FF column (GE Healthcare). The column was equilibrated with 5 column volumes (c.v) of binding buffer: 20 mM Tris-HCl (pH 7.6) and 500 mM NaCl prior to protein injection. Upon complete sample injection and column washing, with approximately 7 - 8 c.v of binding buffer, elution was performed by stepwise increments in imidazole concentration: 0, 20, 100, 200 and 500 mM, which is present in the elution buffer, 20 mM Tris-HCl (pH 7.6) and 500 mM NaCl.

Fractions of approximately 3 - 4 mL were collected and separated per phase of purification: application of sample, column wash, and elution. Upon selection of the fractions that contained the protein of interest (analysis performed *via* SDS-PAGE 12.5% polyacrylamide gel Tris-Tricine buffer system, as described in Section 2.2.7.2 Polyacrylamide Gel Electrophoresis (PAGE)) these were pooled and concentrated as described in Section 2.7.4 Protein Concentration and Buffer Exchange.

2.2.6.2 GST-tag Containing Proteins

Proteins containing the GST-tag, CopS_C and HRV3C, were purified using glutathione affinity, on a 5 mL GSTrap FF column (GE Healthcare). The column was equilibrated with 5 c.v of the binding buffer: 10 mM phosphate buffer (pH 7.4) and 150 mM NaCl, and subsequent to injection of 1 c.v of sample there was an incubation period of 1 h[1] at 4°C. Upon complete sample injection and column washing, with approximately 7 - 8 c.v of binding buffer, elution was performed by means of gradient in the presence of reduced glutathione (Sigma) in the elution buffer: 50 mM Tris-HCl (pH 7.6) and 5 – 20 mM reduced glutathione.

Fractions of approximately 3 - 4 mL were collected and separated per phase of purification: application of sample, column wash, and elution. Upon selection of the fractions which contained the protein of interest (analysis performed via polyacrylamide gel 12.5% gel in denaturing gel as described in Section 2.2.7.2 Polyacrylamide Gel Electrophoresis (PAGE))

[1] In extraordinary circumstances this incubation period was shortened to 20 – 30 min so as to complete the purification process without maintaining the samples at -4°C for over 24 h.

these were pooled and concentrated as described in Section 2.7.4 Protein Concentration and Buffer Exchange.

2.2.6.3 Tag Cleavage

A. Production of HRV3C Protease

Large-scale production of the HRV3C protease (in this construct HRV3C is fused to a GST-tag), was carried out as described for the CopR and CopS proteins. Cells were grown at 37°C in 0.5 L of LB media with kanamycin to a final concentration of 25 µg/mL. Before induction with IPTG, to a final concentration of 0.1 mM cells were grown to an OD_{600nm} of approximately 0.6, and harvested by centrifugation after an additional 1.5 h of induction. Cell lysis, supernatant clarification and HRV3C protease purification were carried out as previously described for the CopS_C protein. Concentrated enzyme was stored in 20 mM Tris-HCl buffer pH 7.6, at 4°C.

B. Cleavage and Purification

The His_6-tag and GST-tag were removed by treatment with the HRV3C protease produced *in-house*. Previously purified proteins were incubated with 0.4 mg/mL enzyme at 4°C, and DTT to a final concentration of 1 mM. Progression of the reaction was monitored by SDS-PAGE as described in Section 2.2.7.2 Polyacrylamide Gel Electrophoresis (PAGE).

As cleavage approached completetion or was completed, the cleaved protein was purified from this mixture of tagged and cleaved protein, protease and tag. In the case of proteins containing a GST-tag this purification was achieved by injection of samples into the GSTrap FF column, and for His_6-tag containing proteins the GSTrap column was placed in tandem with the 5 mL HisTrap FF column (GE Healthcare).

These columns were pre-equilibrated with the wash buffer: 50 mM Tris-HCl (7.6), 500 mM NaCl, followed by injection of 1 c.v sample and incubation for 1 h at 4°C until the whole sample had been injected. Column washing was performed by injection of 5 c.v of wash buffer followed by 4 c.v of elution buffer I and then buffer II (see below).

In the case of columns in tandem, the composition of elution buffer I was: 50 mM Tris-HCl (pH 7.6) and 300 mM imidazole, and the composition of elution buffer II was: 50 mM Tris-HCl (pH 7.6), and 10 mM reduced glutathione. In the case of GST-tagged proteins the elution buffer used was 50 mM Tris-HCl (pH 7.6) and 10 mM reduced glutathione.

Once again fractions of approximately 3 - 4 mL were collected and analyzed by SDS-PAGE. The fractions containing the protein of interest were pooled and concentrated as described in Section 2.7.4 Protein Concentration and Buffer Exchange.

2.2.6.4 Protein Concentration and Buffer Exchange

Samples were concentrated under a pressure of 4 psi using a pressure-fugation apparatus (Vivacell, Sartorius) and a membrane with the appropriate cut-off: 5 kDa for CopR_NHis$_6$ and CopR_CHis$_6$, 10 kDa for CopRHis$_6$ and 30 kDa for HRV3C and CopS_CHis$_6$GST. Concentration was preferentially performed to a final volume of approximately 2.5 mL (for buffer exchange) or until initiation of protein precipitation.

A maximum of 2.5 mL of protein sample was applied to a PD10 Desalting Column (Amersham) previously equilibrated with the selected buffer solution, after which the protein sample eluted with 3.5 mL of the buffer of choice listed in Table 2.5.

Table 2.5 Buffers chosen for medium to long term storage of the protein samples upon purification. The CopRHis$_6$ protein and respective C-terminal domain were maintained at 4°C.

Protein Sample	Elution Buffer
CopRHis$_6$	50 mM Tris-HCl (pH 7.6), 250 mM NaCl, 20% glycerol
CopR_NHis$_6$	20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 10% glycerol, 1 mM DTT
CopR_CHis$_6$	20 mM Tris-HCl (pH 7.6)
CopS_CHis$_6$GST	20 mM Tris-HCl (pH 7.6)

2.2.6.5 Protein Quantification - Bicinchoninic Acid Assay (BCA Assay)

The Bicinchoninic Acid Assay (BCA; “Protein Assay Kit” from Sigma) was used to determine the total protein content in solution. This method follows the same principles as the Lowry protein assay [113]. The procedure was executed as described by the manufacturer (Sigma) with the following adjustments: to 50 µL of protein sample, 1 mL of working BCA reagent was added and incubated during 30 min at 37°C.

Absorbance (abs.) at 562 nm was measured on a UV spectrophotometer (Shiadzu UV – 160A). Bovine serum Albumin (BSA) (Sigma) was used for protein calibration, in a concentration range from 0 to 300 µg/mL. A typical calibration curve obtained is shown in Figure 2.1.

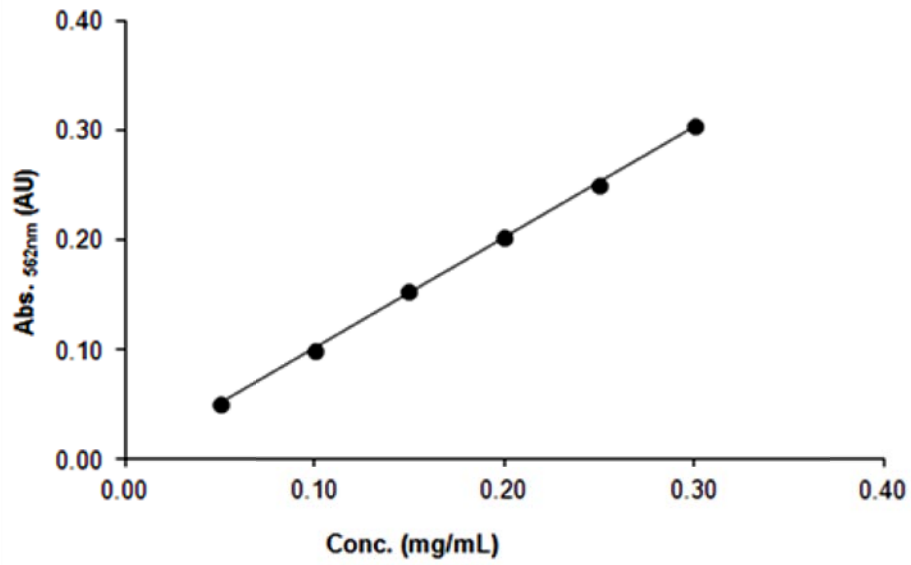


Figure 2.1 Graphic representation of a typical calibration curve obtained in the BCA method for the determination of total protein concentration. Legend: The full circles represent the absorption readings for BSA at various concentrations. The equation for the linear regression is, y = 1.01x, which has a R of 1.00.

2.2.7 Protein Techniques

2.2.7.1 SDS-PAGE

Preparation of samples for electrophoresis encompassed addition of deposition solution (0.1 M Tris-HCl pH 6.8, 0.25 M β-mercaptoethanol, 2.5% SDS, 5% glycerol, 0.2 mM bromophenol blue) β-mercaptoethanol and SDS were used in function of the conditions of the gel, non-denaturing or denaturing, and subsequent subjection to thermal denaturation (100°C for 5 min).

2.2.7.2 Polyacrylamide Gel Electrophoresis (PAGE)

Typically, Tris-Tricine polyacrylamide gels were 0.75 mm, 10/12.5% and 5% acrylamide/bisacrylamide for the separation and concentration gels, respectively. These gels were prepared in either denaturing (SDS-PAGE) or native conditions (PAGE), and electrophoresis was run at constant voltage, 150 V, for 45 to 60 min with anode buffer, 0.2 M Tris-HCl (pH 8.9); and 0.1 M Tris-HCl, 0.1 M Tricine and 0.1% SDS (pH 8.25) as the cathode buffer for SDS-PAGE and 0.1 M Tris-HCl and 0.1 M Tricine as cathode buffer for PAGE. The molecular marker used for SDS-PAGE gels was the LMW-SDS markers (Fermentas, BioPortugal).

The Tris-glycine polyacrylamide gels were prepared exclusively in native conditions and utilize 10% and 5% acrylamide/bisacrylamide for the separation and concentration gels, respectively. Electrophoresis was run at constant voltage, 150 V, for 45 to 60 min with 0.025 M

Tris (pH 8.5), 0.192 M glycine, 0.1% SDS as the electrophoresis buffer (both anode and cathode).

2.2.7.3 Protein Staining Procedure

Gels were stained for the presence of protein with Coomassie blue (Coomassie Blue R-250) 5 g/L in a 45% methanol, 7.5% acetic acid solution, which has a detection limit of the order of nmol of protein. Gels were stained for approximately 30 min, and distained in 45% methanol and 7.5% acetic acid until the blue background disappeared from the gel.

Images were recorded with a VWR genosmart gel documentation system and dried for storage.

2.3. Biochemical Characterization

2.3.1 Analytical Size-exclusion Chromatography

Apparent molecular mass for individual proteins was determined on a gel filtration column, Superdex 75 (GE Healthcare) (10 mm (Ø) X 300 mm), equilibrated with 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl.

For the calibration of this column, various protein samples (Gel filtration calibration kit LMW and HMW, GE Healthcare) with established molecular masses were injected. Each sample was prepared as described in the GE Healthcare protocol and injected in a 1:4 ratio in the equilibration buffer. Depending on the protein, a subset of the proteins listed in Table 2.6 was used for calibration. Elution was performed with a 0.5 mL/min flow rate over 1.5 c.v for all samples (calibration and studied proteins).

Table 2.6 Proteins from gel filtration calibration kits (HMW and LMW from GE Healthcare) utilized for determination of the calibration curves used to determine apparent molecular masses throughout this work.

Protein	Concentration (mg/mL)	Molecular Mass (kDa)
Aprotinin	2	6.5
Horse cytochrome *c*	2	12.4
Ribonuclease A	3	13.7
Chymotrypsinogen A	2	25.0
Carbonic anhydrase	3	29.0
Ovalbumin	4	43.0
Albumin	5	66.0

2.3.2 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra of purified proteins were acquired using a 0.1 cm quartz cell, at 25°C, on a JASCO CD spectrometer, model J-815. The collected spectra covered a wavelength range of 200-240 nm with a 20 nm/min scanning speed; a bandwidth of 2 nm, a response time of 4 s, and 0.1 nm data pitch. Final spectra were the sum of 8 scans. The CD data was converted from machine units of milidegrees to per residue molar absorption units of CD, measured in mdeg $M^{-1}cm^{-1}$ by application of the equation presented in Figure 2.2. Deconvolution of spectra was performed utilizing the graphic data interface K_2D_2 [114].

$$\Delta\varepsilon = \theta \times \frac{0.1 \times \mathbf{MRW}}{\mathbf{P} \times \mathbf{Conc} \times 3298}$$

Figure 2.2 Equation for the conversion of mean residue elipticity (θ), to per residue molar absorption units of circular dichroism (Δε). MRW refers to the mean residue weight, which is the protein mass units (Da)/number of residues. P is the path length in cm and conc. is the concentration of the protein sample in mg/mL.

2.3.3 Biochemical Characterization of CopR

2.3.3.1 Stability Assays

Stability of the full-length CopR protein was independently assayed twice throughout this work. On the first occasion DTT to a final concentration of 1 mM was added to the CopR1 sample. In the second instance, glycerol and salt, NaCl, were added to aliquots of 200 µL of CopR2, CopRHis_6 in 50 mM Tris-HCl (pH 7.6) with 150 mM NaCl. The stability conditions tested are presented in Table 2.7, and upon addition of the stability agent and salt, samples were diluted to 600 µL, gently homogenized and placed on ice for approximately 8 h. After this period the samples were centrifuged for 30 min at 14,000 xg, RT and 5 µL were used for SDS-PAGE analysis.

Table 2.7 Conditions tested for CopR protein stability. Original samples were in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and were used as a control. The effect of a stability agent and osmolarity were assayed.

	NaCl (mM)	Glycerol (%)
Control	150	0
Assay	50	10
	50	20
	50	30
	250	10
	250	20
	250	30

2.3.3.2 *In vitro* Phosphorylation Assay

Approximately 3 µM of purified CopR protein was incubated with the phosphorylation buffer: 50 mM Tris-HCl (pH 7.6) containing 200 mM KCl, 20 mM $MgCl_2$, 30 mM acetyl phosphate to a total reaction volume of 75 µL. Phosphorylation was carried out at RT for 30 min, and 20 µL aliquots were removed and stopped at different time periods, 1 h, 2 h, and 3 h by addition of PAGE sample buffer and placed on ice. A control reaction with 15 µL of CopR was also prepared. The samples were then analyzed on a 10% PAGE Tris-Glycine gel and the proteins were visualized by Coomassie staining.

2.3.3.3 Electrophoretic Mobility Shift Assay (EMSA)

The proposed promoter DNA fragment, *pro*, was generated by PCR as described previously. This dsDNA was quantified spectrophotometrically by application of the Beer Lambert law. At a wavelength of 260 nm, the average extinction coefficient for double-stranded DNA is 0.020 $(\mu g/mL)^{-1}$ cm^{-1}.

Samples of a 1:200 dilution, to a minimum of 400 µL were prepared of the purified dsDNA and the absorbance was registered on a UV spectrophotometer (Shimadzu, UV-160A) in a quartz cell with a 0.5 cm path length at RT. A spectrum of the sample was taken in the range of 220 – 320 nm (Figure 2.3).

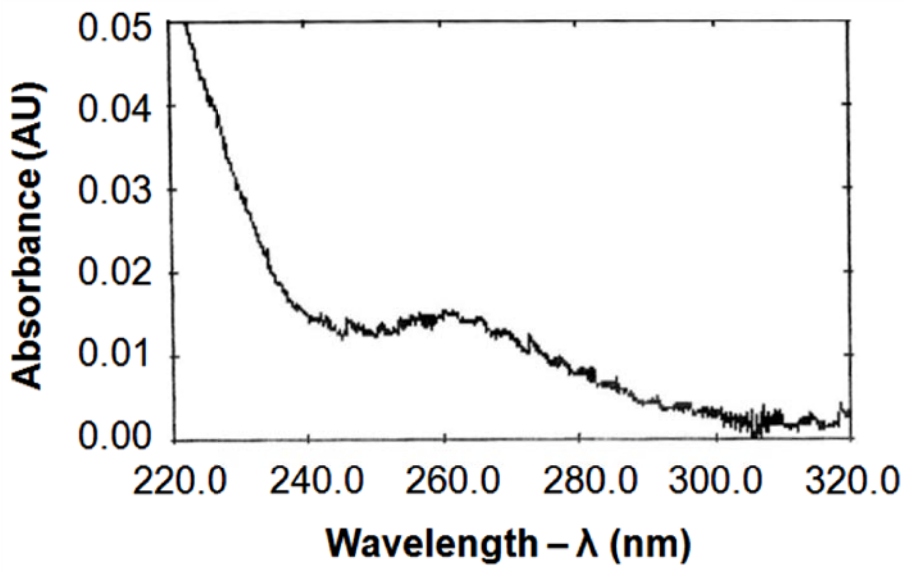


Figure 2.3 Optical density of a DNA sample diluted 1:200 of the *pro* fragment in function of the wavelength, after purification. The calculated concentration of this sample is 0.55 mM based on the absorbance at 260 nm. Legend: AU corresponds to arbitrary absorbance units. The absorbance spectrum was measured on a UV spectrophotometer (Shiadzu UV – 160A).

Upon 3 h phosphorylation of CopR, the DNA was mixed with CopR, in a 1:1 ratio and incubated at 37°C for 20 min in 25 µL 5X binding buffer (final concentrations of 100 mM Tris-HCl (pH 7.6); 100 mM KCl; 50 mM $MgCl_2$, 12.5% glycerol; 1.5 mg/mL BSA and 5 mM DTT). The samples were then separated by electrophoresis in nondenaturing 5% Tris borate (TB) gels.

A. Tris-Borate Gels

The EMSA reactions were analyzed by electrophoresis through nondenaturing 5% (unless otherwise mentioned) polyacrylamide gels (acrylamide/bisacrylamide [40:1.1] in 1X TB buffer). The gels were pre-run for an hour at 120 V at 4°C, upon sample loading they were run at 120 V for 1 - 3 h (until the dye-front reached the end of the gel), and visualized by Sybr Safe (Invitrogen) as described previously and/or silver stained. The protocol used for silver staining is described in Appendix A.2.

CHAPTER 3 – RESULTS AND DISCUSSION

3. Results and Discussion

3.1 *Marinobacter aquaeolei* VT8

3.1.1 Bioinformatic Analysis

Two-component systems (TCS) are, as was described in the introduction, widely distributed throughout bacteria, of which the *Marinobacter* genus is no exception. A cursory analysis of the NCBI database revealed that the genome of the strain used in this work, *M. hydrocarbonoclasticus* 617, is not sequenced, but that of a close relative, *M. aquaeolei* VT8 is. Previous work showed that several sequenced genes from *M. hydrocarbonoclasticus* 617 share almost 100% identity with those of *M. aquaeolei* VT8. For this reason, all posterior bioinformatic analysis was performed for the latter and assumed to be translatable to *M. hydrocarbonoclasticus* 617.

A primary analysis of the chromosomal genome of *M. aquaeolei* VT8, utilizing the P2CS database [16], which contains a compilation of the TCS genes within completely sequenced genomes, showed the presence of approximately 96 proteins involved in TCS, 42 Response Regulators (RR) (Figure 3.1.A.), 54 Histidine Kinases (HK) (Figure 3.1.B.) dispersed throughout the genome, and a sole phosphotransfer protein.

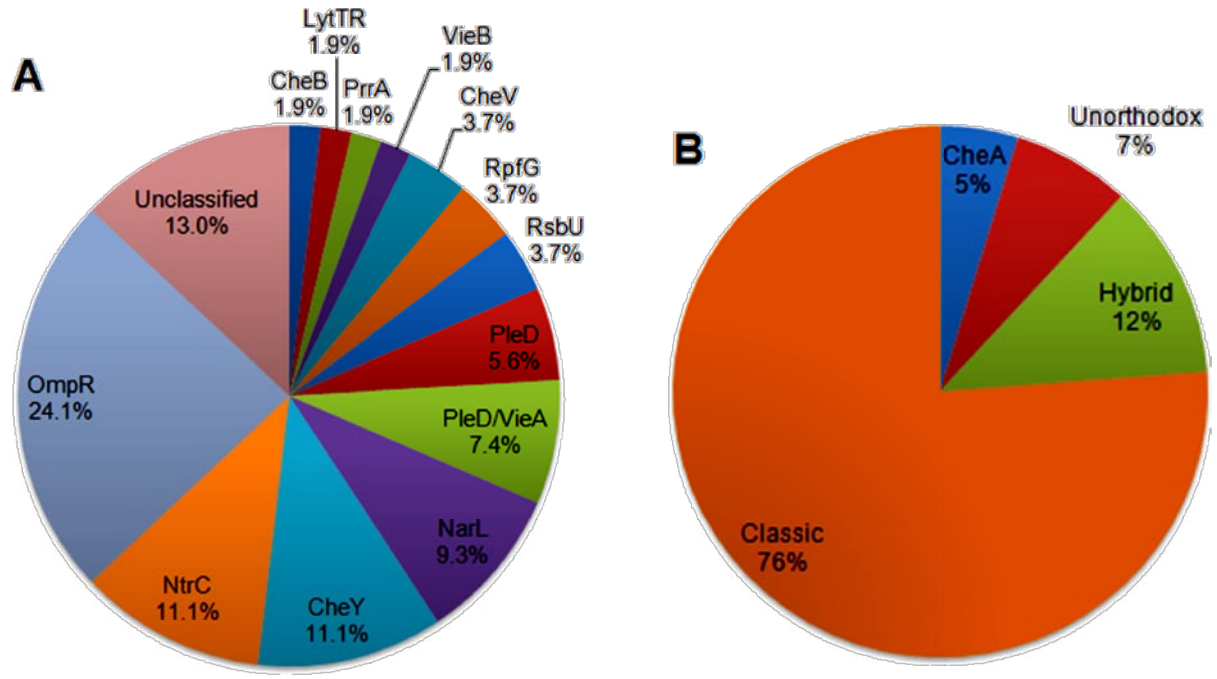


Figure 3.1 Graphic representation of the distribution of **(A)** RR and **(B)** HK classes present in the *M.aquaeolei* VT8 chromosomal genome. Data obtained, from P2CS, for the compilation of these graphs are presented in Appendix C.1 - Tables 6.3 and 6.4.

The previous graphs emphasize the diversity of classes within the TCS components of *M.aquaeolei* VT8, but show that proteins of the OmpR sub-family of RR (24.1%) and the classic HK (76%) are the most prevalent.

Previous bioinformatic analysis and preliminary studies identified a putative chromosomal copper resistance operon, *copXAB* (open reading frames maqu_0125, maqu_0126 and maqu_0127) in *M. aquaeolei* VT8 (Figure 3.2). Hypothetically this operon is regulated by the two-component system, *copSR* (ORF maqu_0123 and maqu_0124), located upstream of the copper resistance operon (Figure 3.2).

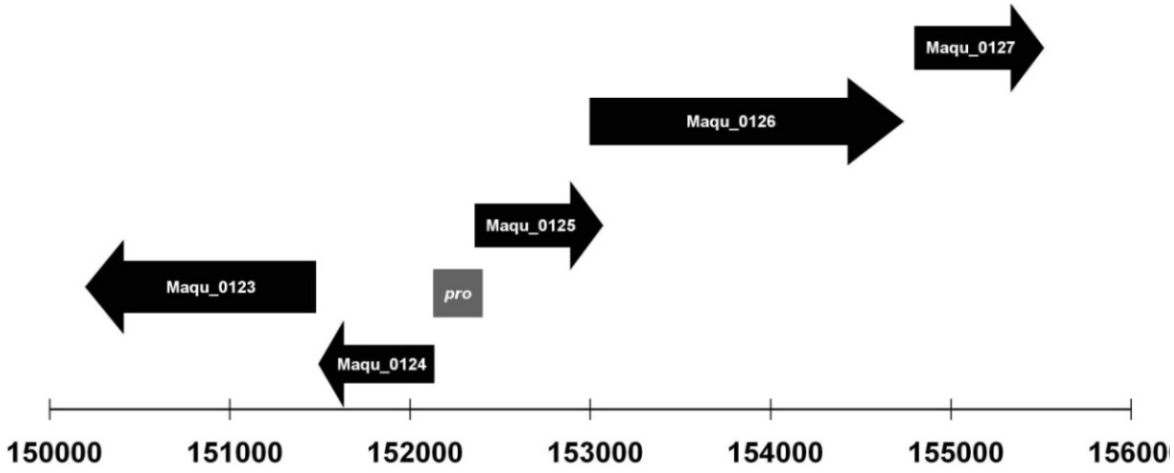


Figure 3.2 Schematic representation of a portion of the chromosomal DNA from *M. aquaeolei* VT8. The two open reading frames, maqu_0123 and maqu_0124, form the *copSR* operon, proteins which will belong to a two-component signal transduction system. The maqu_125, maqu_126 and maqu_127 open reading frames encode structural proteins presumably involved in copper ion resistance. *pro* is the stretch of DNA which is assumed, in this work, to contain the promoter region for both operons.

The CopS and CopR proteins form a TCS, in which CopS, the sensor HK, is proposed to be anchored in the inner cell membrane. This protein undergoes autophosphorylation at a conserved His residue in an ATP dependent manner, in response to an environmental change, in this case it is expected to be an alteration in copper concentration. This information is then conveyed to the interior of the cell via transfer of the resulting high-energy phosphoryl group to a conserved aspartyl residue of the CopR protein, the intracellular response regulator protein. This protein is in turn activated by this phosphorylation, which typically enhances its binding to target sites on DNA and promotes transcription of the structural genes of the copper resistance system, *copXAB* (Figure 3.3).

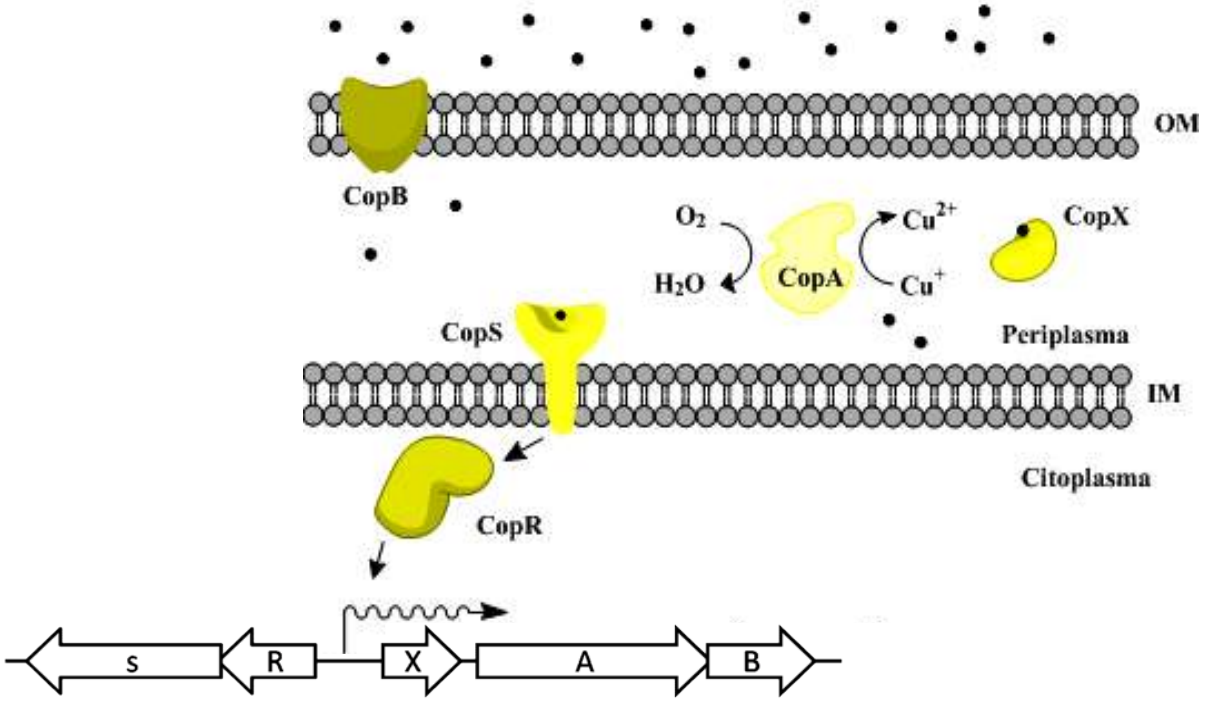


Figure 3.3 Schematic representation of a possible mechanism for copper resistance in *M. hydrocarbonoclasticus* 617. Dashed lines represent protein-protein and protein-DNA interactions.

The *copSR* operon could to some extent be considered a classic/model two-component system, in such that the CopS protein, is a proposed classic integral membrane bound HK, and the CopR protein belongs to the largest and most studied family of RRs, the OmpR/PhoB family of winged helix-turn-helix (wHTH) transcription factors.

However, each system, and essentially each sensor/effector domain is unique. This fact is due to the inherent variability of these systems which ensures specificity in protein-protein interactions, the stimuli sensed and the DNA sequence to which the transcription factor binds.

Also, given the difficulty of studying these proteins, the number of structures, and hence knowledge of these systems is limited. Therefore, a detailed study of each component of this TCS and their interaction becomes important to elucidate, add upon, and/or validate the current knowledge. A better understanding of the copper resistance mechanism in this bacterium could also prove to be beneficial for biotechnological applications.

Table 3.1 Putative *copSRXAB* operon gene motifs and predicted roles, data obtained from KEGG [115].

Protein Name	CopS	CopR	CopX	CopA	CopB
ORF ID	Maqu_0123	Maqu_0124	Maqu_0125	Maqu_0126	Maqu_0127
Genomic Location	150214-151542	151529-152194	152394-152927	153014-154723	154729-155469
Orientation	←	←	→	→	→
Predicted Role	Integral membrane sensor signal transduction histidine kinase	Two component transcriptional regulator	Copper binding protein	CopA family copper resistance protein	Copper resistance protein B
Pfam Motif(s)	HATPase_c, HAMP, HisKA, DUF677	Response_reg, Trans_reg_C	Copper-bind, Cu-oxidase_2, SoxE	Cu-oxidase_3, Cu-oxidase_2, Copper-bind, SH3_2	CopB

Chapter 3 – Results and Discussion
Response Regulator CopR

3.2. Response Regulator – CopR

3.2.1 Bioinformatic analysis

3.2.1.1 Analysis of the Primary Sequence and Protein Homology

The maqu_0124 ORF of *M.aquaeolei* VT8 (Accession no. YP_957418.1), designated throughout as *copR*, encodes the CopR protein which is annotated as a two-component transcription regulator of the OmpR/PhoB response regulator family, and wHTH sub-family.

As the schematic representation of the operon showed (Figure 3.2) it is intuitive to assume that this transcription factor regulates not only the structural genes, *copXAB* (ORF maqu_0125, maqu_0126 and maqu_0127), but also the operon which includes itself and its cognate sensor kinase, identified as *copS*, maqu_0123. This hypothesis can be verified through quantitative real-time PCR experiments and DNA binding experiments (electrophoretic mobility shift assays and DNA foot-printing).

A bioinformatic approach was used to analyze the primary sequence and the information that can subsequently be derived. Translation, as described in Chapter 2 Materials and Methods - Section 2.1.2, of the proposed *copR* gene, which consists of 666 nucleotides with a GC content of 57%, rendered various reading frames of translation (Appendix C.2.1, Figure 6.5). The first of these corresponds to the logical primary sequence for CopR (Figure 3.4). This choice was based on the assumption that the first amino acid should be a methionine (Met), the last codon should correspond to a stop codon (TAA, TAG or TGA) and the protein sequence is the longest uninterrupted ORF produced. These assumptions are considered valid and are used in all subsequent DNA sequence translations.

```
ATGCGTTTATTGCTCGTTGAAGACGACCGTTTGCTGGCCGAGGGGCTGGTCAGACAGCTGGAAAAAGCGGGGTTTAGCATCGAC
 M  R  L  L  L  V  E  D  D  R  L  L  A  E  G  L  V  R  Q  L  E  K  A  G  F  S  I  D
CACACGTCCAGTGCCCGTGAGGCCCAGATTCTGGGGGAGCAGGAGGATTATCGTGCCGCTGTTCTCGATCTCGGCCTGCCGGAT
 H  T  S  S  A  R  E  A  Q  I  L  G  E  Q  E  D  Y  R  A  A  V  L  D  L  G  L  P  D
GGCAACGGGCTGGAGGTTTTGAAACGATGGCGATCGAAGCTCGTCGAGTGCCCGGTATTGGTCCTCACCGCCAGAGGCGACTGG
 G  N  G  L  E  V  L  K  R  W  R  S  K  L  V  E  C  P  V  L  V  L  T  A  R  G  D  W
CAGGACAAGGTCAATGGACTGAAAGCAGGGGCGGATGACTATCTGGCCAAACCGTTCCAGACAGAAGAACTGATCGCCCGCATC
 Q  D  K  V  N  G  L  K  A  G  A  D  D  Y  L  A  K  P  F  Q  T  E  E  L  I  A  R  I
AATGCACTCATACGCCGCAGCGAAGGGCGAGTGCACTCCCAGGTTAAAGCCGGTGGCTTCGAACTGGACGAAAATCGCCAGAGC
 N  A  L  I  R  R  S  E  G  R  V  H  S  Q  V  K  A  G  G  F  E  L  D  E  N  R  Q  S
CTGCGGACAGAAGAAGGAGCAGAACACGCCCTGACGGGTACTGAGTTCCGCCTGCTACGATGCCTGATGAGCCGTCCGGGCCAC
 L  R  T  E  E  G  A  E  H  A  L  T  G  T  E  F  R  L  L  R  C  L  M  S  R  P  G  H
ATCTTTTCCAAGGAACAGCTAATGGAGCAGCTATACAACCTGGATGAGAGCCCCAGCGAAAACGTGATTGAGGCGTATATTCGG
 I  F  S  K  E  Q  L  M  E  Q  L  Y  N  L  D  E  S  P  S  E  N  V  I  E  A  Y  I  R
CGCTTGAGAAAGCTGGTCGGCAACGAAACGATCACCACACGCCGTGGCCAGGGATACATGTTCAATGCCCAACGTTAA
 R  L  R  K  L  V  G  N  E  T  I  T  T  R  R  G  Q  G  Y  M  F  N  A  Q  R  -
```

Figure 3.4 DNA and primary sequence of the *copR* gene of 666 nucleotides from *M. aquaeolei* VT8, maqu_0124. Highlighted in red are the start (ATG) and stop (TAA) codons.

The ProtParam tool, from the ExPASy database [94], in conjunction with data obtained from the P2CS database [21], provided a consensus on the expected molecular mass of the protein, 25 kDa, as well as the theoretical isoelectric point (pI), approximately 6, and an

overview of the amino acid usage, detailed in Appendix C.3, Table 6.5. At this point of the analysis, only the presence of two cysteines in the primary sequence is of worthwhile mention, however additional data from this analysis shall be relevant in forthcoming sections.

For prediction of the subcellular localization of the CopR protein, the PSORTb bioinformatic tool [97], which was further corroborated by other similar programs (for details see Appendix C.4, Figure 6.6), predicted, as expected for a prokaryotic transcription factor which interacts with DNA, that the CopR protein is predominantly present within the cytoplasm.

The modularity of TCS proteins allows a common regulatory scheme to be used by proteins with a wide variety of effector domains. To attest such a modularity, SMART (Simple Modular Architecture Research Tool) [39] was used to analyze the protein primary sequence and revealed the presence of two clearly defined domains identified in Figure 3.5. The length and presence of these domains was confirmed with the bioinformatic programs, Pfam [40] and the Conserved Domain Database (CDD) [100].

- The conserved N-terminal domain, which is the regulatory site of the protein and spans residues 3 - 117. It is in this domain that phosphorylation occurs via the HK domain of the HK protein, as well as, the conformational alterations on the surfaces of each monomer, which lead to dimerization.
- The variable C-terminal domain spans residues 151 - 217. This is the effector domain of the protein, as it is in this region that the amino acids which interact with the dsDNA are located, allowing the protein to regulate transcription of the structural genes.

Figure 3.5 Schematic representation of the CopR protein of 221 amino acids, which is predicted to present two domains: The receiver (amino acids 3 – 117) and the effector (amino acids 151 – 217) domains.

By conducting a BLAST search [95], as described in Chapter 2 Materials and Methods - Section 2.1.2, one can hypothetically identify the proteins that perform similar function to CopR, as these would have the highest sequence identity (Figure 3.6).

Figure 3.6 ClustalW multiple sequence alignment of proteins which share above 50% identity with the primary sequence of the RR, CopR, from *M.aquaeolei* VT8. Legend: ABO_1365 from *Alcanivorax borkumensis* SK2, Mmc1_0312 from *Magnetococcus* sp. MC-1, MELB17_11654 from *Marinobacter* sp. ELB17, MDG893_10211 from *M. algicola* DG893, MAMP_00174 from *Methylophaga aminisulfidivorans* MP, PST_0851 from *Pseudomonas stutzeri* A1501, Alvin_0011 from *Allochromatium vinosum* DSM 180. For the refseq identifiers see Appendix B4, Table 6.2. Asterisks (*), colons (:) and stops (.) below the sequence indicate identity, high conservation or conservation of the amino acids, respectively. Boxed regions correspond to the receiver (grey) and effector (green) domains. Amino acids in green form the presumed four antiparallel β-sheets between the receiver and effector domains characteristic of the OmpR/PhoB family of RRs.

The protein refseq annotation for each of the recovered sequences indicates that they all function as response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain. As one would expect, both domains are highly conserved within the protein sequences of Figure 3.6. The effector domain (highlighted in green) shows great conservation, presumably due to the common fold that this region of the protein must adopt in order to interact with DNA.

However, the region which corresponds to the four β-sheets that precede the effector domain and are characteristic of the OmpR/PhoB subfamily does not figure in the predicted region of the effector domain and one can verify that there is a low sequence similarity in this region (represented in green in Figure 3.6). It could be suggested that this local low sequence similarity be due to the function of this region. Presumably, it is here that the receiver and effector domain, upon phosphorylation and consequent activation, interact, which differs even between structurally similar proteins, as this alignment(Figure 3.6), between sequence that share above 50% identity, show.

3.2.1.2 Secondary Structure Prediction and Model Structure of CopR

Secondary structure predictions are still not 100% reliable, hence the best way to obtain a global and somewhat bias free prediction is to utilize as many of these programs as possible. This information is vital when there is no tertiary structure information available. Prediction of protein secondary structure then provides a platform for mechanistic and structural analysis, which is essential to interpret experimental data.

Although various other programs were used to predict the secondary structure (Appendix C.6.1 – Figure 6.7), those were, for the most part consistent with the results presented here. Figure 3.7 presents the secondary structure prediction for CopR by the Porter prediction server [101].

```
        10         20         30         40         50         60         70         80
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL VECPVLVLTA
RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS LRTEEGAEHA LTGTEFRLLR
CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR RGQGYMFNAQ R
```

Figure 3.7 Secondary structure prediction, performed by the Porter prediction server [101], for CopR based on its primary sequence. In blue and red are the predicted β-sheets and α-helices, respectively.

The overall receiver domain topology, $(\alpha\beta)_5$, and effector domain appear to be defined, as the secondary structures, which characterize these domains are present in the prediction. The effector domain, which presents a variation on the winged helix topology, is consistent with proteins of the wHTH sub family; three α-helices flanked on both sides by antiparallel β-sheets, and with an N-terminal four-stranded β-sheet and a C-terminal hairpin.

The work of Menon S. and Wang S., published in June 2011 [116], highlighted the existence, to date, of solely 6 structures of proteins of the OmpR/PhoB DNA-binding protein family. These were isolated from two organisms: DrrD (PBD ID: 1KGS), DrrB (PBD ID: 1P2F) and PhoP (PBD ID: 3R0J) from *T. maritima*, and PrrA (PBD ID: 1YS6), MtrA (PBD ID: 2GWR), and RegX3 (PBD ID: 2OQR) from *Mycobacterium tuberculosis*. The primary sequences of these proteins were retrieved from PDB and aligned utilizing ClustalW [96], Figure 3.8.

This alignment was performed with the objective of identifying the protein with known structure which shares the highest sequence identity with the CopR primary sequence for homology modelling. Based on the obtained scores, the DrrD RR protein of *T. maritima* (PDB ID: 1KGS, 36% sequence homology, Appendix C.8, Figure 6.9) was used as a template for the SWISS-model software [107].

```
CopR              ----------------------MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQ 42
1KGS (36%)        MN-------------------VRVLVVEDERDLADLITEALKKEMFTVDVCYDGEEGMYMALN 44
1YS6 (34%)        MDTGVTS--------------PRVLVVDDDSDVLASLERGLRLSGFEVATAVDGAEALRSATE 49
2GWR (32%)        MDTMRQR---------------ILVVDDDASLAEMLTIVLRGEGFDTAVIGDGTQALTAVRE 47
3R0J (31%)        MRKGVDLVTAGTPGENTTPEARVLVVDDEANIVELLSVSLKFQGFEVYTATNGAQALDRARE 62
2OQR (31%)        ----ATS---------------VLIVEDEESLADPLAFLLRKEGFEATVVTDGPAALAEFDR 43
1P2F (28%)        ------------------MMWKIAVVDDDKNILKKVSEKLQQLG-RVKTFLTGEDFLND--E 42
                                        : :*:*:  :   :   *.          .     .

CopR              EDYRAAVLDLGLPDGNGLEVLKRWRSKLVECPVLVLTARGDWQDKVNGLKAGADDYLAKPFQ 102
1KGS (36%)        EPFDVVILDIMLPVHDGWEILKSMRESGVNTPVLMLTALSDVEYRVKGLNMGADDYLPKPFD 104
1YS6 (34%)        NRPDAIVLDINMPVLDGVSVVTALRAMDNDVPVCVLSARSSVDDRVAGLEAGADDYLVKPFV 109
2GWR (32%)        LRPDLVLLDLMLPGMNGIDVCRVLRAD-SGVPIVMLTAKTDTVDVVLGLESGADDYIMKPFK 104
3R0J (31%)        TRPDAVILDVMMPGMDGFGVLRRLRADGIDAPALFLTARDSLQDKIAGLTLGGDDYVTKPFS 122
2OQR (31%)        AGADIVLLDLMLPGMSGTDVCKQLRAR-SSVPVIMVTARDSEIDKVVGLELGADDYVTKPYS 122
1P2F (28%)        EAFHVVVLDVMLPDYSGYEICRMIKETRPETWVILLTLLSDDESVLKGFEAGADDYVTKPFN 102
                  :**: :*  .*  :    :         .::   .    : *:  *.***: **:

CopR              TEELIARINALIRRSEGRVHSQ----VKAGGFELDENRQSLRTEEGAEHALTGTEFRLLRCL 158
1KGS (36%)        LRELIARVRALIRRKSESKSTK----LVCGDLILDTATK-KAYRGSKEIDLTKKEYQILEYL 159
1YS6 (34%)        LAELVARVKALLRRRG-STATSSSETITVGPLEVDIPGR-RARVNGVDVDLTKREFDLLAVL 167
2GWR (32%)        PKELVARVRARLRRNDDEPAEM----LSIADVEIDVPAH-KVTRNGEQISLTPLEFDLLVAL 156
3R0J (31%)        LEEVVARLRVILRRAGKGNKEPRNVRLTFADIELDEETH-EVWKAGQPVSLSPTEFTLLRYF 181
2OQR (31%)        ARELIARIRAVLRRGGDDDSEMSDGVLESGPVRMDVERH-VVSVNGDTITLPLKEFDLLEYL 161
1P2F (28%)        PEILLARVKRFLEREKKGLYDF-------GDLKIDATGF-TVFLKGKRIHLPKKEFEILLFL 153
                  ::**:.  :.*               . . :*          .    *.  *: :*  :

CopR              MSRPGHIFSKEQLMEQLYNLDESPSENVIEAYIRRLRKLVGNET-----ITTRRGQGYMFNAQR 221
1KGS (36%)        VMNKNRVVTKEELQEHLWSFDDEVFSDVLRSHIKNLRKKVDKGFK-KKIIHTVRGIGYVARDE- 225
1YS6 (34%)        AEHKTAVLSRAQLLELVWGYDFAADTNVVDVFIGYLRRKLEAGGG-PRLLHTVRGVGFVLRMQ- 233
2GWR (32%)        ARKPRQVFTRDVLLEQVWGYRHPADTRLVNVHVQRLRAKVEKDPENPTVVLTVRGVGYKAGPP- 228
3R0J (31%)        VINAGTVLSKPKILDHVWRYDFGGDVNVVESYVSYLRRKIDTGE--KRLLHTLRGVGYVLREP- 246
2OQR (31%)        MRNSGRVLTRGQLIDRVWGADYVGDTKTLDVHVKRLRSKIEADPANPVHLVTVRGLGYKLE--- 226
1P2F (28%)        AENAGKVVTREKLLETFW--EDPVSPRVVDTVIKRIRKAIEDDPNRPRYIKTIWGVGYMFTG-- 217
                  .   :.::  : : .:          :   :  :*  :         : *     * *:
```

Figure 3.8 ClustalW multiple sequence alignment of the six full-length structurally determined proteins of the wHTH subfamily, aligned against the primary sequence of the response regulator, CopR, from *M.aquaeolei* VT8. Conserved residues and sequence motifs identified by multiple sequence alignment. Legend: Asterisks (*), colons (:) and stops (.) below the sequence indicate identity, high conservation or conservation of the amino acids, respectively. For the accession numbers see Appendix B.4 - Table 6.2. Boxed regions correspond to the receiver (grey) and effector (green) domains of CopR.

The obtained model structure, Figure 3.9, was viewed in Chimera [108] and its energy was minimized as described in Chapter 2 Materials and Methods - Section 1.3. It is important to point out that this model is of the protein in the non-phosphorylated state, not bound to DNA, and thus not in a postulated activate conformation (phosphorylated dimer).

This model was verified in terms of validity using the z-score obtained from ProSA [110] and the Ramachandran plot from PROCHECK [109]. ProSA utilizes only the C-α atoms of the input structure and, as such, the z-score obtained for the overall model may be used as a measure of scores typical for native proteins of a similar size in the PDB database. In this case, the z-score is of -6.88, Figure 6.13 A in Appendix C.9, which is well within the values obtained for structures determined by either NMR or X-ray crystallography. PROCHECK evaluated the overall quality of the structure, which was good, with only one amino acid, Asp56, falling into the disallowed region of the Ramachandran plot (Appendix C.9.1 - Figure 6.10).

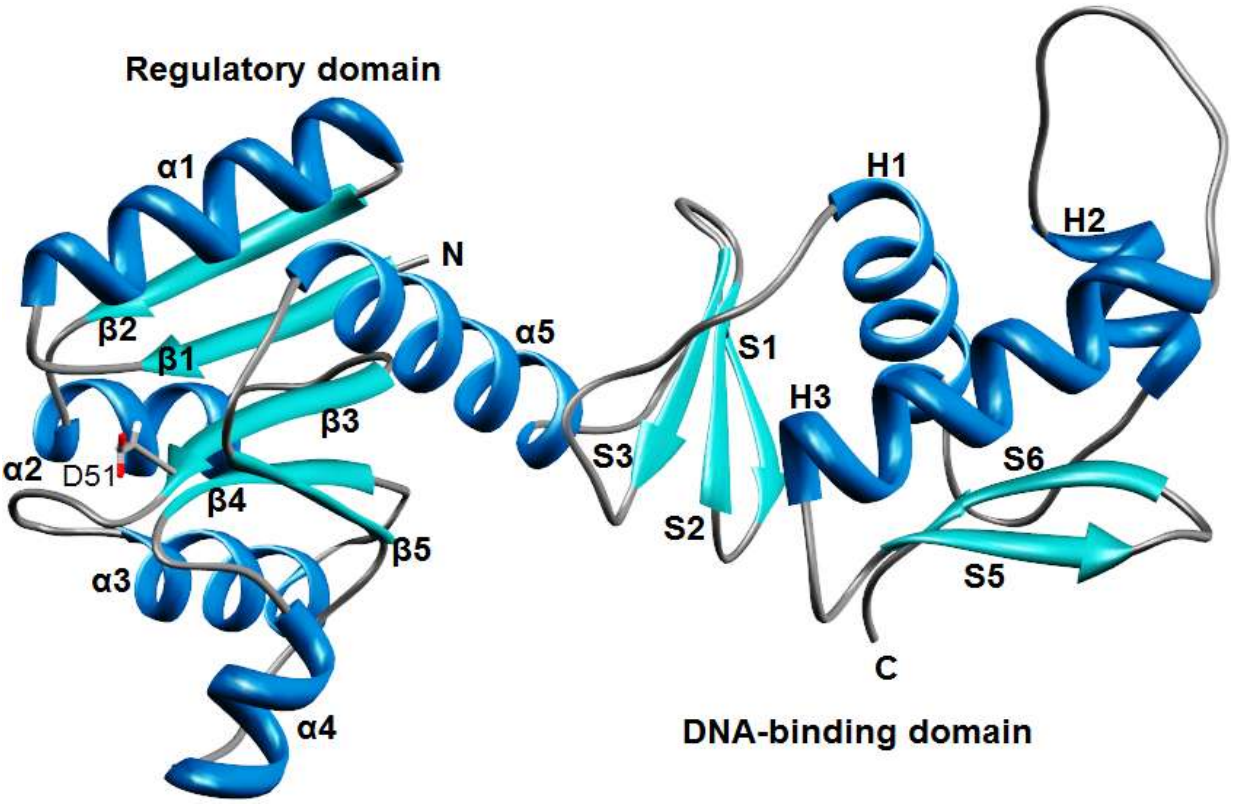


Figure 3.9 Model structure of CopR from *M. aquaeolei* VT8 using the Swiss-Model software [109] and the DrrD RR protein of *T. maritima* (PDB ID: 1KGS) as template. N and C indicate the N- and C-termini of the model, respectively. Secondary structural elements are labelled as α, H (α-helices) and β, S (β-sheets). The putative phosphorylation site, D51 is also labelled and coloured by element. Figure prepared using Chimera software [108].

Before considering the model structure of CopR, one must stress that this is merely a model of the CopR structure, since the primary sequence identity is only 36%, the limit acceptable for structure modelling [117,118] and as such is subject to errors.

The secondary structure, predicted based on the primary sequence of CopR (Figure 3.7), can be compared with the one obtained from the model structure (Figure 3.10). The comparison identified small variations, of about 2 amino acids, between the predictions in respect to the beginning and end of structural elements. As previously stated, predictions are subject to errors and many servers indicate that α-helix prediction is more reliable than predictions of β-sheets. As such, these variations are most probably not significant.

```
        10         20         30         40         50         60         70         80
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL VECPVLVLTA
RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS LRTEEGAEHA LTGTEFRLLR
CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR RGQGYRMFNA QR
```

Figure 3.10 Representation of the predicted secondary structure of CopR from *M.aquaeolei* VT8 from the PROCHECK [109] evaluation of the SWISS-model [109] prediction of the CopR structural model. β-sheet and α-helix predictions are represented in the primary sequence by blue and red, respectively. Amino acid Asp56, which falls into a disallowed region of the Ramachandran plot, is underlined.

A short β-sheet, D97-K101, from the modelling prediction is not present in the bioinformatic prediction. However, this segment contains a structurally conserved amino acid, Y98 which belongs to a stretch of three consecutive hydrophobic residues (Y98, L99 and A100), which supports the proposal that this region is not random coil but indeed structured. The inverse occurs in the short segment from I169-S171. This region, immediately precedes or begins the α2 helix and contains a hydrophobic residue followed by an aromatic and a nucleophilic residue (I169, F170, S171). While these amino acids may form a helical structure, it is more instinctive to accredit them with the formation of a structural "kink" in this region, necessary for positioning helix α2.

It is also important to keep in mind that the structure of CopR is modelled from the RR DrrD from *T. maritima*, and as such the structural elements predicted from homology modelling are heavily derived from this structure. Also, many prediction servers base their predictions on datasets of sequence similar proteins. Therefore, deviations from the typical architecture, which occur in DrrD, such as α4 being roughly perpendicular rather than parallel to the other helices [66], is also observed in the CopR model. Buckler *et al*. [66] hypothesised in 2002 that this orientation of the α4 helix is stabilized by a residue corresponding to Y88 in DrrD. However, of the 269 sequences of putative members of the OmpR/PhoB family, only three members (including DrrD) have hydrophobic or aromatic residues at the position corresponding to Y88 (two with Y and one with I). While 253 have D, E, or N, which is the case of CopR (D86), the remaining are polar (T, Q, and S), alanine (one occurrence), or lysine (one occurrence). Thus, this alternative orientation of α4 is probably uncommon within the OmpR/PhoB family and probably not a major conformation in CopR.

The linker region [119] seems to be equally influenced by the DrrD structure, as the size of the linker is short (5 amino acids) when compared to the linkers of 9 and 15 amino acids of PhoB and OmpR respectively [120]. This is significantly shorter than linkers of typical multidomain response regulators. This may be partly due to the presumed function of the linker in these proteins. The linker region, which is quite flexible in structures of homologs, should function to tether the two domains without significantly contributing to the packing forces that stabilize their interaction in the inactive state. When comparing the secondary structure prediction from Porter [101], as well as the ones from several other programs (though these are merely predictions), the predicted secondary structures are congruent and in this region predict random coil spanning 8 residues. This is one of the least conserved regions of primary sequence in the family, and thus the model may be biased.

Clarification of the structural anomaly presented by the Asp56 residue is possible when the structural model is analyzed in conjunction with the secondary structure and the information gleaned from other structures of the same family. The position occupied by this residue within

the secondary structure is sterically crowded due to the 180° turn that takes place in the space of three amino acids (Pro, Asp, Gly), shown in Figure 3.11.

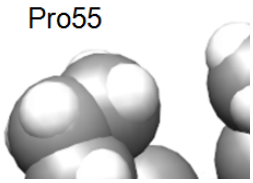


Figure 3.11 Schematic representation of the 180° turn which occurs in the space of three amino acids (Pro55, Asp56 and Gly57) from the model structure of *M. aquaeolei* VT8, CopR. These residues are represented as spheres, which highlights the steric impediments of the Asp56 caused by Pro55. Atoms are coloured by element: carbon in grey, nitrogen in blue, oxygen in red and hydrogens in white. Figure prepared using the Chimera software [108].

Elucidation of the unfavourable conformation adopted by Asp56 is necessary, given the strong conservation for structural reasons. This turn, though unfavourable, is highly conserved throughout bacterial response regulators of this family, so one can assume that it is essential. Also, given the position of this turn, *ca.* 4 residues from the phosphorylation site, these amino acids may be necessary in RR-HK transient interactions, recognition or for phosphorylation. Whichever is the case, the unfavourable conformation adopted by this amino acid must be of importance.

3.2.2 Construction of the Heterologous Expression Plasmid for CopR

Two different plasmids were constructed for CopR expression, pCopR1 and pCopR2. pCopR1 was obtained previously by the author during her B.Sc. thesis and pCopR2 was constructed during this work. Therefore, only for pCopR2 will the cloning process be fully described, while in the case of pCopR1, the work focused on the optimization of the expression conditions and on CopR1 purification (Strategies 1 and 2).

These expression plasmids, though of the pET family, differ in the strategy that was used for their construction. In the first case, pCopR1 was obtained using restriction enzyme cloning, while pCopR2 was obtained using the Ligation Independent Cloning (LIC) approach. Besides the differences in the cloning strategy, these plasmids also differ in the fact that CopR1

is expressed without a fusion tag, while CopR2 is expressed as a N-terminal His_6-tag fusion protein, to enhance its solubility and facilitate purification.

3.2.2.1 Heterologous Expression and Purification of CopR1

A. Verification of *copR* Gene Sequence in pCopR1

As previously stated, the pCopR1 plasmid, represented schematically in Figure 3.12, was constructed beforehand.

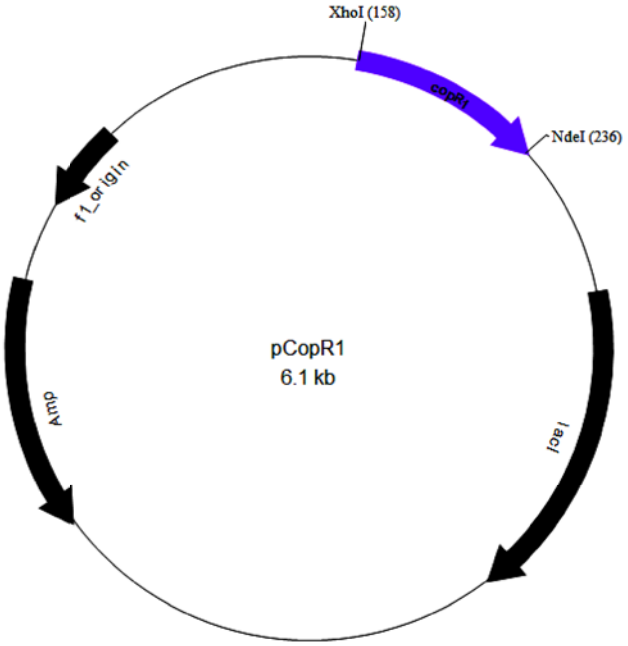


Figure 3.12 Expression plasmid pCopR1 (6.1kb). Legend: *copR1* gene with 666 bp in purple. Cloning at the multiple cloning site, between NdeI (located at 158 bp in pET21-c (+)) and XhoI (located at 236 bp in pET21-c (+)). Image created using BVTech plasmid software, version 3.1 (BV Tech Inc., Bellevue, WA).

However, to ensure that the plasmid used in the studies contained the desired sequence, the plasmid was digested with the restriction enzymes used for cloning to verify that the *copR1* gene, of 666 bp, was inserted into the expression vector pET21-c (+) (map of the vector in Appendix B.1 - Figure 6.2).

To this end, a non-expression host, *E.coli* Giga Blue, was transformed with this plasmid and 4 of the colonies obtained were used to inoculate LB medium supplemented with 100 µg/mL ampicillin for plasmid amplification and isolation. The extracted plasmids were digested with the restriction enzymes, NdeI and XhoI, in order to evaluate insertion of the DNA fragment containing *copR* into the expression plasmid (Figure 3.13).

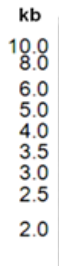


Figure 3.13 Clone screening for insertion of the desired DNA fragment, *copR1* of approximately 750 bp. Legend: **M**: 1 kb DNA ladder; **Lanes 1 to 4:** Digestion of the isolated plasmid DNA from clones 1 - 4 obtained from the transformation of *E.coli* Giga Blue with pCopR1. The plasmid DNA was digested with NdeI and XhoI as described in Chapter 2 Materials and Methods. The linear form of the pET 21-c (+) vector appears as a band between 5 and 6 kb. 1% agarose gel, electrophoresis at 100 mV for approximately, 20 min. Gel was stained with SybrSafe and visualized under UV light.

Upon analysis of Figure 3.13 it seems that all the digested clones have a DNA fragment insert of approximately 750 bp, which corresponds approximately to the expected length of *copR1*, 663 bp. Also, the linear form of the plasmid without the gene is observed between 5 and 6 kb, with the expected length of 5.1 kb (the resolution in this part of the gel is low, due to the 1% agarose gel used).

Examination of the DNA sequence, obtained by sequencing the chosen expression vector (Clone 4 of Figure 3.13), confirmed that the inserted DNA fragment shares 98% identity with maqu_0124 from *M. aquaeolei* VT8, and the protein sequence shares 100% identity, when an alignment was performed with ClustalW (Figure 3.14).

```
                 M  R  L  L  L  V  E  D  D  R  L  L  A  E  G  L  V  R  Q  L  E  K  A
copR1            ATGCGTTTATTGCTCGTTGAAGACGACCGTTTGCTGGCCGAGGGGCTGGTCCGCCAGCTGGAAAAAGCG 69
maqu_0124        ATGCGTTTATTGCTCGTTGAAGACGACCGTTTGCTGGCCGAGGGGCTGGTCAGACAGCTGGAAAAAGCG 69
                 *************************************************.*.***************

                 G  F  S  I  D  H  T  S  S  A  R  E  A  Q  I  L  G  E  Q  E  D  Y  R
copR1            GGGTTTAGCATCGACCACACGTCCAGTGCCCGTGAGGCCCAGATTCTGGGGGAGCAGGAGGATTATCGT 126
maqu_0124        GGGTTTAGCATCGACCACACGTCCAGTGCCCGTGAGGCCCAGATTCTGGGGGAGCAGGAGGATTATCGT 126
                 ************************************************************************

                 A  A  V  L  D  L  G  L  P  D  G  N  G  L  E  V  L  K  R  W  R  S  K
copR1            GCCGCTGTTCTCGATCTCGGCCTGCCGGATGGCAACGGGCTGGAGGTTTTGAAACGATGGAGATCGAAG 189
maqu_0124        GCCGCTGTTCTCGATCTCGGCCTGCCGGATGGCAACGGGCTGGAGGTTTTGAAACGATGGCGATCGAAG 189
                 ************************************************************.********

                 L  V  E  C  P  V  L  V  L  T  A  R  G  D  W  Q  D  K  V  N  G  L  K
copR1            CTCGTCGAGTGCCCGGTATTGGTCCTCACCGCCAGAGGCGACTGGCAGGACAAGGTCAATGGACTGAAA 252
maqu_0124        CTCGTCGAGTGCCCGGTATTGGTCCTCACCGCCAGAGGCGACTGGCAGGACAAGGTCAATGGACTGAAA 252
                 ************************************************************************

                 A  G  A  D  D  Y  L  A  K  P  F  Q  T  E  E  L  I  A  R  I  N  A  L
copR1            GCAGGGGCGGATGACTATCTGGCGAAACCGTTCCAGACAGAAGAACTGATCGCCCGCATCAATGCACTC 315
maqu_0124        GCAGGGGCGGATGACTATCTGGCCAAACCGTTCCAGACAGAAGAACTGATCGCCCGCATCAATGCACTC 315
                 ************************************************************************

                 I  R  R  S  E  G  R  V  H  S  Q  V  K  A  G  G  F  E  L  D  E  N  R
copR1            ATACGCCGCAGCGAAGGGCGAGTGCACTCCCAGGTTAAAGCCGGTGGCTTCGAACTGGACGAAAATCGC 378
maqu_0124        ATACGCCGCAGCGAAGGGCGAGTGCACTCCCAGGTTAAAGCCGGTGGCTTCGAACTGGACGAAAATCGC 378
                 ************************************************************************

                 Q  S  L  R  T  E  E  G  A  E  H  A  L  T  G  T  E  F  R  L  L  R  C
copR1            CAGAGCCTGCGGACAGAAGAAGGAGCAGAACACGCCCTGACGGGTACTGAGTTCCGACTGCTACGATGC 441
maqu_0124        CAGAGCCTGCGGACAGAAGAAGGAGCAGAACACGCCCTGACGGGTACTGAGTTCCGCCTGCTACGATGC 441
                 ************************************************************.***********

                 L  M  S  R  P  G  H  I  F  S  K  E  Q  L  M  E  Q  L  Y  N  L  D  E
copR1            CTGATGAGCCGTCCGGGCCACATCTTTTCCAAGGAACAGCTCATGGAGCAGCTATACAACCTGGATGAG 504
maqu_0124        CTGATGAGCCGTCCGGGCCACATCTTTTCCAAGGAACAGCTAATGGAGCAGCTATACAACCTGGATGAG 504
                 ****************************************.***************************

                 S  P  S  E  N  V  I  E  A  Y  I  R  R  L  R  K  L  V  G  N  E  T  I
copR1            AGCCCCAGCGAAAACGTGATTGAGGCGTATATTCGGCGCTTGAGAAAGCTGGTCGGCAACGAAACGATA 567
maqu_0124        AGCCCCAGCGAAAACGTGATTGAGGCGTATATTCGGCGCTTGAGAAAGCTGGTCGGCAACGAAACGATC 567
                 ************************************************************************.

                 T  T  R  R  G  Q  G  Y  M  F  N  A  Q  R  -
copR1            ACCACACGCCGTGGCCAGGGATACATGTTCAATGCCCAACGTTAA 666
maqu_0124        ACCACACGCCGTGGCCAGGGATACATGTTCAATGCCCAACGTTAA 666
                 *********************************************
```

Figure 3.14 Alignment of the DNA sequences, and translated primary sequence, obtained by sequencing the selected vector, against the reference gene *M. aquaeolei,* Maqu_0124. Legend: Asterisks (*) and stops (.) below the sequence indicate identity or conservation of the amino acids, respectively. Alignment was performed using ClustalW [96].

The expression plasmid, clone 4, which was sent to DNA sequencing was then used to transform the expression host *E.coli* BL21(DE3), as detailed in Chapter 2 Materials and Methods.

B. Heterologous Test Expressions and Solubility Assays

Various growth conditions were tested to determine the optimal expression conditions for CopR1. In addition to the work presented in this thesis, there was prior work performed with this clone which had shown that expression in LB medium, and inducing expression with 0.5 mM

IPTG at an OD_{600nm} of approximately 0.7, during 2 h at 37°C, produced CopR in the form of inclusion bodies, which were unstable after refolding using the dialysis approach.

In an attempt to circumvent this additional problem, some of the data retrieved from those studies, such as, shorter expression times and lower concentrations of IPTG (0.1 mM) for induction at 0.6 OD_{600nm} were used. In this work, other factors were tested, such as lower temperature of growth upon induction. Since it had been previously observed that there was basal expression of this gene, this condition was also tested on growths without induction for 2 to 16 hours. Relative to growth media, the primary growth medium used for cell growth continued to be the rich LB medium. The detailed growth conditions that were tested are listed in Chapter 2 Materials and Methods – Section 2.2.5.2, Table 2.3.

Since the aim was to determine the expression conditions in which CopR1 was soluble, a screening method for the ideal expression conditions for CopR1 solubility was a priority, and as such the growths were subjected to the BugBuster protocol (Chapter 2 Materials and Methods - Section 2.5.2.2). These samples were then analyzed by SDS-PAGE, which clearly shows that the optimal CopR1 expression occurs at shorter growth times and at a lower growth temperature, 20°C *vs.* RT after induction (Figure 3.15).

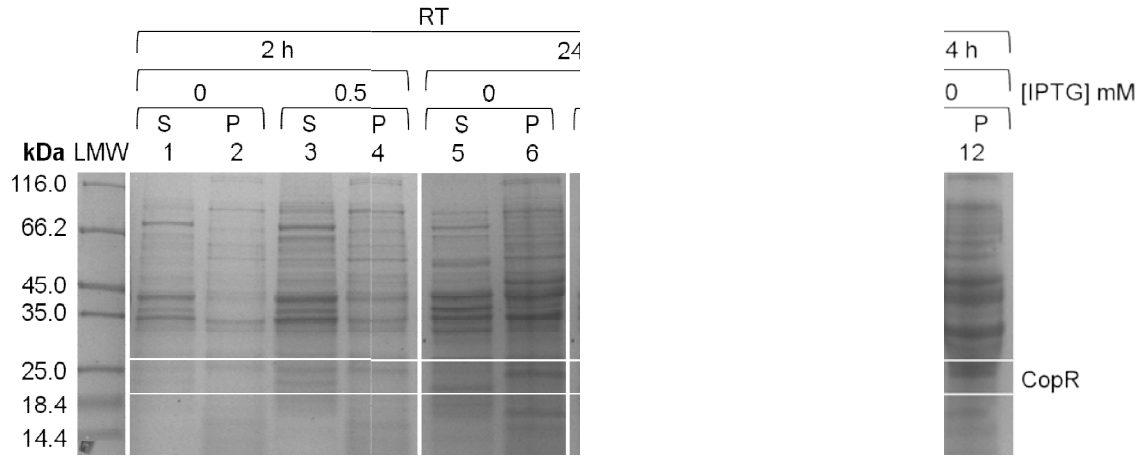


Figure 3.15 Solubility test of CopR1 (25 kDa) obtained under different growth conditions. Legend: **LMW:** Low molecular weight marker; **Lane 1 and 3:** Soluble extracts of cells grown at RT for 2 h, without and with 0.5 mM IPTG induction, respectively; **Lane 5 and 7:** Soluble extracts of cells grown at RT for 24 h without and with 0.5 mM IPTG induction, respectively; **Lane 9:** Soluble extract of cells grown at 20°C for 2 h with 0.5 mM induction; **Lane 11:** Soluble extract of cells grown at 20°C for 24 h without induction; **Lanes 2, 4, 6, 8, 10 and 12:** Pellet of the cells in the lanes to their left. The S and P above the lanes identify the soluble and pellet fractions of the growths, respectively. The boxed region corresponds to the expected electrophoretic mobility of a protein of 25 kDa. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

Expression of CopR1, which appears as a band at 25 kDa, is not independent of the expression conditions. As one can see from Figure 3.15, expression of CopR1 is quite varied, ranging from insignificant, lanes 1/2, 3/4 or 11/12; to inclusion bodies in the pellet, lanes 5/6, 7/8

or a respectable amount, the majority of which is present in the soluble extract, when compared to the expression of *E.coli* genes (35 or 45 kDa), lanes 9/10.

Again, in this work it is shown that at lower expression temperatures, expression of CopR1 occurs within the first hours of induction, after which it degradation seems to occur. Also, expression without induction at higher temperatures, lanes 1/2 and 5/6 from Figure 3.15, produces insufficient quantity or insoluble forms of CopR1. The best expression conditions to obtain soluble CopR1 seems to be at 20°C after induction. In fact, the majority of the soluble protein is to be found in the soluble extract of growths performed at 20°C for 2 h upon induction with 0.5 mM IPTG (Figure 3.15, lane 9).

As a caveat, it should be mentioned that the degree of cell lysis when using the BugBuster protocol varies. As a result, CopR1 present in the pellet can be either in inclusion bodies and in part in the soluble form, considering that the *E.coli* cells might not be completely disrupted.

Therefore, after this analysis the growth conditions for heterologous expression of CopR1 were considered to be LB medium at 20°C for 2 h after induction with 0.5 mM IPTG at OD_{600nm} of 0.6 (Figure 3.16).

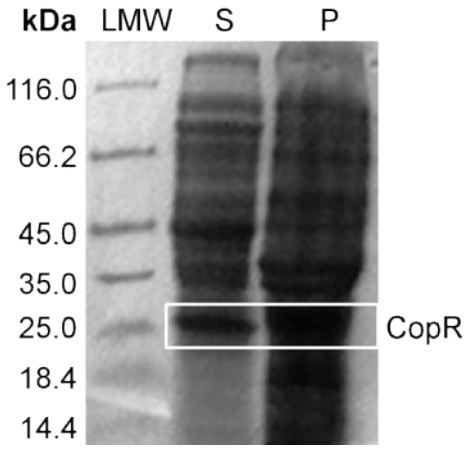


Figure 3.16 SDS-PAGE of the soluble extract and insoluble fraction of cells obtained from over-expression of CopR1 (growth conditions: induction at OD_{600nm} at 0.6 with 0.5 mM IPTG and expression for 2 h at 20°C). Legend: **LMW:** Low molecular weight marker; **Lane S:** Soluble extract of cells transformed with pCopR1 obtained after ultracentrifugation of the disrupted cells; **Lane P:** Pellet of cells transformed with pCopR1 obtained after ultracentrifugation of the disrupted cells. The boxed region indicates CopR1 (25 kDa). SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

C. Purification of CopR1

E.coli is a bacteria and as such does not have a nucleus, but, surrounding the cytoplasm is a discreet space, the periplasm. CopR is a cytoplasmatic protein and upon cell lysis, the soluble extract contains not only CopR1 but all the soluble proteins present in the cytoplasm, as well as in the periplasm. In order to purify CopR1 from the soluble extract two strategies were used, as detailed in Table 3.2:

Table 3.2 Strategies employed for CopR1 purification.

CopR1	
Strategy 1	**Strategy 2**
Recombinant protein 0.5 mM IPTG, 2 h, 20˚C, 210 rpm Anionic exchange Gel filtration	Recombinant protein 0.5 mM IPTG, 2 h, 20˚C, 210 rpm Anionic exchange Anionic exchange

The soluble extracts used for each strategy were obtained independently. For Strategy 1 the growths, used for the previous BugBuster protocol, which had soluble CopR1 were pooled, centrifuged and employed as the soluble extract. For Strategy 2, the soluble extract from cells grown in the optimal expression conditions found, 20°C for 2 h after induction with 0.5 mM IPTG, were used.

C1. Protein Purification – Strategy 1

In Strategy 1, the obtained soluble extract was loaded onto an anionic exchange column with a DEAE-52 matrix, since CopR1 has a pI of approximately 6.0, and thus is negatively charged at pH 7.6. The fractions were analysed by SDS-PAGE and those containing CopR1 were pooled (Figure 3.17B, Lane A), concentrated and loaded onto a gel filtration chromatography column (Superdex 75). After this second chromatographic step, two fractions were obtained, Lanes B and C in Figure 3.17 B.

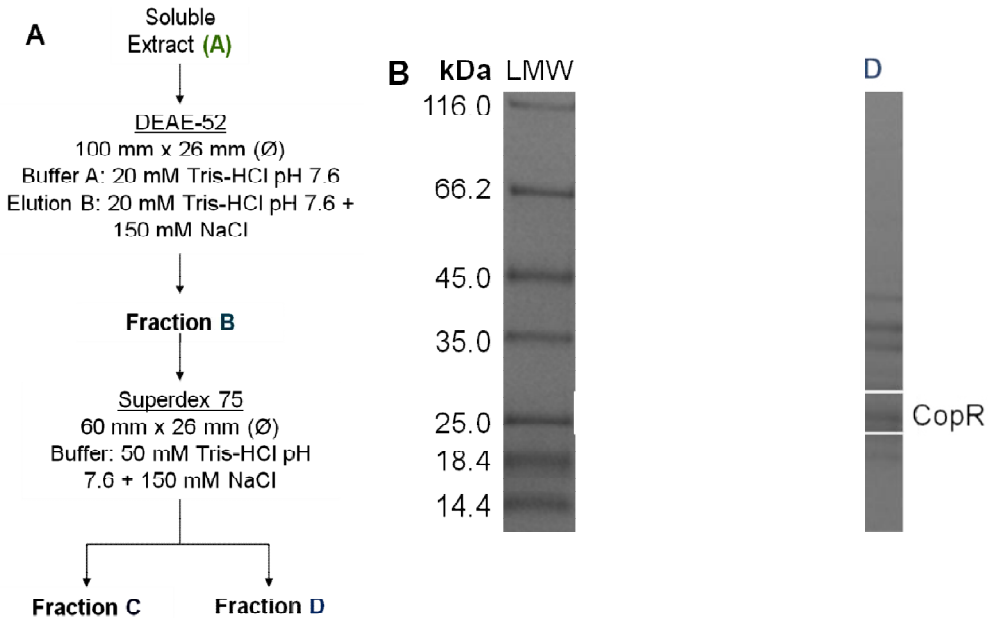


Figure 3.17 A) Flow-chart of the purification of CopR1 (25 kDa) using Strategy 1. **B)** SDS-PAGE of the fractions obtained throughout the purification using Strategy 1. **Lanes A, B and C**: Samples of the fractions identified in A. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

However, in neither of these fractions was CopR1 considered pure to be used for biochemical characterization, and the amount of CopR1 present, when compared to that of the contaminants, precluded further purification.

C2. Protein Purification – Strategy 2

In the second strategy used for CopR1 purification, the soluble cell extract (Figure 3.18B, Lane A) was loaded onto an anionic-exchange chromatography, DEAE-52 matrix. The fractions obtained were analyzed using SDS-PAGE, and the ones containing CopR1 were pooled into fractions B and C and concentrated (Figure 3.18, Lanes B and C). Fraction B was buffer exchanged and loaded onto another anionic exchange chromatography column, Resource-Q. However, similarly to what was observed in Strategy 1, the resulting fraction, D (Figure 3.18B, Lane D), did not present the purity required for biochemical characterization. The chromatograms and gels for the purification steps of both strategies are presented in Appendix C.12, Figures 6.15 – 6.17.

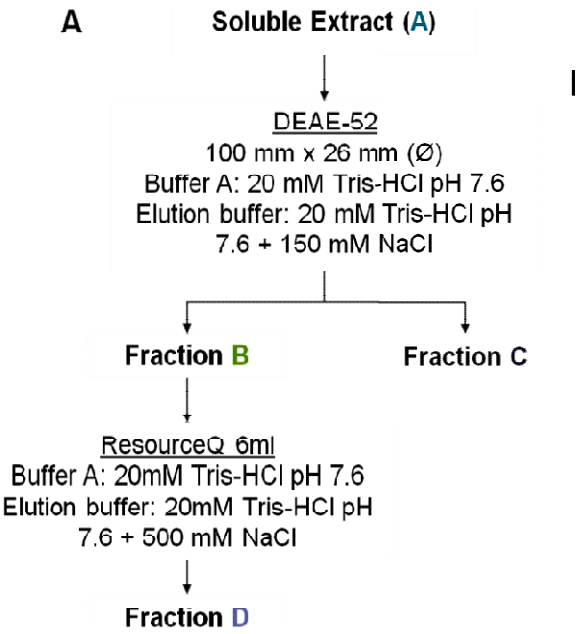


Figure 3.18 A) Flow-chart of the purification of CopR1 (25 kDa) using Strategy 2. **B)** SDS-PAGE of the fractions obtained throughout the purification using Strategy 2. **Lanes A, B, C and D**: Samples of the fractions identified in A. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In conclusion, both strategies used in the purification of CopR1 were not efficient, resulting in impure samples with a very low yield in CopR1 protein. Therefore, a different strategy was chosen for protein purification, the use of fusion tag. In this case a His_6-tag located at the N-terminus of CopR enables the use of affinity chromatography in the purification of this

protein. In addition, as mentioned in the beginning of this Chapter, a different cloning approach will be used, one that is ligation independent – the LIC method.

3.2.2.2 Heterologous Expression and Purification of CopR2

A. Construction of Expression Vector

The primers listed in Table 2.1 (Chapter 2 Materials and Methods - Section 2.2) were used in a PCR to amplify the DNA fragment which encodes the putative CopR2 protein. The primers for the *copR2* gene did not include restriction sites, given its posterior insertion into expression vector pET-30 EK/LIC, using the LIC approach.

Figure 3.19 PCR amplification of the *copR2* gene using the primers and PCR program described in Chapter 2 Materials and Methods – Section 2.2. Legend: **M**: DNA ladder of 1 kb; **Lane 1**: Amplified PCR product of approximately 750 bp indicated by the arrow. 1% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

PCR amplification of the *copR2* gene (663 bp) resulted in a fragment of approximately 750 bp (Figure 3.19), which was subsequently purified and cloned into the pET-30 EK/LIC expression vector (map of the vector in Appendix B.2 - Figure 6.3), and constituted the expression vector depicted schematically in Figure 3.20.

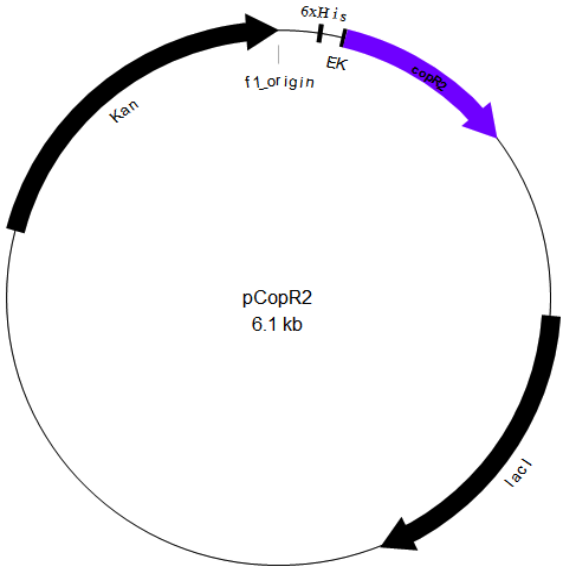


Figure 3.20 Cloning vector pCopR2 (6.1 kb). Legend: *copR2* gene with 663 bp in purple. Cloning at the LIC site into pET-30 EK/LIC. Image created using BVTech plasmid, version 3.1 (BV Tech Inc., Bellevue, WA).

The resulting 6.1 kb expression vector, pCopR2, was transformed into a non-expression host, *E.coli* Giga Blue, and 4 of the colonies obtained were used to inoculate LB medium supplemented with 25 µg/mL kanamycin for plasmid amplification and isolation (Figure 3.21).

Figure 3.21 Clone screening. Legend: **M**: DNA ladder of 1 kb; **Lanes 1 to 4**: Plasmid DNA isolated from clone 1 - 4 colonies obtained from the transformation of *E.coli* Giga Blue with pCopR2. The arrow indicates the linear form of the plasmid DNA, of approximately 6 kb. Clones 3 and 4 were sent for DNA sequencing to check the DNA sequence of the inserted fragment. 0.8% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

Figure 3.21 shows the three forms of the expression vector pCopR2, the nicked circle (migrates with apparent molecular mass of 8 kb), linear (approximately 6 kb) which is signalled with an arrow and supercoiled (migrates with an apparent molecular mass between 4 and 5 kb).

Analysis of the DNA sequence obtained by sequencing the chosen plasmids (clones 3 and 4 from Figure 3.21), confirmed that the DNA sequence shares 99 and 100% identity and that the translated protein sequences for this DNA fragment and both share 100% identity with *M. aquaeolei* VT8, Maqu_0124 (Figure 3.22). The plasmid DNA from clone 4 100% DNA identity was used to transform the expression host *E.coli* BL21(DE3), as detailed in Chapter 2 Materials and Methods.

```
                                                                          74
Maqu_0124   MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQEDYRAAVLDLGLPDGNGLEVLKRWRSKLVECP
CopR2       MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQEDYRAAVLDLGLPDGNGLEVLKRWRSKLVECP
            ************************************************************************

                                                                         148
Maqu_0124   VLVLTARGDWQDKVNGLKAGADDYLAKPFQTEELIARINALIRRSEGRVHSQVKAGGFELDENRQSLRTEEGAE
CopR2       VLVLTARGDWQDKVNGLKAGADDYLAKPFQTEELIARINALIRRSEGRVHSQVKAGGFELDENRQSLRTEEGAE
            ************************************************************************

                                                                     221
Maqu_0124   HALTGTEFRLLRCLMSRPGHIFSKEQLMEQLYNLDESPSENVIEAYIRRLRKLVGNETITTRRGQGYMFNAQR
CopR2       HALTGTEFRLLRCLMSRPGHIFSKEQLMEQLYNLDESPSENVIEAYIRRLRKLVGNETITTRRGQGYMFNAQR
            ************************************************************************
```

Figure 3.22 Alignment of the amino acid sequences obtained by translation of the sequenced selected plasmids (ExPASy translate tool [94]), *copR2*, against the reference protein from *M. aquaeolei* VT8, Maqu_0124. Alignment was performed using ClustalW [96]. Legend: Asterisks (*) below the sequence indicate identity of the amino acids.

B. Heterologous Test Expressions and Solubility Assays

In order to establish the growth conditions for maximum protein production, different lengths of induction, 0 to approximately 16 hours were tested, using 0.1, 0.5 and 1.0 mM IPTG for induction, at 16 or 37°C, as the *copR2* gene is under the control of a T7 *lac* promoter (Appendix B.2, Figure 6.3. B). Cell growth was performed in small scale (50 mL), as described in Chapter 2 Material and Methods, with a typical expression profile in LB media, as exemplified by Figures 3.23 and 3.24.

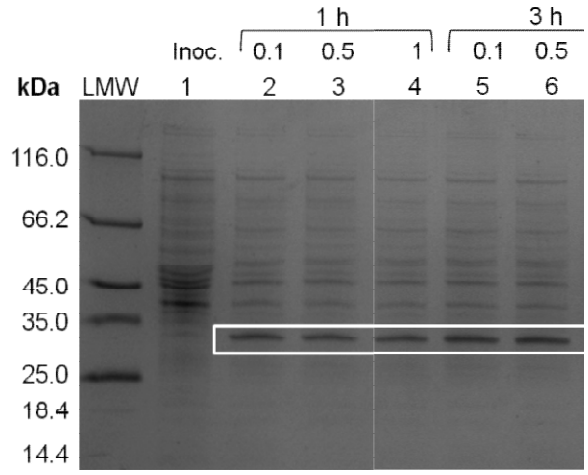


Figure 3.23 SDS-PAGE analysis of the expression profile of cells transformed with pCopR2 in LB growth medium, at 37°C for CopR2 expression (31 kDa). Legend: **LMW**: Low molecular weight marker; **Lane 1**: Inoculum; **Lane 2**: 1 h growth of cells after induction with 0.1 mM IPTG; **Lane 3**: 1 h growth of cells after induction with 0.5 mM IPTG; **Lane 4**: 1 h growth of cells after induction with 1 mM IPTG; **Lane 5**: 3 h growth of cells after induction with 0.1 mM IPTG; **Lane 6**: 3 h growth of cells after induction with 0.5 mM IPTG; **Lane 7**: 3 h growth of cells after induction with 1 mM IPTG; **Lane 8**: 5 h growth of cells after induction with 0.1 mM IPTG; **Lane 9**: 5 h growth of cells after induction with 0.5 mM IPTG; **Lane 10**: 5 h growth of cells after induction with 1 mM IPTG; **Lane 11**: 16 h growth of cells after induction with 0.1 mM IPTG; **Lane 12**: 16 h growth of cells after induction with 0.5 mM IPTG; **Lane 13**: 16 h growth of cells after induction with 1 mM IPTG. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

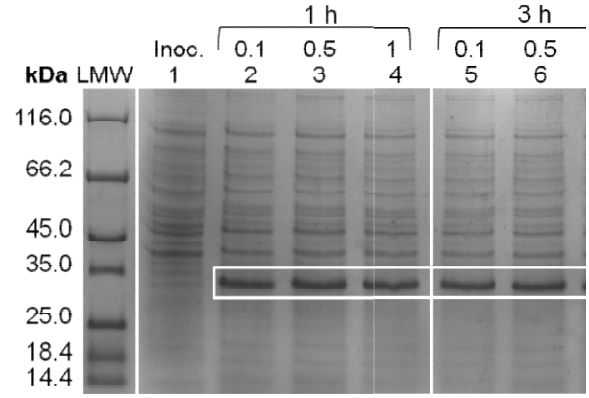


Figure 3.24 SDS-PAGE analysis of the expression profile of cells transformed with pCopR2 in LB growth medium, at 16°C for CopR2 expression (31 kDa). Legend: **LMW:** Low molecular weight marker; **Lane 1**: Inoculum; **Lane 2**: 1 h growth of cells after induction with 0.1 mM IPTG; **Lane 3**: 1 h growth of cells after induction with 0.5 mM IPTG; **Lane 4**: 1 h growth of cells after induction with 1 mM IPTG; **Lane 5**: 3 h growth of cells after induction with 0.1 mM IPTG; **Lane 6**: 3 h growth of cells after induction with 0.5 mM IPTG; **Lane 7**: 3 h growth of cells after induction with 1 mM IPTG; **Lane 8**: 5 h growth of cells after induction with 0.1 mM IPTG; **Lane 9**: 5 h growth of cells after induction with 0.5 mM IPTG; **Lane 10**: 5 h growth of cells after induction with 1 mM IPTG; **Lane 11**: 16 h growth of cells after induction with 0.1 mM IPTG; **Lane 12**: 16 h growth of cells after induction with 0.5 mM IPTG; **Lane 13**: 16 h growth of cells after induction with 1 mM IPTG. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In order to analyze these results one needs to take into account that CopRHis$_6$ (CopR2) is expected to have a molecular mass of 31 kDa (Appendix C.13.1, Figure 6.18), since the molecular mass of CopR is 25 kDa and the additional tags included in the protein due to the cloning procedure have a total molecular mass of 6 kDa. Therefore, since the amount of cells

loaded onto each of the gels was normalized, it is possible to conclude that CopRHis$_6$ is the predominant protein in *E.coli* cells grown at either 16°C or 37ºC, when compared with the native *E.coli* proteins. The comparison between the two temperatures is not conclusive, as all the bands for cells transformed with pCopR2 at 16/37ºC seem equally intense.

Analysis of Figures 3.23 and 3.24 appears to show that, as was verified in previous work, degradation of CopRHis$_6$ occurs when the cells are grown for 16 h after induction, an observation which is more pronounced when growths are performed at 37°C. The CopRHis$_6$ protein band intensity almost does not change, though *E.coli* protein bands are more intense (Lanes 11-13, Figure 3.23). This can be attributed to the fact that CopR is a transcription factor, and thus a usual target for rapid degradation.

The second stage of this analysis is to determine in which growth conditions the CopRHis$_6$ is present in the soluble fraction. As was referenced before this protein is unstable, so, before advancing to large scale over-expression of CopRHis$_6$, 1 mL of each growth condition tested was subjected to a lysis buffer as described in Chapter 2 Materials and Methods – Section 2.2.5.2, using the BugBuster protocol. These samples were then analyzed by SDS-PAGE (Figures 3.25 and 3.26) that clearly shows that optimal CopRHis$_6$ expression as a soluble protein occurs at shorter growth times, independent of the studied temperatures.

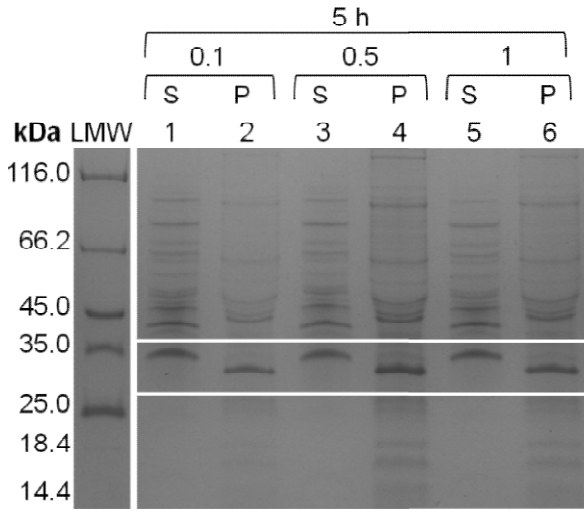


Figure 3.25 Solubility of CopRHis$_6$ (31 kDa) in LB medium under various conditions, at 37°C. Legend: **LMW:** Low molecular weight marker; **Lane 1:** Soluble extract of cells grown for 5 h after 0.1 mM IPTG induction; **Lane 2:** Pellet of cells grown for 5 h after 0.1 mM IPTG induction; **Lane 3:** Soluble extract of cells grown for 5 h after 0.5 mM IPTG induction; **Lane 4:** Pellet of cells grown for 5 h after 0.5 mM IPTG induction; **Lane 5:** Soluble extract of cells grown for 5 h after 1 mM IPTG induction; **Lane 6:** Pellet of cells grown for 5 h after 1 mM IPTG induction; **Lane 7:** Soluble extract of cells grown for 16 h after 0.1 mM IPTG induction; **Lane 8:** Pellet of cells grown for 16 h after 0.1 mM IPTG induction; **Lane 9:** Soluble extract of cells grown for 16 h after 0.5 mM IPTG induction; **Lane 10:** Pellet of cells grown for 16 h after 0.5 mM IPTG induction; **Lane 11:** Soluble extract of cells grown for 16 h after 1 mM IPTG induction; **Lane 12:** Pellet of cells grown for 16 h after 1 mM IPTG induction. Above the lane identifiers is the concentration of IPTG used for induction as well as length of induction. The S and P above lanes correspond to the soluble and pellet fractions. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

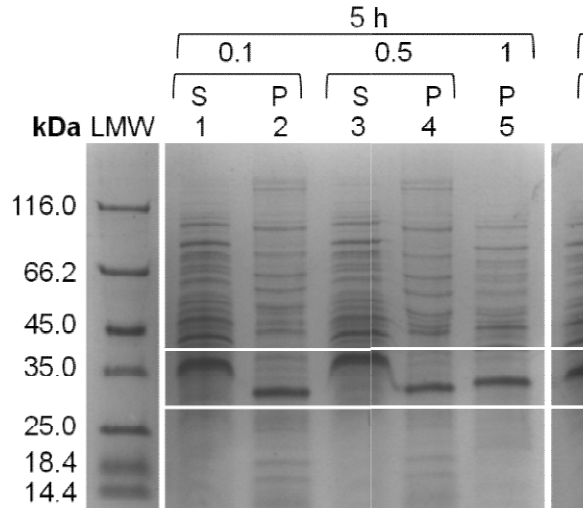


Figure 3.26 Solubility of CopRHis$_6$ (31 kDa) in LB media under various conditions, at 16°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Soluble extract of cells grown for 5 h with 0.1 mM IPTG induction; **Lane 2**: Pellet of cells grown for 5 h with 0.1 mM IPTG induction; **Lane 3**: Soluble extract of cells grown for 5 h with 0.5 mM IPTG induction; **Lane 4**: Pellet of cells grown for 5 h with 0.5 mM IPTG induction; **Lane 5**: Pellet of cells grown for 5 h with 1 mM IPTG induction; **Lane 6**: Soluble extract of cells grown for 16 h with 0.1 mM IPTG induction; **Lane 7**: Pellet of cells grown for 16 h with 0.1 mM IPTG induction; **Lane 8**: Soluble extract of cells grown for 16 h with 0.5 mM IPTG induction; **Lane 9**: Pellet of cells grown for 16 h with 0.5 mM IPTG induction; **Lane 10**: Soluble extract of cells grown for 16 h with 1 mM IPTG induction; **Lane 11**: Pellet of cells grown for 16 h with 1 mM IPTG induction. Above the lane identifiers is the concentration of IPTG used for induction as well as length of induction. The S and P above lanes correspond to the soluble and pellet fractions. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In all growths, at 37ºC (Figure 3.25) and 16ºC (Figure 3.26), CopRHis$_6$ is present in the soluble fraction. Though, its electrophoretic mobility is different from that observed in Figures 3.23 and 3.24, and also different from CopRHis$_6$ present in inclusion bodies, migrating as a protein of higher molecular mass (Figures 3.25 and 3.26, Lanes identified with pellet).

Since previously it was shown that higher temperature growth was not so successful in obtaining soluble CopR, the conditions that were chosen to perform large scale growth were: induction at OD_{600nm} of 0.6, with 0.1 mM IPTG and growth for 5 or 16 h at 16ºC in 500 mL medium. Both growths were then subjected to cell lysis and clarification by ultracentrifugation. Samples of the soluble extracts and pellets from these two growths were run on SDS-PAGE to verify the solubility of the fusion protein, CopRHis$_6$ (Figure 3.27).

The CopRHis$_6$ is present in the soluble extract of both conditions, with almost no protein being present in the pellet, and the relative amount of CopRHis$_6$ is comparable in the 16h growth.

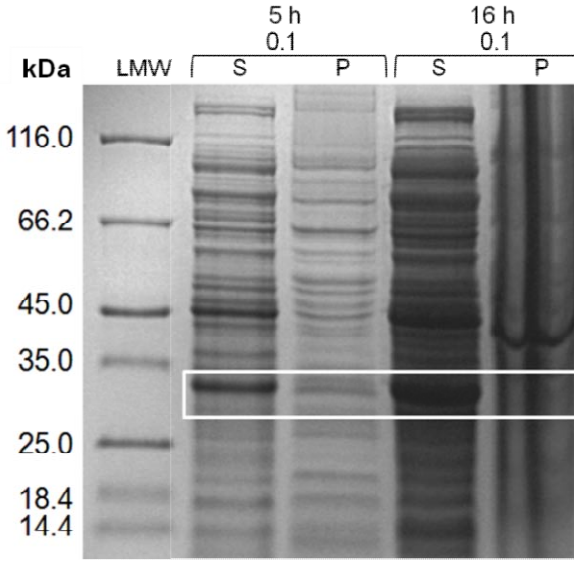


Figure 3.27 Solubility of CopRHis$_6$ (31 kDa) in LB media upon induction with 0.1 mM IPTG, at 16°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Soluble extract of cells grown for 5 h with 0.1 mM IPTG induction; **Lane 2**: Pellet of cells grown for 5 h with 0.1 mM IPTG induction; **Lane 3**: Soluble extract of cells grown for 16 h with 0.1 mM IPTG induction; **Lane 4**: Pellet of cells grown for 16 h with 0.1 mM IPTG induction. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

C. Purification of CopRHis$_6$

CopRHis$_6$ is present in the soluble extract of both the previous growths, thus these two fractions were pooled and centrifuged before protein purification. As mentioned, CopRHis$_6$ contains a N-terminal His$_6$-tag, which simplifies purification, enabling separation by affinity chromatography, using a Ni^{2+}-Sepharose matrix (His-Trap FF).

After loading the soluble extract, the unbound proteins are washed, similarly to an ionic exchange chromatography in order to elute the proteins that do not bind to the matrix, and then eluted in steps of imidazole (20 – 500 mM), which competes with the proteins bound to the matrix (through their His-tag) for the Ni^{2+} in the matrix, since imidazole is homologous of the histidine side chain. All the fractions that were collected were analyzed by SDS-PAGE to determine which contained CopRHis$_6$ (Figure 3.28).

Analysis of the flow-through sample, Lane 3 of Figure 3.28, indicates column saturation as there seems to be desired protein in this fraction, however, there is also an *E.coli* protein of a similar molecular weight expressed in these conditions.

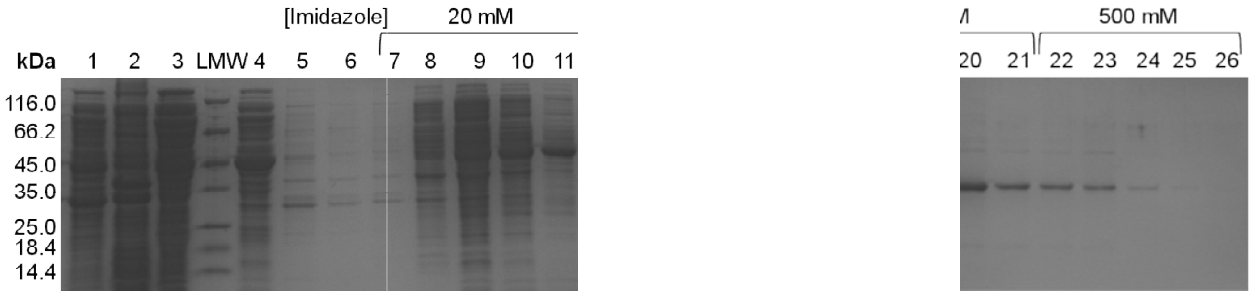


Figure 3.28 Purification of the CopRHis$_6$ fusion protein (31 kDa) by Ni^{2+}-sepharose affinity chromatography. Legend: **Lane1**: Sample of the soluble extract loaded onto the column; **Lane 2**: Resuspended pellet from centrifugation; **Lane 3**: Column flow-through during sample application; **LMW**: Low molecular weight marker; **Lanes 4 to 6**: Samples of fractions after elution with 5, 25 and 50 mL column washing with 20 mM Tris-HCl (pH 7.6), 500 mM NaCl; **Lanes 7 to 11**: Samples of fractions after elution with 4, 8, 12, 16 and 20 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 20 mM Imidazole; **Lanes 12 to 16**: Samples of fractions after elution with 4, 8, 12, 16 and 20 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 100 mM Imidazole; **Lanes 17 to 21**: Samples of fractions after elution with 4, 8, 12, 16 and 20 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 200 mM Imidazole; **Lanes 22 to 26**: Samples of fractions after elution with 4, 8, 12, 16 and 20 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 500 mM Imidazole. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In this case, the pooled fractions were the ones corresponding to the samples loaded into lanes 16 through to 24 (Fraction B, higher purity) and the ones between 14 and 15, constituted Fraction C (lower purity) (Figure 3.29B). However, it is important to point out that these are the results of one of the three growths that were performed during the studies presented here. As in this case, two CopRHis$_6$ fractions were obtained, which correspond to two different purity indexes, one with a higher and another with a lower purity index as shown in Figure 3.29B, Lanes B and C.

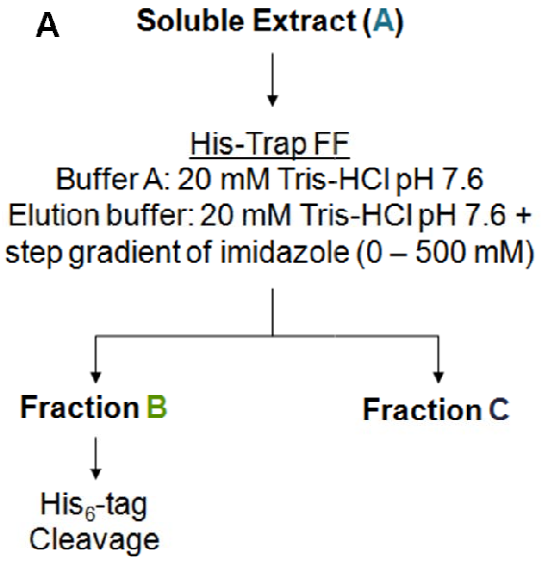


Figure 3.29 A) Flow-chart of the purification of the CopRHis$_6$ fusion protein, CopR2 (31 kDa). **B)** SDS-PAGE of the samples from fractions obtained using the strategy in the flow-chart. **Lanes A, B and C**: Samples of the fractions identified in A. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

As the resulting samples from strategies 1 and 2, for purification of the CopR1 protein were not of sufficient purity, their protein yield was not calculated, but that of Strategy 3 for the purification of CopR2 was, Table 3.3.

Table 3.3 Strategies utilized for the purification of the CopR protein (25 kDa) from *M.aquaeolei* VT8. Not listed are the individual expression conditions tested for Strategies 1, 2 and 3, which are explained in detail in previous sections. Strategy 1 and 2 differ only in method of purification whilst strategy 3 relied on LIC instead of restriction enzymes as a cloning methodology.

	CopR1		CopR2
	Strategy 1	**Strategy 2**	**Strategy 3**
	• Recombinant protein • 0.5 mM IPTG, 2 h, 20˚C, 210 rpm • Anionic exchange • Gel filtration	• Recombinant protein • 0.5 mM IPTG, 2 h, 20˚C, 210 rpm • Anionic exchange • Anionic exchange	• Recombinant protein with His_6-tag • 0.1mM IPTG, 24 h, 16˚C, 210 rpm • Affinity chromatography
Average Yield (per L LB)	-	-	6.0 mg (Fraction B)

D. Cleavage of the His_6-tag in CopRHis_6

Although the His_6-tag itself is small, in the CopRHis_6 construct not only a His_6-tag was added to the N-terminus of CopR, but also a S-tag and other residues that in total added 6 kDa to CopR, due to the pET30-EK/LIC vector chosen for expression. This large stretch of amino acids may interfere with the CopR structure and function, and thus this fusion tag was removed before performing the functional assays, using a protease cleavage site introduced in the forward primer, which is recognized by the HRV3C protease (Appendix C.13.1 – Figure 6.18). The HRV3C protease (produced as a GST-fusion construct) was also expressed and purified *in-house*, and a typical purification of this protein is presented in Appendix A.3.

D1. Tag cleavage and Purification of CopR

It is known that enzymatic cleavage is dependent of the fusion protein itself (folding which will change the accessibility to the cleavage site) and on the cleavage conditions used [121]. In a first approach the enzymatic reaction conditions were assayed to verify the influence of the temperature and the time required for this process when using the HRV3C protease. In Figure 3.30 the enzymatic tag digestion at RT and 4ºC at different incubation times is shown. The results obtained evidence that at RT there is unfavourable degradation of the CopR protein, as evidenced by multiple bands with molecular masses inferior to 31 kDa, in lane 13 of Figure 3.30.

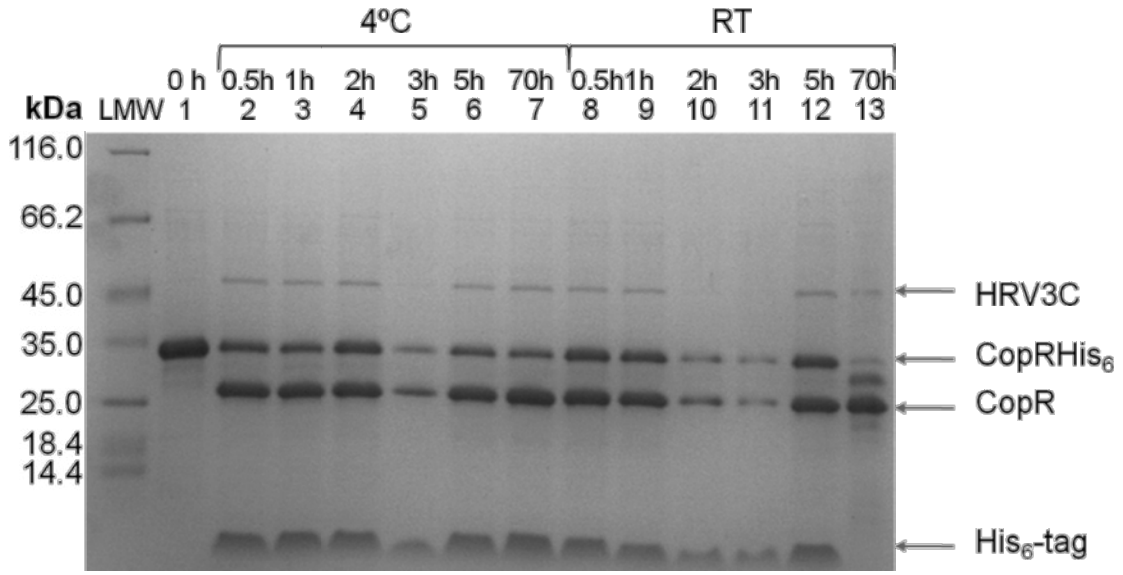


Figure 3.30 Assay of the reaction conditions for enzymatic His_6-tag cleavage of the $CopRHis_6$ protein(31 kDa), with 0.4 mg/mL HRV3C protease (54 kDa) in 20 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Sample prior to protease addition; **Lanes 2 to 7**: $CopRHis_6$ enzymatic cleavage assay at 4°C after 0.5, 1, 2, 3, 5 and 70 hours; **Lanes 8 to 13**: $CopRHis_6$ enzymatic cleavage assay at 25°C after 0.5, 1, 2, 3, 5 and 70 hours. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

On a posterior $CopRHis_6$ sample, the conditions for enzymatic cleavage, 4°C, 0.4 mg/mL HRV3C protease in 20 mM Tris-HCl (pH 7.6), 150 mM NaCl were used and cleavage was allowed to occur until a clear majority of $CopRHis_6$ was digested, Figure 3.31.

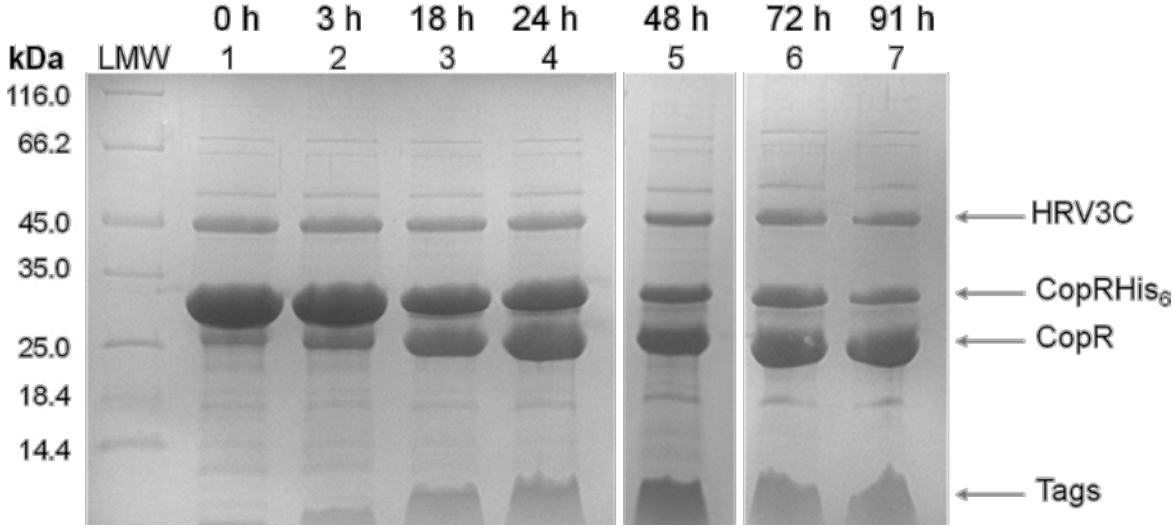


Figure 3.31 Enzymatic His_6-tag cleavage of the $CopRHis_6$ (31 kDa) protein, with 0.4 mg/mL HRV3C protease (54 kDa) in 20 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: **LMW:** Low molecular weight marker; **Lane 1**: Sample upon protease addition; **Lanes 2 to 7**: $CopRHis_6$ protease digestion at 4°C after 3, 18, 24, 48, 72 and 91 hours. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In order to obtain the digested, pure CopR protein sample, another purification step is necessary. In this case, given the affinity of the His_6-tag and the non-cleaved protein for the Ni^{2+}

affinity column, and the HRV3C protease for the GSTrap column, these two were used in tandem for purification, being the majority of the CopR protein eluted with the washing buffer (Figure 3.32).

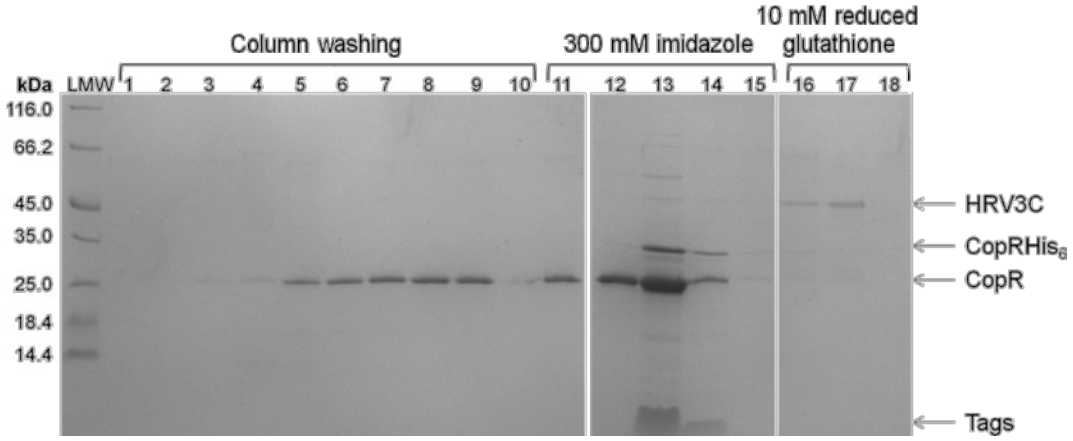


Figure 3.32 Purification of the CopR protein (25 kDa) by Ni^{2+}-sepharose and GST affinity chromatography. Legend: **LMW**: Low molecular weight marker; **Lane1 to 10**: Samples of fractions after elution with 2.5, 5, 7.5, 10, 12.5, 15, 17.5, 20, 22.5 and 25 mL column washing with 50 mM Tris-HCl (pH 7.6), 500 mM NaCl; **Lanes 11 to 15**: Samples of fractions after elution with 4, 8, 12, 16 and 20 mL of 50 mM Tris-HCl (pH 7.6) and 300 mM Imidazole; **Lanes 16 to 18**: Samples of fractions after elution with 4, 12 and 20 mL of 50 mM Tris-HCl (pH 7.6) and 10 mM reduced glutathione. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

The fractions which were pooled to form the completely purified CopR sample were the ones corresponding to the samples loaded into lanes 3 to 12. This fraction was then concentrated (Figure 3.33, Lane C) and used for the posterior studies.

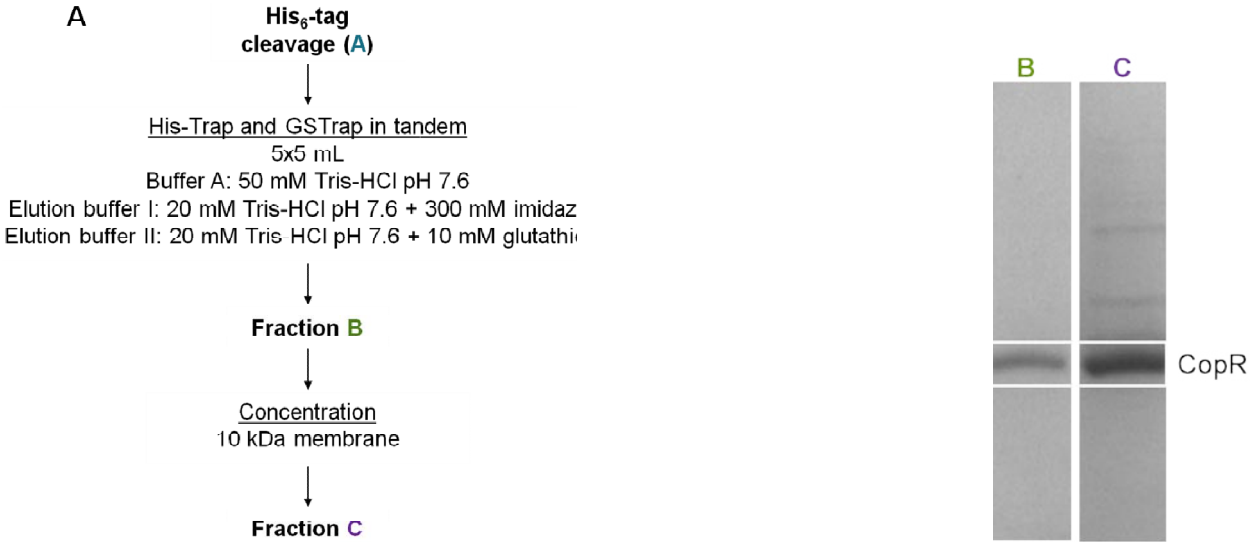


Figure 3.33 A) Flow-chart of the purification of the digested CopR protein (25 kDa). **B)** SDS-PAGE of the fractions obtained using the strategy in the flow-chart. **Lanes A, B and C**: Samples of the fractions identified in A. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

E. Determination of CopR Concentration

For determination of CopR yield, the concentrations of the protein prior to, and upon His_6-tag cleavage and posterior purification, concentration and buffer exchange, are necessary. These values were determined using the BCA method described in Chapter 2 Materials and Methods - Section 2.2.6.5.

In LB medium, the range of purified protein from the soluble cellular extract was between 3.69 and 6.44 mg per liter of LB, which given the observed purity of the sample (Figure 3.33B, Lane B) corresponds in it's entirety to CopRHis_6. After tag cleavage this yield was reduced 96.6 % (yields between 0.14 and 0.23 mg per liter of LB), as expected, since there were three purification steps, and also the majority of CopR without fusion tag was eluted with imidazole with its tag (Figure 3.32, lane 13). This reduction was observed in all the three independent purifications and digestions performed throughout the work.

The affinity of CopR to the Ni^{2+} agarose matrix can be attributed to its aminoacid content (Appendix C.3, Table 6.5), or just its association with its cleaved tag. However, this fact was not tested and greatly reduces the yield of pure CopR obtained at the end of the procedure, and thus requires optimization, such as finding other strategies to separate these proteins after cleavage, though neither ion-exchange or size-exclusion chromatography could be used, due to the similar properties of the proteins to separates (pI and MM).

3.2.3 Biochemical Characterization of CopR

The biochemical characterization of the CopR protein was performed on the B fractions obtained from the purification strategy, *i.e.*, the fractions with a higher purity ratio determined by SDS-PAGE.

This characterization consisted in the determination of the apparent molecular mass in 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl and verification of the formation of complexes under various conditions. The CopR protein was also subjected to *in vitro* phosphorylation and electrophoretic mobility shift assays. Circular dichroism was used to evaluate the overall state of folding of the protein.

3.2.3.1 Stability Studies

The amino acid content of CopR according to its primary sequence (Appendix C.3, Table 6.5), shows that there are two cysteine residues, Cys73 and Cys161. As an initial approach to stabilize CopR, DTT was added to a final concentration of 1 mM. This was tested on CopR1 (CopR expressed without a fusion tag) but the rate of degradation/precipitation was not visibly affected (data not shown). Precipitation was in fact so significant, even in the presence of DTT,

that during purification using a Superdex 75 column the quantity of protein retrieved was considered almost negligible.

The second approach was more methodical, in such that equal fractions of the same sample of $CopRHis_6$ (prior to tag cleavage) were incubated in different conditions of stability agent (glycerol) and osmolarity (NaCl), as explained in Chapter 2 Materials and Methods – Section 2.3.3.1 and compared by SDS-PAGE (Figure 3.34). From previous manipulation of the CopR protein it was known that the protein is remarkably unstable, and precipitates with time. In order to preserve the quantity of soluble protein for future studies, this assay was carried out using the recovered precipitated CopR. Therefore, there is no real way to quantify the amount of $CopRHis_6$ in each of these samples without expending the sample itself. The retrieved pellet was resuspended and homogenized as much as possible in 1.4 mL of 50 mM Tris-HCl (pH 7.6) and 250 mM NaCl. From this resuspension, a 200 μL sample was used for each of the assayed conditions. Each of the samples was incubated on ice for 8 h and subsequently centrifuged for 30 min at 14,000 xg, and only a sample of the supernatant was loaded onto the SDS-PAGE (Figure 3.34).

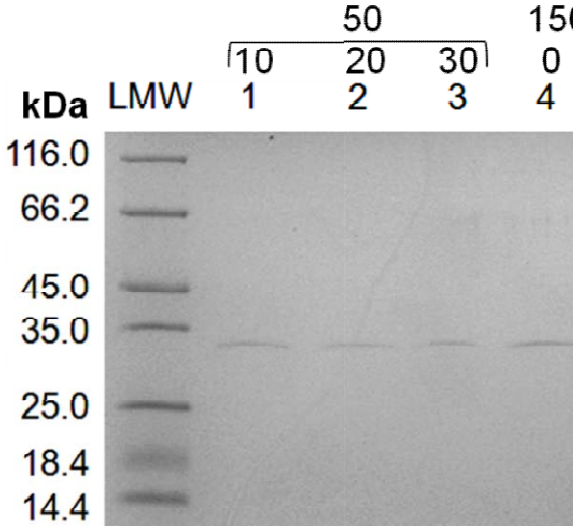


Figure 3.34 Stability assay of the $CopRHis_6$ protein (31 kDa) in various buffers. Legend: **LMW**: Low molecular weight marker; **Lane 4**: Control, $CopRHis_6$ in 50 mM Tris-HCl (pH 7.6), 50 mM NaCl; **Lanes 1 to 3**: $CopRHis_6$ in 50 mM Tris-HCl (pH 7.6), 50 mM NaCl; **Lanes 5 to 7**: $CopRHis_6$ in 50 mM Tris-HCl (pH 7.6), 250 mM NaCl; **Lanes 1 and 5, 2 and 6, and 3 and 7**: Glycerol was added to 10%, 20% or 30% respectively. 200 μL of $CopRHis_6$ was used in each lane and diluted 1:3. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

The increase in intensity of the bands in lanes 5 to 7 from Figure 3.34 is consistent with an increase in the quantity of soluble $CopRHis_6$. So, from this point forward, all $CopRHis_6$ and CopR samples were maintained in 50 mM Tris-HCl (pH 7.6), 250 mM NaCl and 10% glycerol.

As an additional measure freeze/thaw cycles were also reduced by freezing half the sample and maintaining the remaining sample at 4°C, for use.

3.2.3.2 *In vitro* Phosphorylation Assay

It is generally accepted that phosphorylation of the conserved Asp residue induces conformational changes throughout the structure of the RR, which in turn leads to formation of a homodimer [122]. This dimer forms an active complex which will interact with DNA, where each monomer recognizes a half-site of 2-fold symmetric DNA sequence [75].

For *in vitro* phosphorylation in the presence of Mg^{2+} the pros and cons of two small phosphodonor molecules were taken into account before the final choice of using acetyl phosphate was made. The choice of using acetyl phosphate lies in the experimental evidence that for CheY (a stand alone receiver RR), in contrast to phosphoramidates, autophosphorylation with acetyl phosphates is largely pH independent [123]. Due to the previously referred instability of CopR at the time it seemed a logical choice not to alter the conditions in which it was, more, soluble. Moreover, the biological relevance of acetyl phosphate and the evidence that RRs autophosphorylate *in vivo* using this molecule [124] only served to cement this choice.

It was shown by Mayover *et al.*, 1999 [46] that phosphorylation of the RR CheY is inhibited by increasing ionic strength. Given that both the receiver domain active site and small molecule phosphodonors are highly charged, the effect of ionic strength that perhaps inhibits interactions is not surprising. However, this effect has only been, to my knowledge, demonstrated for CheY, but throughout this process it is worthwhile to keep in mind that increased ionic strength, as that needed for CopR solubility, might be a hindrance for interdomain interactions.

Various *in vitro* phosphorylation assays were consulted [72,93,125–129] whilst planning this assay, as phosphorylation is sensitive to the buffer components (*i.e.* ionic strength, pH, Mg^{2+} concentration, acetyl phosphate concentration, etc.), which need to be optimized, for each protein.

Though phosphorylation may not be pH dependent in some cases, acetyl phosphate is affected by the pH of the buffer. As Figure 3.35 shows, phosphorylation was attempted at pH 7, 8, 8.5 and 9. However, at pH's above 7, the acetyl phosphate included in the buffer precipitated in a matter of hours, becoming irreversibly insoluble.

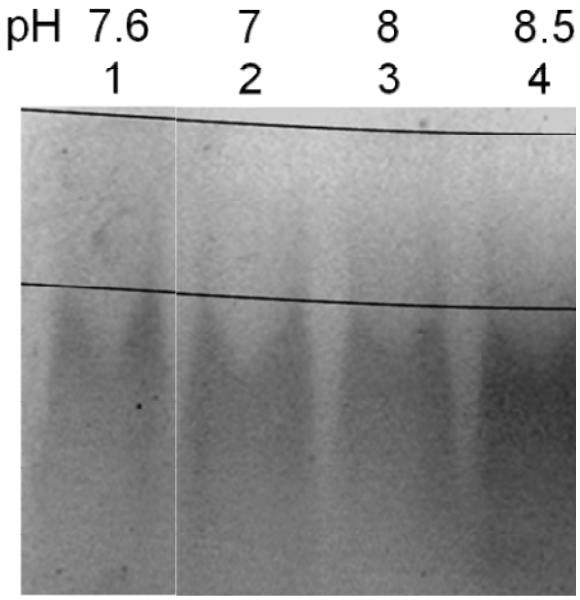


Figure 3.35 Phosphorylation assay of the CopR protein at different pHs. Legend: **Lane 1**: Control CopR protein in 50 mM Tris-HCl (pH 7.6), 250 mM NaCl and 10% glycerol; **Lanes 2 to 5**: Attempted phosphorylation of CopR with the phosphorylation buffer described in Chapter 2 Materials and Methods – Section 2.3.3.2 at pH 7, 8, 8.5 and 9, respectively. The black lines represent the slight curve of the gel and serve for comparison. PAGE prepared in Tris-Glycine buffer (10%, 150 V, 1 h) and stained with Coomassie-blue.

In Figure 3.35, the buffers were used within an hour of preparation, and it seems that phosphorylation in these conditions does not occur as, independent of the pH of the phosphorylation buffer, the protein migrates with the same apparent electrophoretic mobility. On a native PAGE of this type, phosphorylated samples are expected to migrate slower than unphosphorylated protein samples, given their superior molecular mass or the alteration of over-all protein charge.

Various attempts were made to phosphorylate CopR (data not shown), the results presented in Figure 3.35 were merely one of the conditions tested. Gel conditions were also manipulated, nonetheless, there was never an appreciable, consistent, reproducible variation in migration for CopR upon *in vitro* phosphorylation, when compared with the native form.

It is also important to mention that it has been proposed by several researchers (Mukhopadhyay A., personal communication to Pauleta SR) that response regulators may be phosphorylated in the cytoplasm of *E.coli*, an hypothesis which shall be tested at a later date and would explain the result obtained here. Another explanation could be the low stability of the CopR dimer that during migration in the gel could be disrupted.

3.2.3.3 Molecular Mass of CopR

A. Reducing and Non-reducing SDS-PAGE and PAGE

As CopR contains two cysteines, even before tag cleavage, it is possible that intra or intermolecular disulphide bonds could be formed during expression/purification. With the use of two gels, in reducing and non-reducing conditions, Figure 3.36 one can quickly assess the presence and nature of these bonds, as well as possible states of oligomerization.

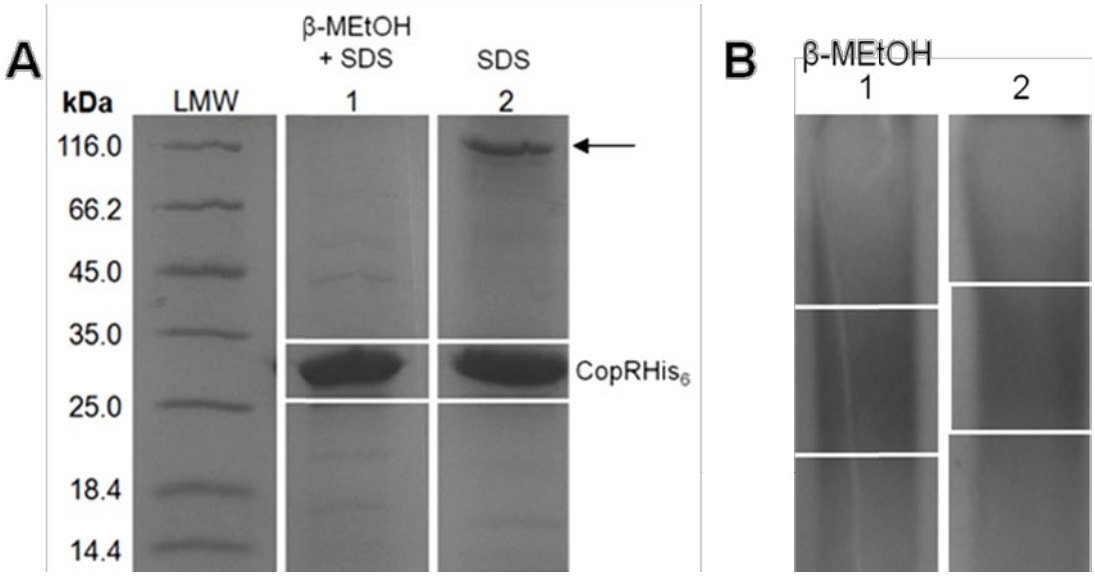


Figure 3.36 A) Analysis of intermolecular disulfide bond presence and oligomeric association *via* SDS PAGE in reducing and non-reducing conditions, of the $CopRHis_6$ protein (31 kDa). Legend: **LMW**: Low molecular weight marker; **Lane 1**: $CopRHis_6$ sample in the presence of SDS and β-MEtOH; **Lane 2**: $CopRHis_6$ sample in the presence of SDS. White boxed region identifies the $CopRHis_6$ protein monomer of 31 kDa. SDS-PAGE prepared in Tris-Glycine buffer (10%, 150 V, 1 h) and stained with Coomassie-blue. **B)** Analysis of intramolecular disulfide bond presence *via* PAGE in reducing and non-reducing conditions, of the $CopRHis_6$ protein (31 kDa). Legend: **Lane 1**: $CopRHis_6$ sample in the presence of β-MEtOH; **Lane 2**: $CopRHis_6$ sample in native, non-reducing, non-denaturing conditions. White boxes highlight the region of the gel in each lane with greater intensity. PAGE prepared in Tris-Glycine buffer (10%, 150 V, 1 h) and stained with Coomassie-blue.

Analyzing Figure 3.36 A, the band of $CopRHis_6$ in a reduced and denatured form, identified in the boxed region, appears between 25 and 35 kDa, consistent with its expected molecular mass of 31 kDa (CopR before His_6-tag removal). In the sample without β-MEtOH, a reducing agent that reacts with disulfide bonds, a band at approximately 116 kDa is now visible (identified with an arrow), which could pertain to a $CopRHis_6$ trimer. Nonetheless, even if this is the case, the $CopRHis_6$ protein is present and purified mostly as a monomeric protein.

The native gel, PAGE of Figure 3.36 B, was performed in the same conditions as the gel in panel A, to access the presence of intramolecular disulfide bonds. Independent of the low resolution of the gel and the diffused CopR band in PAGE, a common problem when resolving the CopR protein in a PAGE, it seems that in the presence of the reducing agent $CopRHis_6$ migrates with a higher electrophoretic mobility, when compared with the same protein in non-reducing conditions in the same gel (white boxes).

These results lead to the conclusion that the CopRHis$_6$ protein is obtained as a monomer, in which the two cysteine residues present in the sequence do not form intermolecular bonds. In fact, upon analysis of the CopR model structure, it seems that these two residues are somewhat buried within the protein, Figure 3.37, and thus not solvent exposed to enable the formation of disulfide bonds, intermolecular or intramolecular, thus the high molecular band observed in non-reducing PAGE can be attributed to intermolecular disulfide bond formation in partially unfolded protein.

Figure 3.37 Model structure of CopR depicted as a surface for cysteine localization purposes. Legend: Cysteines 73 and 161 are shown in the inserts. Atoms are coloured by element, carbon in white, oxygen in red, nitrogen in blue and sulphur in yellow. The methionines present in this protein are shown in orange to avoid confusion with the cysteines, as they too have a sulphur atom but which is not available for disulfide bond formation. Figure prepared using Chimera software [108].

B. Apparent Molecular Mass of CopR

Further assessment of oligomeric states, and determination of the apparent molecular mass of this protein was performed using a pre-packed Superdex 75 10/300 GL, as described in Chapter 2 Materials and Methods – Section 2.3.1. The obtained chromatograms are shown in Figures 3.38 and 3.39, and the calibration curve that was used for apparent molecular mass determination is presented as the insert of Figure 3.38.

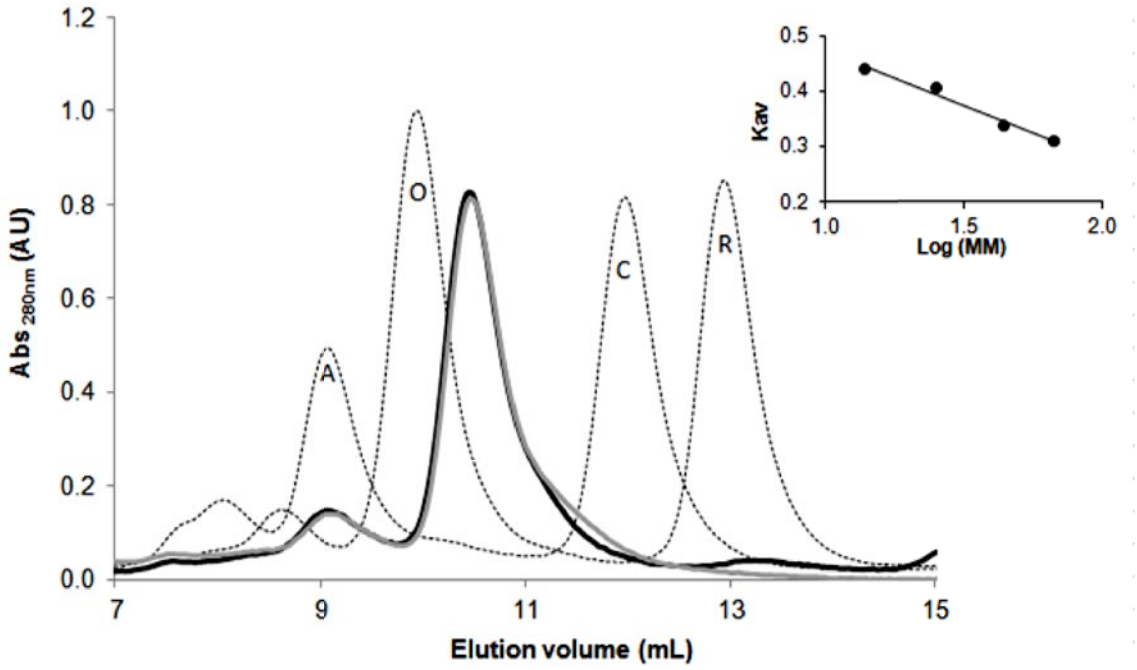


Figure 3.38 Elution profiles of calibration proteins on a Superdex 75 10/300 GL. Proteins were eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: **A**: Albumin, 66.8 kDa; **O**: Ovalbumin, 43.0 kDa; **C**: Chymotrypsinogen A, 25.0 kDa; **R**: Ribonuclease A, 13.7 kDa. Grey curve was obtained for CopRHis$_6$ at 280 nm. Black curve was obtained for the same protein in phosphorylation inducing conditions, at 280 nm. Insert: Calibration curve for the molecular weight determination using LMW Gel Filtration Calibration Kit in Superdex 75 10/300 GL. Trend line equation obtained, Kav = -0.20 log (MW) + 0.67, with an R of 0.99. The profiles correspond to mix 1: A+C and mix 2: O + R.

Phosphorylation of 1.4 mg/mL CopRHis$_6$ was performed by incubating CopR with phosphorylation buffer to the final concentrations described previously (Chapter 2 Materials and Methods - Section 2.3.3.2). The reaction was performed for 3 hours at RT, prior to column injection. Despite the fact that the prior assay for phosphorylation (Chapter 3 Results and Discussion - Section 3.2.3.2) was not conclusive, this could be due to gel conditions which perturb the fragile complex formed (as mentioned), whilst this method can be considered minimally invasive.

Figure 3.38 shows that the elution profiles of both forms of CopRHis$_6$ are almost identical, presenting with almost exactly identical elution volumes (Elution volumes presented in Appendix C.13.4, Table 6.11). The estimated molecular mass for CopRHis$_6$ and CopRHis$_6$-P was determined to be 40 kDa *vs.* the theoretically determined, 31 kDa in the form of monomer. This 7.5 kDa difference could simply be due to the shape that CopRHis$_6$ adopts within the column (non-globular) or could be due to a more complex effect of monomer-dimer equilibrium. Such an ambiguity could be clarified with the determination of the apparent molecular mass as a function of the ionic strength, which was not performed for this protein given its inherent instability, and therefore tendency to precipitate within the columns. A similar study was performed on CopR after tag cleavage, which is presented in Figure 3.39.

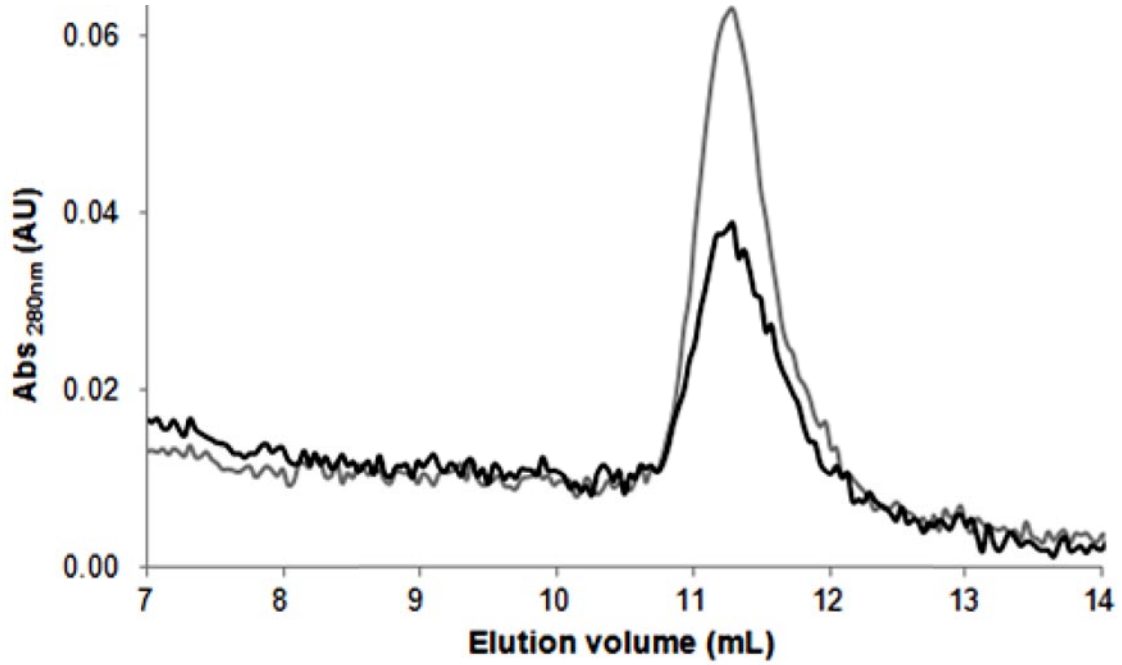


Figure 3.39 Elution profile of heterologously expressed and purified CopR protein in a Superdex 75 10/300 GL. Proteins were eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: Grey curve was obtained for CopR at 280 nm. Black curve was obtained for the same protein in phosphorylation inducing conditions, at 280 nm. Profiles correspond to independent injections.

The relative intensities between CopR and CopR-P is approximately double (0.039 and 0.063 AU for CopR and CopR-P, respectively) and both proteins elute with an apparent molecular mass of 28 kDa in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl pH 7.6, Figure 3.39. However, this observation does not seem relevant, as the intensity might only reflect the amount of protein loaded onto the column, and does not imply that at least a portion of the CopR protein exists in the form of a dimer, which would appear as a peak eluted at a smaller elution volume (not observed).

3.2.3.4 Effect of pH on CopR-P

A wide variety of conditions are expected to affect the RR-DNA interaction, and for the OmpR/PhoB family the formation of a dimer seems to be essential for recognition of the transcription factor binding sequence. Assays for the phosphorylation of the CopR protein included the influence of pH during phosphorylation but due to the variations of pH within PAGE these effects could have been obscured. As such, these assays were also performed in the Superdex 75 column, Figure 3.40.

The pH indicated in the graph of Figure 3.40 correspond to the pH of the kinase buffer used to phosphorylate the CopR protein. Phosphorylation was again performed as previously described before injection into the column which was equilibrated with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl pH 7.6.

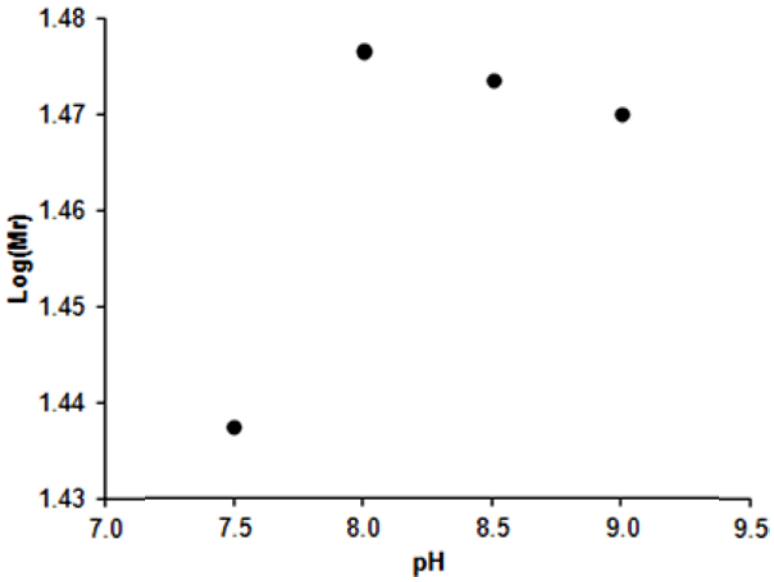


Figure 3.40 Representation of the log(Mr) of the CopR protein in phosphorylation inducing conditions with kinase buffers of different pH's (7.5, 8.0, 8.5 and 9.0).

The individual elution volumes are presented in Appendix C.13.4, Table 6.17, and it is possible to see that upon phosphorylation of CopR in a kinase buffer at pH 8.0, the molecular mass (determined as 30 kDa) is superior to that of the protein phosphorylated at pH 7.5 (determined as 27 kDa), for example, though in gel filtration this difference is usually not considered significant due to the error associated with this determination. Having mentioned this, the difference in the log (Mr) observed for the different assays, which were obtained in consecutive runs with the matrix equilibrated in the same buffer, can be considered different.

This small increase in molecular mass can be attributed to the phosphorylation (or partial phosphorylation) of the CopR protein, which alters the tertiary structure and as such reflects on the elution volume of the protein. This presumption leads to incubation of the CopR protein in a kinase buffer of pH 8.0 for RR-DNA interaction assays.

3.2.3.5 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy was used to demonstrate that the obtained CopR protein presented secondary structure. CD spectra were collected from 190 to 240 nm and analyzed using a neural network analysis, K_2D_2 [114] (Figure 3.41).

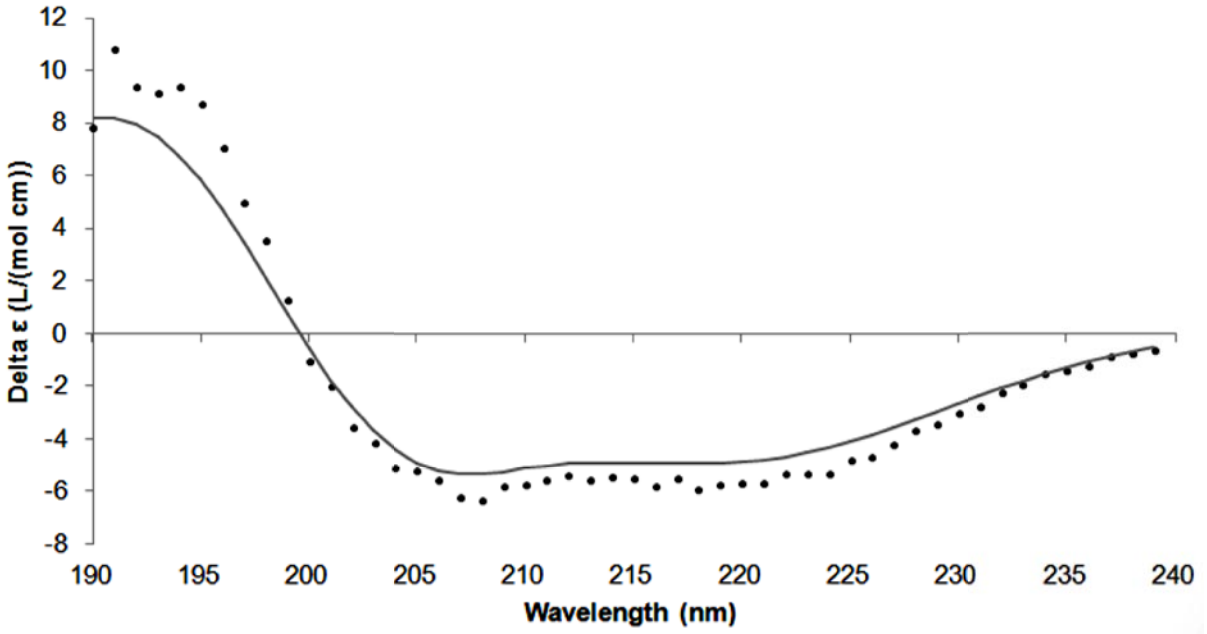


Figure 3.41 Theoretical (-----) and experimental (· · · · ·) CD spectra of the CopR protein (3.69 mg/L) in 5 mM phosphate buffer (pH 7.4), 250 mM NaCl and 10% glycerol. The theoretical CD spectrum was obtained by inserting the experimental data into the K_2D_2 analysis tool [114].

The experimental data seems to be of reasonable quality, and could be fitted to the theoretical curve of a protein composed of 40% α-helices, 12% β-sheets, and 49% random structure, values that are closer to the ones obtained for the predicted secondary structure *vs.* those of the secondary structure obtained for the model structure (Table 3.4).

Table 3.4 Secondary structure composition of the CopR protein determined by prediction servers, homology and deconvolution (using K_2D_2 [114]) of the experimental data obtained by CD.

	CD Deconvolution (%)	Prediction (%)	Homology (%)
α-helix	40	38	31
β-sheet	12	21	25
Random coil	49	41	44

Due to the fact that the concentration of CopR used in this study is lower (12.5 μg/mL) than the recommended (between 0.1 to 0.2 mg/mL), the results obtained have a qualitative rather then a quantitative character. Nevertheless, these results establish that the purified CopR protein retains secondary structure.

3.2.3.6 CopR - Promoter Binding Studies

A. Bioinformatic Analysis of the Promoter Region

In prokaryotes, some co-regulated genes are organized into operons on neighbouring *loci* to be transcribed together *via* a single promoter region. There are four conserved features of bacterial promoters [130]: The transcription start site (TSS); the -10 sequence; the -35 sequence; and the separation between these -10 and -35 sequences.

The σ^{70} family of sigma factors (Table 3.5) promote binding of the RNAP I to generally expressed gene promoters, so the promoter region of the *copSRXAB* operon in question should show -10 and -35 elements typical of eubacterial σ^{70} promoters [131].

Table 3.5 Specific σ factors bound to the RNAP I increase the affinity for the -10 and -35 hexamers. *E.coli* σ factors recognize promoters with different consensus sequences. N – A/T/C/G; Pu - Purine

Sigma factors recognize promoters by their consensus sequences				
Use	**Factor**	**-35 Sequence**	**Separation**	**-10 Sequence**
σ^{70}	General	TTGACA	16-18 bp	TATAAT
σ^{32}	Heat shock	CCCTTGAA	13-15 bp	CCCGATNT
σ^{54}	Nitrogen	CTGGNA (-24)	6 bp	TTGCA (-12)
σ^{28}	Flagellar	CTAAA	15 bp	GCCGATAA
σ^{H}	Unknown	AGGANPuPu	11-12 bp	GCTGAATCA

Taking these sequences into account, and the observation that the maqu_0125, maqu_0126 and maqu_0127 genes seem to form an operon, the promoter region upstream of *copX* (maqu_0125) was analysed.

A)

AGATCGAGAACAGCGGCACGATAATCCTCCTGCTCCCCCAGAATCTGGGCCTCACGGGCACTGGACGTGTGGTCGATGCTAAACCC
CGCTTTTTCCAGCTGTCTGACCAGCCCCTCGGCCAGCAAACGGTCGTCTTCAACGAGCAATAAACGCATGTATCCGTCCTTAACGG
TGCTTGTTTCAGCCGCCAGTGTCGTCGATCAACCTGAATTGTGCCTGAACGCAAGAAAGTGCCCGTTTTTCAGAAAGAATTCAGGT
TGGTGAGAGAAAGTGTCGACTGATTCTGGACATTGCTCAAGATTCAGTACGAATCAGCCGTGTTCTCAAAACCATTATTCCACAGG
AGACAAACGGATGAATGCATCAAACCTGATTGTGGCCGGCCTGGCTTTTTCGATGTCGGTATCGGCGTTTGCGGCTGGCACCCATG
GCGGTGGTCACGGCCATGGTGCTTCCATTGGCGAACCGGGAAAAGCCTCAGAAGCCAGCCGGACCATA

B)

GTAGCCGATGCCCTGCTTTACAAAATCACCGGTATCTATCTCTACGCGATCAACCGTCAGGTCGTACTCGCCAGCCAACGCCAGTG
CGGGCATCAGCAGGCACAGCACTACTGCACCCAAGCGAATGTTCATGTGTTTTGCTCCAAACGCACTTTATCGGGGCTGACAGGAG
GGTGTAGCGTCAGCCCATAACGTCATGGTTAAAGCCATCTACCGGCTTATTCGAAATTCACGTCACCGTACATGCCAGACTGGTAG
TGCCCTGGTACGTTACATGCGAATTCGATATTGCCCTGATTGGCAAACGTCCAGACCACTTCCCGGCTCTGGCCCGGCTCCAGCAG
GACGCTGTTGGGATCGTCATGTTTCATCGAGTGGCCGTTGCCCATGTCCATGTTCATCATGTCATGATTCAACTTGTTACCCTGGA
TCACGCCATGCTCGACCATCATTCTCATTTCTTTCTGGTGGGCCTCATGCATACCCGGAGTCCCGATG

Figure 3.42 Analysis of the DNA sequence located between maqu_0124 and maqu_126 from *M. aquaeolei* VT8. Legend: ATG and CAT in red denote the start codons for A) *copR*, *copX* or B) *copX, copA* genes. The *copX* gene is underlined and *copR* is in grey. The putative promoter regions are in orange. The highlighted yellow regions correspond to regions similar to the cop box of OmpR (Appendix C.11.3) **A)** 5' – 3' DNA sequence corresponding to the *pro* DNA fragment and approximately 150 bp in the 5' and 3' direction. **B)** 5'-3' DNA sequence corresponding to the region between the putative maqu_0125 and maqu_0126 ORFs.

This region was also analysed for the presence of DNA binding sequences identified for other copper response regulators (see Chapter 1 Introduction – Section 1.3). However, these sequences seem to be absent in this region, as thus it could be that the CopR binding sequence is different.

There appear to be two separate promoter regions within the *copSRXAB* operons. With the aid of BPROM (Softberry, Inc., Mount Kisco, NY) these promoter regions were identified and are localized between the *copR* and *copX* genes (Figure 3.42 A) and within the *copX* gene before *copA* (Figure 3.42 B) and belong to the σ^{70} family.

In the first case, these regions correspond to -10 and -31 with respect to the *copX* transcription start site (TSS) which appears 46 bp upstream of the *copX* start codon, Figure 3.42 A, and as such should function as the promoter sequence for *copX, since copAB* as identified here seems to present its own promoter region. The region in question, highlighted in grey in Figure 3.42 A, was amplified by PCR, as described in Chapter 2 Materials and Methods - Section 2.2.

For the *copRS* operon, the promoter regions correspond to -12 and -35 with respect to the TSS, which seems to be absent or was not identified using these bioinformatic tools. However, within the *copX* gene another promoter region was identified in this study, which might correspond to the promoter region of *copAB*. This second promoter region was not previously identified and thus is not present in the DNA sequence isolated and denominated *pro*.

B. DNA Concentration Determination

The DNA fragment was obtained as described in Chapter 2 Materials and Methods - Section 2.2, and Figure 3.43, Lane 1 shows the amplification of a fragment with approximately 200 bp, identified by the white boxed region. This indeed corresponds to the expected fragment size of 202 bp.

In order for this dsDNA fragment to be used in the electrophoretic mobility shift assays, it was purified (Figure 3.43, Lane 2) and quantified as described in Chapter 2 Materials and Methods - Section 3.2.

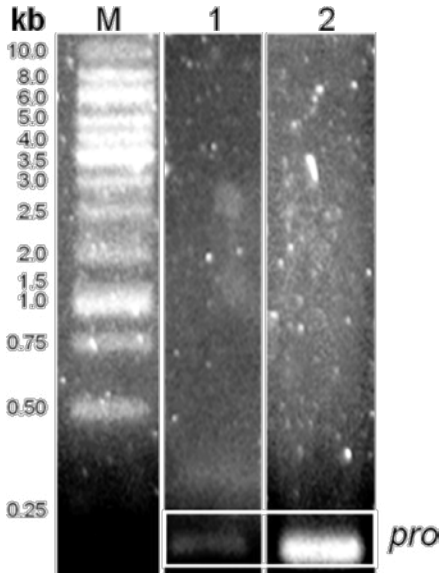


Figure 3.43 PCR amplification and purification of the *pro* DNA fragment (202 bp), using the primers and PCR program described in Chapter 2 Materials and Methods - Section 2.2. Legend: **M**: DNA ladder of 1 kb; **Lane 1**: Amplified PCR product of approximately 200 bp. **Lane 2**: Purified dsDNA fragment resultant from PCR amplification. 1.8% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

As needed, PCRs were performed to obtain this fragment for EMSA, however the concentrations in the assays remained constant.

C. Evaluation of CopR-DNA Interaction by Gel Filtration

The binding of CopR to *pro*DNA was studied using two approaches, the determination of the elution profile of CopR in the presence of *pro*DNA and the other was Electrophoretic mobility shift assays (EMSA). Therefore, CopR was injected in a 5:1 proportion into the column with the *pro* fragment, the DNA region of 202 bp between the start coding of *copR* and *copX*, which is proposed to contain the binding sequence of CopR (Figure 3.44). This experiment tested the hypothesis that as in the case of OmpR, dimerization is dependent on the presence of the DNA fragment [132].

Also, it is essential to refer to the fact that in this assay the ligation buffer did not contain BSA. In appendix however, is also the chromatogram for the DNA alone, Appendix C.13.5, Figure 6.21.

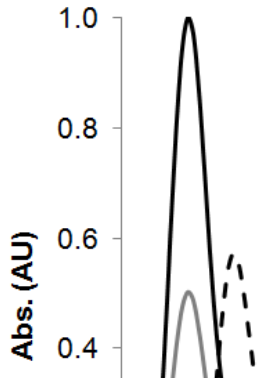


Figure 3.44 Elution profiles of CopR-P in the presence and absence of the DNA fragment *pro* (5:1 ratio) in a Superdex 75 10/300 GL. Biomolecules were eluted with 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl. Legend: The dotted grey and black lines represent the control elution profile of CopR-P alone (elution profile at 280 nm) and *pro*DNA (eluition profile at 260 nm), respectively. For clarity a constant (0.05 A) was added to these profiles. The continuous black and grey profiles correspond to the elution profile of CopR-P in the presence of DNA at 260 nm and 280 nm, respectively.

The elution profiles of the CopR-P protein (dotted grey line) and the *pro*DNA fragment (dotted black line) allow for comparison of the elution profiles when both components were injected simultaneously. The determined molecular mass of the CopR-P protein, when injected independently, is 30 kDa, which is in agreement with the previous results.

Given the exclusion limits of the Superdex 75 column (3 – 70 kDa), the *pro*DNA fragment was eluted as a molecule that exceeds these limits with an elution volume of 7.8 mL (comparing this value with albumin of 67 kDa with an elution volume of 9.1 mL). However, analysis of the results shows that the elution volume of the DNA is altered in the presence of CopR-P (black continuous line, when compared with the dotted line for the free *pro*DNA). Due to the difficulty in determining the molecular mass of this complex, and of the isolated DNA, this peak shift may occur due to an interaction of the DNA fragment with the CopR-P either as monomer or dimer. These results indicate the formation of a complex, though weak. However, to completely validate the hypothesis that this increase in molecular mass of the DNA band is due to the formation of the CopR-*pro*DNA complex, assays with different CopR-P:*pro*DNA ratios should be performed, as well as assays in which the concentrations/components are similar to the ones used in the binding assay (see below).

Another injection was performed in which the previous conditions were maintained and BSA was added to the ligation buffer to a final reaction concentration of 75 µg/mL (Figure 3.45).

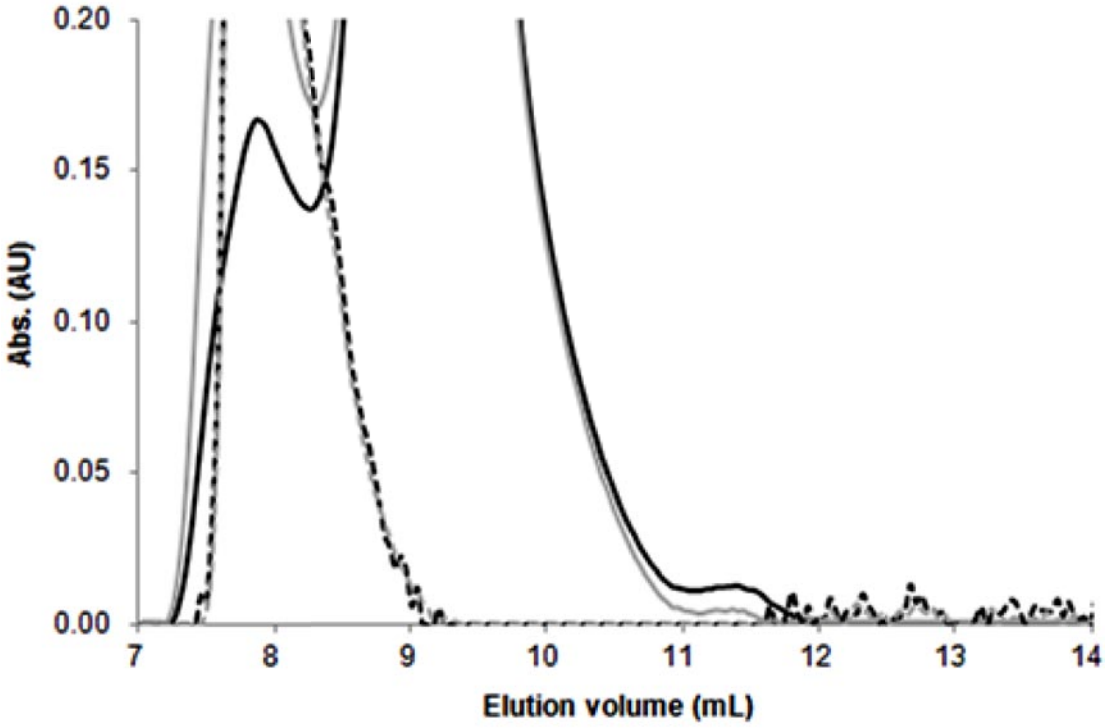


Figure 3.45 Elution profiles of the DNA fragment pro (5:1 ratio) in the presence and absence of CopR-P in a Superdex 75 10/300 GL in the presence of BSA (concentration of 75 μg/mL) in the ligation buffer. Biomolecules were eluted with 50 mM Tris-HCl (pH 7.6) and 150 mM NaCl. Legend: The dotted grey and black lines represent the control elution profile of the pro DNA fragment alone elution profile in black and grey measured at 280 and 260 nm, respectively. The continuous black and grey profiles correspond to the elution profile of the pro DNA fragment in the presence of CopR at 280 nm and 260 nm, respectively.

The results obtained for this assay are not as clear as the ones mentioned before, as the concentration of BSA (eluted with an elution volume of 9.1 mL) in the reaction is several times higher than those of the DNA and CopR-P components, masking their elution profiles and thus the interpertation of a possible interaction. Thus, the determination of the elution volume is morecomplex. However, the comparison of the elution volumes of the DNA in the presence or absence of CopR-P in these conditions (7.5 and 7.9 mL, respectively), may indicate that the DNA-CopR-P complex might not be formed due to the presence of BSA in the ligation buffer.

D. Evaluation of CopR-DNA Interaction by EMSA

The relatively strong binding necessary for initiation of transcription is a result of two different types of interactions, sequence–specific with specific DNA bases, interaction of the protein recognition helix (α3) with the major groove, and non–sequence specific interactions with the DNA backbone, in this case namely the wing interactions with the minor grooves [64]. The low affinity ($\sim 10^{-3}$–10^{-5} M) non-sequence specific binding to DNA allows RRs (and generally all transcription factors) to slide along the DNA and find their target sites, whereas the high affinity ($\sim 10^{-8}$–10^{-12} M) sequence–specific interactions allow the immobilization of the RR to its target site for sufficient time to allow regulation of transcription [130].

Usually EMSAs are performed with different ratios of protein to DNA, and also in the presence of competitor DNA, but in this initial stage our goal was to identify the conditions in which protein-DNA interactions occurred. Given the fact that in some RR the formation of the dimer is strictly necessary for DNA interaction, and in others dimerization occurs only in the presence of the DNA, in all EMSAs the CopR protein was incubated with a low molecular mass phosphodonor, acetyl phosphate, in order to assure CopR phosphorylation.

The protein-DNA association will lead to an altered electrophoretic mobility of CopR, resulting in a band shift when the sample is loaded onto a native, non-reducing PAGE.

The appearance of smears in EMSA gels is not uncommon due to the interactions between the protein and DNA [133]. In the case of a strong interaction between protein and DNA two distinct bands are observed, corresponding to the protein:DNA complex and to the free DNA (in the case of lower protein:DNA ratios). However, because of the dissociation that inevitably occurs during electrophoresis, and because this free (released) DNA can never run as the free DNA, a faint smear may be seen between the two major bands [133]. In contrast, a weak DNA–protein interaction should produce a fainter band corresponding to the protein:DNA complex and a more intense smear.

Additionally, multiple effects are likely to contribute to the altered stability of the RR-DNA complex in the non-native condition that exists inside the gel matrix during electrophoresis. These include, for example, restriction of diffusion by the gel matrix, migration of the reactants in the applied electric field, and altered concentration of ions. These assays are therefore subjected to a wide variety of factors. Many of these were tested, such as gel concentration, gel buffer and running conditions, kinase buffer components and concentrations. One effect that seemed quite prominent throughout these assays was the localized variation of ion concentration/pH [134].

The results are not shown for all of the assays, as none of the tested conditions produced a reproducible result with a band shift of any significance. The most promising conditions are those presented in Figure 3.46. Also, calculations were performed for protein titration, given the assumption that CopR binds as a dimer, however, given the reduced concentration of CopR after His_6 tag digestion, these assays in PAGE were not feasible.

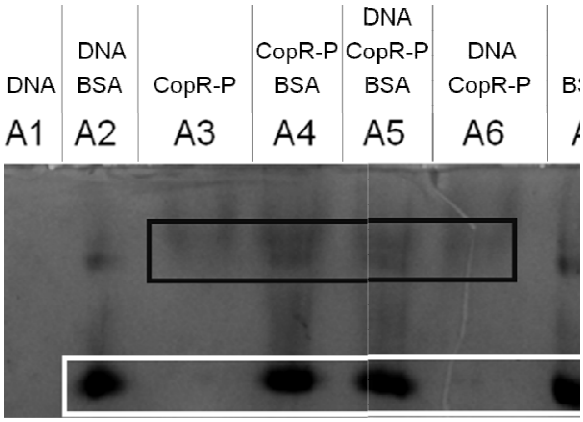


Figure 3.46 EMSA for CopR interaction with the putative promoter fragment, *pro*. Legend: **Lanes 1**: 150 ng/mL control *pro* DNA fragment in distilled water; **Lanes 2**: 150 ng/mL *pro* DNA fragment with BSA in kinase buffer; **Lanes 3**: Control 11.5 ng/mL phosphorylated CopR; **Lanes 4**: 11.5 ng/mL phosphorylated CopR with BSA in kinase buffer; **Lanes 5**: phosphorylated CopR in a 1:1 ratio *pro* DNA fragment with BSA in the kinase buffer; **Lanes 6**: 12.7 ng/mL phosphorylated CopR in a 1:1 ratio with *pro* DNA fragment; **Lanes 7**: Control kinase buffer with BSA. PAGE Tris-Glycine gel prepared in Tris-Borate buffer (10%, 120 V, approximately3 h) **A**: Gel was stained with Silver Staining. **B**: Gel was stained with SybrSafe and visualized under UV light.

Figure 3.46 shows the same PAGE gel stained either for protein (Figure 3.46 A) or DNA (Figure 3.46 B). Various controls were used, and only lanes 5 and 6 correspond to an EMSA assay, in the presence or absence of BSA in the kinase buffer. It is quite clear that in either case there is no change in the DNA fragment electrophoretic mobility (black box in Figure 3.46 B).

Identification of the band which corresponds to the CopR protein is difficulted by the propensity of this protein to diffuse within the gel lane when the gel is run in native conditions. The black box in Figure 3.46 A highlights the presence of a second, more diffused band above the one proposed to be BSA (white box in Figure 3.46 A), which should correspond to the CopR protein. Therefore, the data obtained shows that in the PAGE free DNA and BSA have similar migration, and that under these conditions and detection limit of the SybrSafe it is not possible to detect a band corresponding to CopR-P:*pro*DNA.

These results are somewhat complementary to those obtained by the gel filtration. Independent of the presence of BSA or not, the complex does not seem to form, however, if it is possible an assay with a larger CopR-P:DNA ratio would be constructive.

Chapter 3 – Results and Discussion
Response Regulator CopR domains

3.3 Response Regulator Domains - CopR_N and CopR_C

3.3.1 Bioinformatic Analysis

3.3.1.1 Domain Analysis

Given the challenges encountered whilst working with the full-length CopR protein, another strategy was designed in which each domain was expressed separately. The main reasons for adopting this strategy lie in the modular nature of TCS proteins (Figure 3.47 A) and the reported cases of successful domain manipulation [65,135,136]. There is, however, a possible limitation to this approach, as analysis of the propagation of alterations which occur upon phosphorylation/activation, from the receiver to the effector domain, is compromised.

A)

B)

```
        10        20        30        40        50        60        70        80
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL VECPVLVLTA
RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS LRTEEGAEHA LTGTEFRLLR
CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR RGQGYMFNAQ R
```

Figure 3.47 Domain analysis of CopR from *M. aquaeolei* VT8. **A)** Schematic representation of CopR which is predicted to present two domains: The regulatory response domain (amino acids 3 – 114) and the effector domain (amino acids 140 – 217). Predicted receiver and effector domains are in grey and turquoise, respectively. **B)** Primary sequence of CopR highlighted the two domains that were cloned. Blue and red represent β-sheet and α-helix predictions, respectively, from the PROCHECK evaluation of the SWISS-model prediction of the CopR structure.

One of the main concerns when attempting expression of protein domains is their instability, which can arise, for example, if secondary structure is disrupted. To this end, the bioinformatically identified domains (Figure 3.6) were extended (Figure 3.47 B) to ensure that these predicted secondary structures were maintained intact upon cloning. This is an attempt at preventing domain instability, though based on secondary structure predictions and CopR model structure, in the absence of experimentally determined structural data. The defined CopR domains are: CopR_N, amino acids 1-120 and CopR_C, amino acids 122 – 221. The linker region, which spans residues 120-125 was not integrated as part of either domain, as its predicted random coiled structure might decrease the stability of the domains.

3.3.1.2 Protein Homology and Protein Structural Model

The previously performed sequence alignment for the full-length CopR protein is sufficient for identification of the conserved amino acids in each of the domains. Even if the chosen sequences are not the most closely related with the domains, amino acids which are essential for protein function should be easily identified as they would be evolutionarily maintained.

```
CopR            MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQEDYRAAVLDLGLPDGNGL 60
ABO_1365        MRLLLVEDDYLLTNGLSAQLEKAGFSVDTARTAREARHLGQQESYRAGILDLGLPDGNGL 60
PST_0851        MRLLLVEDNVPLADELVASLSRQGYATDWLTDGRDAEYQGATEPYDLIILDLGLPGKPGL 60
Mmc1_0312       MRILVVEDHASLAAGLKKDLGAAGFVVDWAANAEEGAFMGREEPYDAVILDLGLPDDSGL 60
Alvin_0011      MRLLLVEDDPAQIAALLPALNAAGFAVDQAQDGAIGERLGETEPYDVIVLDLGLPKRPGL 60
MAMP_00174      MRLLLVEDDPLLGPNLQQALNKAGFATDLSADGIDGEAMGEIEPYDLIVLDLGLPGKPGL 60
MDG893_10211    MRLLLVEDDRLLADGLSRQLEKAGFSVDTTHTAREAMMLARQEEYRAIILDLGLPDGNGL 60
MELB17_11654    MRLLLVEDDRLLAEGLASQLEKAGFSVDVTGTAKEAMLLGVQEDYRAAVLDLGLPDGNGL 60
                **:*:***.      *   *   *:  *    .  .   .  * *   :******   **

CopR            EVLKRWRSKLVECPVLVLTARGDWQDKVNGLKAGADDYLAKPFQTEELIARINALIRRSE 120
ABO_1365        DVLKQWRTHKVSFPVLILTARGDWQDKVNGLKAGADDYLAKPFQTEELIARLHAIVRRSE 120
PST_0851        EVLHAWRAAGVTTPVLILTARGSWAERIDGLKAGADDYLTKPFHPEELLLRIQALLRRAH 120
Mmc1_0312       NVLRGWRAAGVDVPVIILTAWDAWHQRVDGLQAGGDDYLGKPFHMEELIARLNALIRRRH 120
Alvin_0011      EVLRHWRARGLSLPVLILTARDAWPERVDGLKAGADDYLGKPFHVEELIARLNALTRRAA 120
MAMP_00174      EVLENWRRNENAVPVIILTARDAWEDKVLGFKAGADDYLAKPFQTEELIVRINAVLRRCS 120
MDG893_10211    DVLRKWRKDHIAFPVLILTARGDWHDKVEGLKAGADDYLAKPFQTEELIARLNAIVRRSE 120
MELB17_11654    DVLRKWRQDNANFAVLILTARGDWQDKVSGLKAGADDYLGKPFQAEELIARLNAIVRRSE 120
                :**. **      .*::*** . * ::: *::**.**** ***: ***: *::*: **
```

Figure 3.48 ClustalW[96] multiple sequence alignment of proteins which share above 50% identity with the primary sequence of the full-length RR, CopR, from *M. aquaeolei* VT8. Represented are 120 amino acids of the N-terminus of the protein, hereby identified as CopR_N with the amino acids of the active site identified. Legend: ABO_1365 from *Alcanivorax borkumensis* SK2, Mmc1_0312 from *Magnetococcus* sp. MC-1, MELB17_11654 from *Marinobacter* sp. ELB17, MDG893_10211 from *M. algicola* DG893, MAMP_00174 from *Methylophaga aminisulfidivorans* MP, PST_0851 from *Pseudomonas stutzeri* A1501, Alvin_0011 from *Allochromatium vinosum* DSM 180. For the refseq identifiers see Appendix B.4, Table 6.2. Asterisks (*), colons (:) and stops (.) below the sequence indicate identity, high conservation or conservation of the amino acids, respectively. Highlighted in grey are the amino acids identified as forming the active site, and in black is the presumed phosphorylation site.

As described in the introduction, there are numerous conserved amino acids in the receiver domain [42] which constitute the active site of the protein or are involved in phosphorylation mediated conformational change. These are highly conserved in the regulatory N-terminal domain, (Figure 3.48, highlighted in grey) and form the acidic pocket which coordinates a metal ion (for example Mg^{2+}) required for phosphorylation [30].

Similarly to CopR, determination of a model structure for the receiver and effector domains was performed as described before. For the CopR_N domain, the chosen template is the activated receiver domain of the PhoP protein from *E.coli* (PDB ID: 2PL1), which has 50% sequence identity with CopR_N, Figure 3.49. The Ramachandran plots, as well as the z-score plots, presented in Appendices C.9.2, Figures 6.11 and 6.13 B, serve to support this claim.

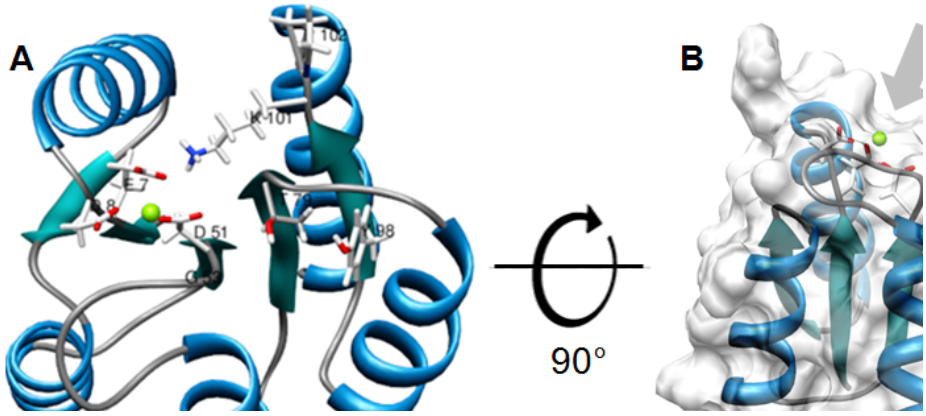


Figure 3.49 Model structure of the receiver domain of the RR CopR from *M. aquaeolei* VT8. **A)** The putatively conserved amino acids involved in phosphorylation-mediated conformational change and the phosphorylation site (D51) are identified as well as the magnesium ion. **B)** Indicates with a grey arrow the relative surface availability of D51 for phosphorylation. In blue are the α-helices, cyan are the β-sheets and in green is the presumed magnesium ion. Figure prepared using Chimera software [108].

In addition to the side chains of the amino acids identified in the active site (Figure 3.48 amino acids highlighted in grey) there is also a highly conserved small residue, Ala80 in the case of CopR, at the end of β4 which allows access, for the HK, Hpt or small molecule phosphodonor, to the phosphorylation site, shown in Figure 3.49 B and indicated by the grey arrow. A conserved threonine or serine, Thr79 for CopR, in the proximity stabilizes the phosphoryl group [137].

Analysis of the conserved features, which are characteristic of the effector domain, is slightly more complex. For the CopR_C domain the closest homolog with a known structure shares a sequence identity of 28% and belongs to the C-terminal domain of OmpR from *E.coli* (PDB ID: 2JPB). This value falls on the threshold accepted for model prediction [117] (Figure 3.50).

Each DNA-binding domain must be involved in specific interactions, such as recognition of a particular DNA sequence, interaction with the regulatory domain and with either the α subunit or the σ^{70} subunit of RNA polymerase [138]. Thus a low level of sequence homology is not entirely unexpected.

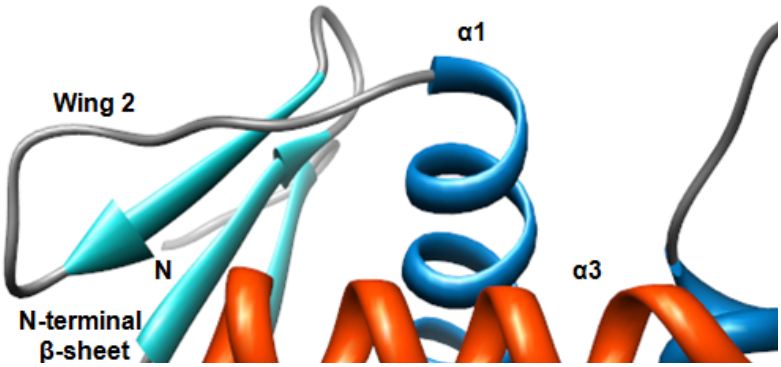


Figure 3.50 Model structure of the effector domain of the RR CopR from M. aquaeolei VT8. In blue are the α-helices, cyan are the β-sheets and in orange is the recognition helix. The a-loop and C-terminal hairpin characteristic of the OmpR/PhoB family are also identified. Figure prepared using Chimera software [108].

It seems that the effector domain of CopR contains a higher frequency, 17 or 22% (CopR_C or CopR_CHis$_6$) *vs.* 16% of the basic/positive amino acids over the full-length protein sequence (Table 3.6). This not only explains the higher theoretical pI, but also supports the rational that this region of the protein interacts with the acidic DNA backbone.

Table 3.6 Compilation of bioinformatic analysis results of calculated molecular masses (MM), pI and frequency of acidic and basic residues. Arginine, lysine and histidine were considered as acidic residues, whilst asparagine and glutamine were considered basic. *CopR without the His$_6$-tag. Data obtained from the ProtParam tool of the ExPASy server [94]. Note: In this case the His-tag was always located at the C-terminus of the protein.

	Amino acids	MM (kDa)	pI	Frequency of residues (%)	
				Acidic	Basic
CopR*	221	25.1	6.0	16	16
CopR_N	120	13.4	5.1	18	15
CopR_C	99	11.5	6.8	14	17

3.3.2 Construction of Heterologous Expression Plasmids for CopR Domains

For expression of each of the CopR domains two expression plasmids were constructed using the restriction enzyme cloning methodology. Expression plasmids, pCopR_NA/B and pCopR_CA/B, were constructed using the pET21 c (+) plasmid, and differ only in the presence or not of a C-terminal fusion tag, His$_6$, respectively.

3.3.2.1 Construction of Expression Plasmid

In order to amplify the *copR_N* and *copR_C* DNA fragments, a PCR was performed using the primers listed in Table 2.1, Chapter 2 Materials and Methods - Section 2.2. Primers for the pCopR_N/CA vectors were designed to incorporate a C-terminal His_6 tag to the protein sequence, which might enhance solubility and aids in protein purification (Figure 3.51, Lanes 1 and 3). The pCopR_N/CB expression vectors were constructed with DNA fragments cloned with restriction enzyme sites for their insertion into the pET-21 c (+) plasmid without the addition of a fusion tag (Figure 3.51, Lanes 2 and 4).

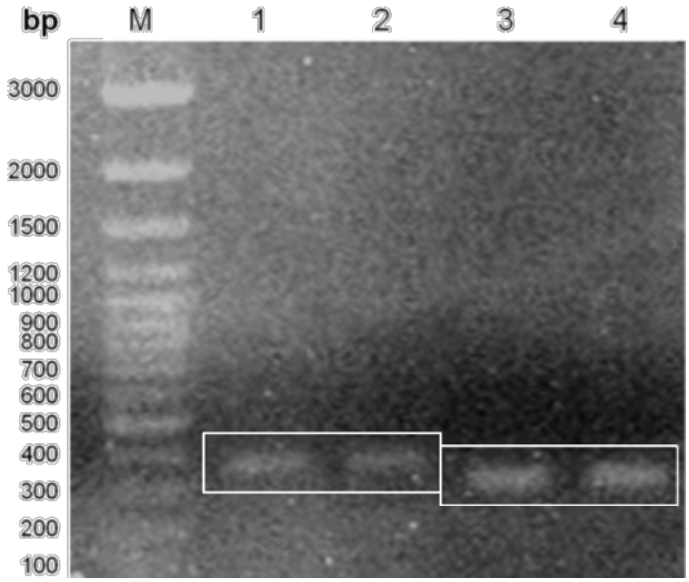


Figure 3.51 PCR amplification of the *copR_N (360 bp)* and *copR_CA/B (300 bp)* genes using the primers and PCR program described in Chapter 2 Materials and Methods – Section 2.2, Table 1.2. Legend: **M**: DNA ladder of 100 bp. **Lanes 1 and 2**: Amplified PCR fragments, *copR_CA* and *copR_CB*. **Lanes 3 and 4**: Amplified PCR fragments, *copR_NA* and *copR_NB*. The white boxed regions indicate the fragments of approximately 350 and 300 bp, respectively. 1% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

The amplified DNA fragments shown in Figure 3.51 correspond to the expected lengths of the desired fragments (*copR_N*, 360 bp and *copR_C*, 303 bp).

These DNA fragments were purified and cloned, using the restriction enzyme methodology, into pET-21 c (+) vectors (map of the vector in Appendix B.1 - Figure 6.2).

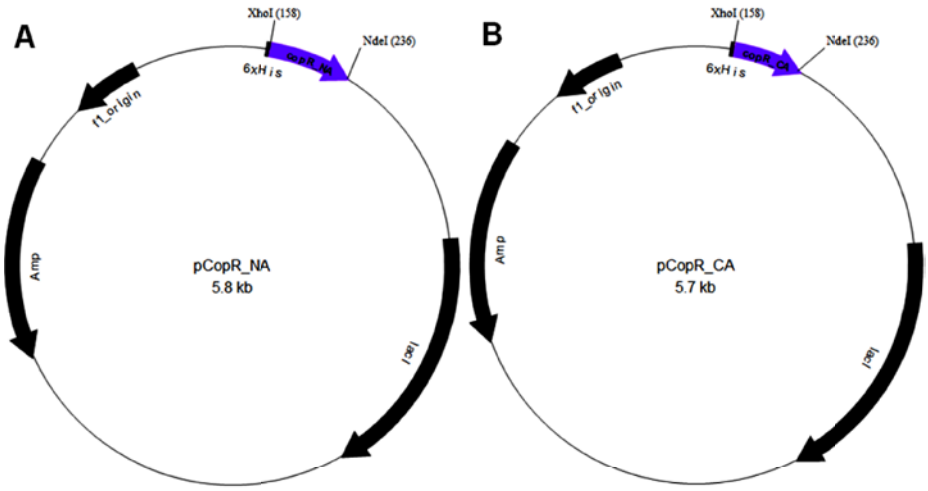


Figure 3.52 Cloning plasmids pCopR_NA (5.8 kb) and pCopR_CA (5.7 kb). The presented plasmids represent Strategy A expression vectors, where the inserted gene is not followed by a stop codon and as such the protein is expressed with a His_6-tag. Strategy B expression vectors are very similar, the distinguishing feature being the presence of the stop codon at the end of the inserted gene. Legend: *copR_NA* gene with 363 bp and *copR_CA* with 303 bp, in purple. Cloning at the restriction enzyme sites of XhoI (158 bp) and NdeI (236 bp) of pET21 c (+). Image created using BVTech plasmid, version 3.1 (BV Tech Inc., Bellevue, WA).

The resulting expression plasmids of 5.8 and 5.7 kb are schematically depicted in Figures 3.52 A and B, respectively. These plasmids, pCopR_NA/B and pCopR_CA/B, were transformed into a non-expression host, *E.coli* Giga Blue, and 2 of the colonies obtained from each plasmid were used to inoculate LB medium supplemented with 100 µg/mL ampicillin for plasmid amplification and isolation (Figure 3.53).

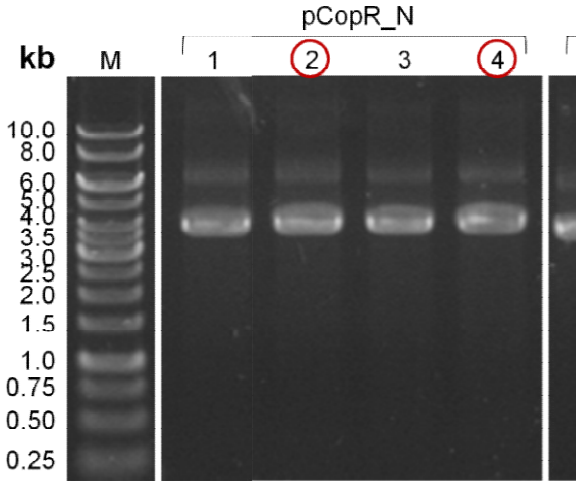


Figure 3.53 Clone screening. Legend: **M**: DNA ladder of 1 kb; **Lanes 1 and 2**: Plasmid DNA obtained from *E.coli* Giga Blue cells transformed with pCopR_N, strategy A plasmids; **Lanes 3 and 4**: Plasmid DNA obtained from *E.coli* Giga Blue cells transformed with pCopR_N, strategy B plasmids; **Lanes 5 and 6**: Plasmid DNA obtained from *E.coli* Giga Blue cells transformed with pCopR_C, strategy A plasmids; **Lanes 7 and 8**: Plasmid DNA obtained from *E.coli* Giga Blue cells transformed with pCopR_C strategy B plasmids. Clones from the lanes identified with a red circle were sent for DNA sequencing to check the DNA sequence of the inserted fragment. 0.8% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel stained with SybrSafe and visualized under UV light.

Upon sequencing the chosen expression vectors, Strategy A vectors pCopR_N2 and pCopR_C2; and strategy B vectors: pCopR_N2 and pCopR_C1, the DNA sequence was analyzed to verify insertion of the correct fragments. Confirmation for each of the domains is shown in Figures 3.54 A and B, where the translated DNA sequence of all of the inserted fragments correspond to the theoretical constructs, and share 100% identity, as do each of the DNA sequences.

A) CopR_N

```
Maqu_0124       MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQEDYRAAVLDLGLPDGNGL 60
CopR_N          MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQEDYRAAVLDLGLPDGNGL 60
                ************************************************************

Maqu_0124       EVLKRWRSKLVECPVLVLTARGDWQDKVNGLKAGADDYLAKPFQTEELIARINALIRRSE 120
CopR_N          EVLKRWRSKLVECPVLVLTARGDWQDKVNGLKAGADDYLAKPFQTEELIARINALIRRSE 120
                ************************************************************
```

B) CopR_C

```
Maqu_0124       VHSQVKAGGFELDENRQSLRTEEGAEHALTGTEFRLLRCLMSRPGHIFSKEQLMEQLYNL 60
CopR_C          MHSQVKAGGFELDENRQSLRTEEGAEHALTGTEFRLLRCLMSRPGHIFSKEQLMEQLYNL 60
                :************************************************************

Maqu_0124       DESPSENVIEAYIRRLRKLVGNETITTRRGQGYMFNAQR- 99
CopR_C          DESPSENVIEAYIRRLRKLVGNETITTRRGQGYMFNAQR- 99
                *****************************************
```

Figure 3.54 Alignment of the amino acid sequences obtained by translating (ExPASy translate tool [94]) the selected sequenced vectors, against the reference gene *M. aquaeolei* VT8, Maqu_0124. Alignment was performed using ClustalW (version 2.0) [96]. Legend: Asterisks (*) and colons (:) below the sequence indicate identity of the amino acids.

The plasmid DNA from the sequenced expression vectors previously identified were used to transform the expression host cells *E.coli* BL21(DE3) as detailed in Chapter 2 Materials and Methods – Section 2.2.5.1.

3.3.3 Expression and Protein Purification

3.3.3.1 Heterologous Test Expressions and Solubility Assays

A. Heterologous Test Expressions for CopR_NA/B

As this is the first attempt at expressing these domains, various conditions were assayed as detailed in Chapter 2 Materials and Methods – Section 2.2.5.2, Table 2.3. Assays for the optimal expression conditions were carried out in 50 mL growths of LB medium with 100 µg/mL ampicillin.

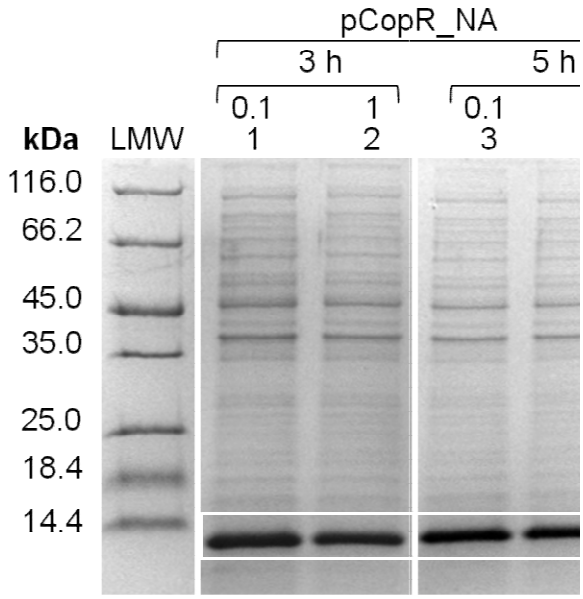


Figure 3.55 SDS-PAGE analysis of the expression profile of cells transformed with pCopR_NA/B (14.5/13.4 kDa) in LB growth medium, at 37°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: 3 h growth of cells transformed with pCopR_NA, after induction with 0.1 mM IPTG; **Lane 2**: 3 h growth of cells transformed with pCopR_NA, after induction with 1 mM IPTG; **Lane 3**: 5 h growth of cells transformed with pCopR_NA, after induction with 0.1 mM IPTG; **Lane 4**: 5 h growth of cells transformed with pCopR_NA, after induction with 1 mM IPTG; **Lane 5**: 3 h growth of cells transformed with pCopR_NB, after induction with 0.1 mM IPTG; **Lane 6**: 3 h growth of cells transformed with pCopR_NB, after induction with 1 mM IPTG; **Lane 7**: 5 h growth of cells transformed with pCopR_NB, after induction with 0.1 mM IPTG; **Lane 8**: 5 h growth of cells transformed with pCopR_NB, after induction with 1 mM IPTG. The white boxed regions show the expression of the CopR_NHis$_6$ and CopR_N proteins. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

For analysis of these expression results one needs to take into account that the theoretically determined molecular masses of CopR_NA (CopR_NHis$_6$) and CopR_NB (CopR_N) are 14.5 and 13.4 kDa, respectively. The amount of cells loaded onto each of the gels was normalized, in order to allow the relative comparison between conditions.

It is apparent from the cells grown at 37°C (Figure 3.55) that CopR_NA/B are the predominant components in the cellular extract, and also that their expression does not seem to be affected by the assayed conditions. So, when the growths were performed at 16°C, the conditions tested were merely growth of cells for 3 h upon induction with 0.1 mM IPTG.

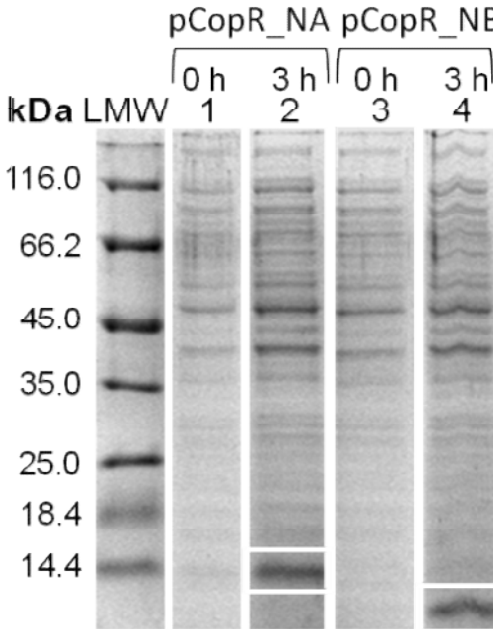


Figure 3.56 SDS-PAGE analysis of the expression profile of cells transformed with pCopR_NA/B (14.5 and 13.4 kDa) in LB growth medium, at 16°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Cells transformed with pCopR_NA, before induction; **Lane 2**: Cells transformed with pCopR_NA, before induction; **Lane 3**: 3 h growth of cells transformed with pCopR_NB, after induction with 0.1 mM IPTG; **Lane 4**: 3 h growth of cells transformed with pCopR_NB, after induction with 0.1 mM IPTG. The boxed regions highlight the desired proteins. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

For the cells grown at 16°C (Figure 3.56) it is apparent that there is only expression of the target proteins upon induction (Lanes 2 and 4), and that their expression after 3 h is significant when compared to the *E.coli* proteins (35 or 45 kDa).

At this point, growth of the desired proteins at 37°C would yield a greater quantity of protein, however, solubility assays are necessary to verify if these are present in the soluble cell extract or not. It is unnecessary to evaluate the solubility of the proteins in all the assayed conditions, given that the quantity of expressed proteins seems to be independent of the concentration of IPTG used and length of induction, only the cells grown at 16 and 37°C for 3 h after induction with 0.1 mM were lysed.

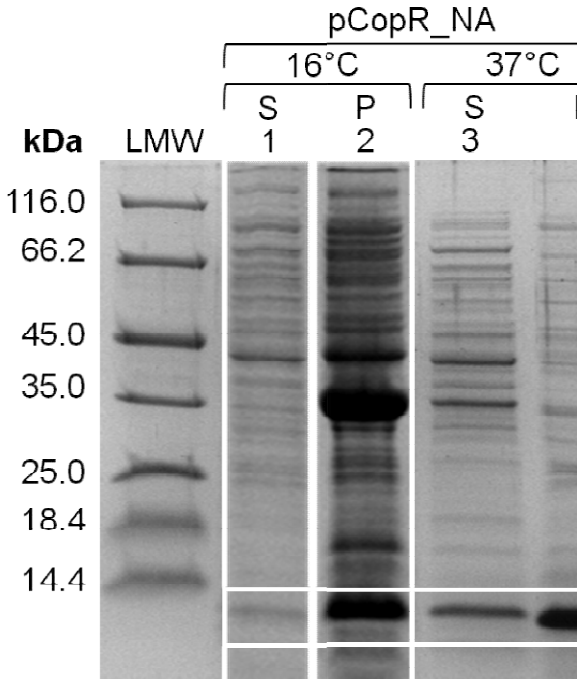


Figure 3.57 Solubility of CopR_NA/B proteins (14.5/13.4 kDa) in LB medium after 0.1 mM IPTG induction and 3 h growth at 37°C. Legend: **LMW:** Low molecular weight marker; **Lane 1**: Soluble extract of cells transformed with pCopR_NA at 16°C; **Lane 2**: Pellet of cells transformed with pCopR_NA at 16°C; **Lane 3**: Soluble extract of cells transformed with pCopR_NA at 37°C; **Lane 4**: Pellet of cells transformed with pCopR_NA at 37°C; **Lane 5**: Soluble extract of cells transformed with pCopR_NB at 16°C; **Lane 6**: Pellet of cells transformed with pCopR_NB at 16°C; **Lane 7**: Soluble extract of cells transformed with pCopR_NB at 37°C; **Lane 8**: Pellet of cells transformed with pCopR_NB at 37°C. The boxed regions correspond to the expressed CopR_NHis$_6$ and CopR_N protein domains. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

For the cells grown at 37°C, cell lysis was performed as described previously using the BugBuster protocol, however for the cells grown at 16°C the cells were lysed in a French press. This alteration of protocol is due to the fact that large-scale growths were performed prior to the solubility assays, and it was assumed that the growths at 37°C would yield insoluble protein as previously observed for the full length protein. However, this was shown to be an incorrect assumption, as Figure 3.57 shows.

The degree of solubility of the target proteins in the cells grown at 37°C (Figure 3.57, Lanes 3,4, 7 and 8) is reduced, with the majority of the protein being in fact found in inclusion bodies or unbroken cells (Lanes 4 and 8). Moreover, a similar result was obtained when protein production was performed at 16ºC (Figure 3.57, Lanes 1, 2, 5 and 6), where cell lysis was inefficient as denoted by the presence of the *E.coli* protein which is expressed at 35 kDa in the pellet (Figure 3.57, Lanes 2 and 6). This greatly reduces the reliability of this assay. Though the preliminary results indicate that optimal expression occurs at 37°C with 3 h growth upon induction with 0.1 mM IPTG additional assays could indicate with more reliability the optimal expression conditions.

Whichever is the case, as was previously stated, the large-scale growth for protein purification was performed prior to this assay and as such cells were grown at 16°C for 3 h upon induction with 0.1 mM IPTG.

A. Heterologous Test Expressions for pCopR_CA/B

Expression assays for the CopR_C domain were performed as those for the CopR_N domain. The theoretical molecular masses of the CopR_CA (CopR_CHis$_6$) and CopR_CB (CopR_C) proteins were determined to be 12.6 and 11.5 kDa, respectively.

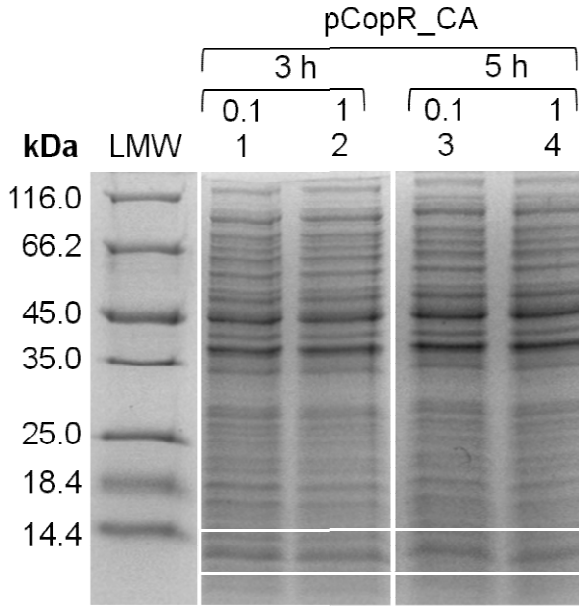


Figure 3.58 SDS-PAGE analysis of the expression profile of cells transformed with pCopR_CA/B (12.5/11.6 kDa) in LB growth medium, at 37°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: 3 h growth of cells transformed with pCopR_CA after induction with 0.1 mM IPTG; **Lane 2**: 3 h growth of cells transformed with pCopR_CA after induction with 1 mM IPTG; **Lane 3**: 5 h growth of cells transformed with pCopR_CA after induction with 0.1 mM IPTG; **Lane 4**: 5 h growth of cells transformed with pCopR_CA after induction with 1 mM IPTG; **Lane 5**: 3 h growth of cells transformed with pCopR_CB after induction with 0.1 mM IPTG; **Lane 6**: 3 h growth of cells transformed with pCopR_CB after induction with 1 mM IPTG; **Lane 7**: 5 h growth of cells transformed with pCopR_CB after induction with 0.1 mM IPTG; **Lane 8**: 5 h growth of cells transformed with pCopR_CB after induction with 1 mM IPTG. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

The results obtained for expression of the CopR_C domain at 37°C (Figure 3.58) are quite different from those obtained for the CopR_N domains at the same temperature. In these growths the CopR_CA/B domains are not the predominant protein, and it seems that expression of the CopR_CHis$_6$ protein (Figure 3.58, Lanes 1 to 4) is more efficient than that of CopR_C (Figure 3.58, Lanes 5 to 8).

Once more, when the growths at 16°C (Figure 3.59) were performed, only 3 h growth after induction with 0.1 mM IPTG were tested for both expression plasmids. However, in this

case the quantity of cells loaded onto the gel was insufficient for identification of CopR_C expression.

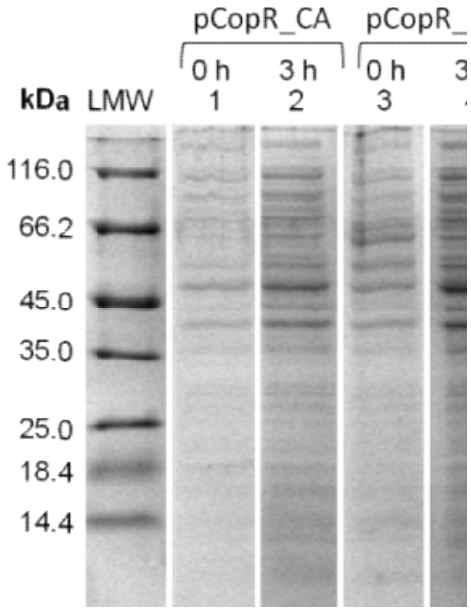


Figure 3.59 SDS-PAGE analysis of the expression profile of cells transformed with pCopR_CA/B (12.6/11.5 kDa) in LB growth medium, at 16°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Cells transformed with pCopR_CA before induction; **Lane 2**: Cells transformed with pCopR_CB, before induction; **Lane 3**: 3 h growth of cells transformed with pCopR_CA, after induction with 0.1 mM IPTG; **Lane 4**: 3 h growth of cells transformed with pCopR_CB, after induction with 0.1 mM IPTG. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

For identification of the expression conditions which produce soluble forms of the CopR_C domain, samples of 1 mL of cells from each of the growths were subjected to the BugBuster protocol and a larger quantity of cells were loaded onto the gel in order to visualize the presence, or not, of the desired protein and its solubility (Figure 3.60).

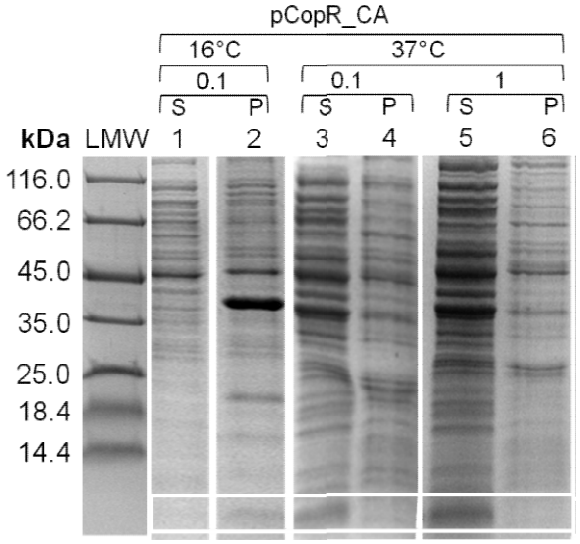


Figure 3.60 Solubility of CopR_CA/B domains (12.6/11.5 kDa) in LB medium under various conditions, after 3 h growth. Legend: **LMW:** Low molecular weight marker; **Lane 1**: Soluble extract of cells transformed with pCopR_CA after 0.1 mM IPTG induction at 16°C; **Lane 2**: Pellet of cells transformed with pCopR_CA after 0.1 mM IPTG induction at 16°C; **Lane 3**: Soluble extract of cells transformed with pCopR_CA after 0.1 mM IPTG induction at 37°C; **Lane 4**: Pellet of cells transformed with pCopR_CA after 0.1 mM IPTG induction at 37°C; **Lane 5**: Soluble extract of cells transformed with pCopR_CA after 1 mM IPTG induction at 37°C; **Lane 6**: Pellet of cells transformed with pCopR_CA after 1 mM IPTG induction at 37°C; **Lane 7**: Soluble extract of cells transformed with pCopR_CB after 0.1 mM IPTG induction at 16°C; **Lane 8**: Pellet of cells transformed with pCopR_CB after 0.1 mM IPTG induction at 16°C, **Lane 9**: Soluble extract of cells transformed with pCopR_CB after 0.1 mM IPTG induction at 37°C; **Lane 10**: Pellet of cells transformed with pCopR_CB after 0.1 mM IPTG induction at 37°C; **Lane 11**: Soluble extract of cells transformed with pCopR_CB after 1 mM IPTG induction at 37°C; **Lane 12**: Pellet of cells transformed with pCopR_CB after 1 mM IPTG induction at 37°C. The white boxes highlight the electrophoretic mobility of the desired proteins. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

At 16°C the produced proteins seem to be expressed in an insoluble form or are present in unbroken cells, being found in the pellet (Lanes 2 and 8 of Figure 3.60). Also, it seems that cells transformed with pCopR_CB produced lesser amounts of soluble protein, when compared to those transformed with pCopR_CA.

As previously, the large scale growths for protein purification were performed prior to this analysis and the chosen conditions for expression were 3 h growth upon induction with 0.1 mM IPTG at 37°C.

B. Expression of CopR_NHis$_6$ and CopR_CHis$_6$

Taking into account the previously obtained results, the plasmids of choice for large scale growth were those obtained by Strategy A, or in other words, the plasmids in which the expressed proteins are His$_6$ tag fusion proteins. This was due to the reduced quantity of these proteins in the soluble extract of the cell growths.

For the CopR_NA or CopR_NHis$_6$ protein, the large scale growth conditions were induction at an OD_{600nm} of 0.6, with 0.1 mM IPTG and growth at 16ºC for 3 h. Large scale growth

for the CopR_CA or CopR_CHis$_6$ protein was induction at 0.6 OD_{600nm}, with 0.1 mM IPTG and grown at 16ºC for 3 h. These conditions do not correspond to the optimal expression conditions, as described previously.

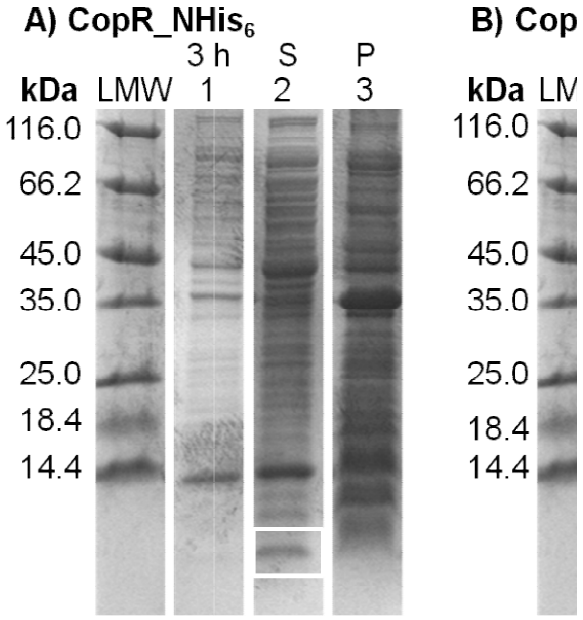


Figure 3.61 Expression and solubility of CopR_NHis$_6$ (Panel A) and CopR_CHis$_6$ (Panel B) in LB media upon cell growth at 16°C for 3 h after induction with 0.1 mM IPTG. **A) LMW**: Low molecular weight marker; **Lane 1**: Cells transformed with pCopR_NA after growth; **Lane 2**: Soluble extract of cells transformed with pCopR_NA; **Lane 3**: Pellet of cells transformed with pCopR_NA. B) **Lane 1**: Cells transformed with pCopR_CA after growth; **Lane 2**: Soluble extract of cells transformed with pCopR_CA; **Lane 6**: Pellet of cells transformed with pCopR_CA. The desired proteins of 14.5 and 12.6 kDa, CopR_NHis$_6$ and CopR_CHis$_6$, respectively are highlighted by the white boxes. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

Irrespective of the low expression of these domains in the determined conditions (Figure 3.61, bands identified) purification is feasible due to the presence of the His$_6$ tag.

3.3.3.2 Purification of CopR_NHis$_6$ and CopR_CHis$_6$

As mentioned previously, the proteins resulting from expression with the Strategy A are fusion proteins, containing the desired protein sequence, as well as, a His$_6$-tag at the C-terminus. This simplifies purification as affinity chromatography can be used (a Ni^{2+}-Sepharose matrix (His-Trap FF)).

A. Purification of CopR_NHis$_6$

After loading the soluble extract, unbound proteins were washed and eluted in steps of imidazole. All the collected fractions were analyzed by SDS-PAGE to determine which contained the desired protein (Figure 3.62).

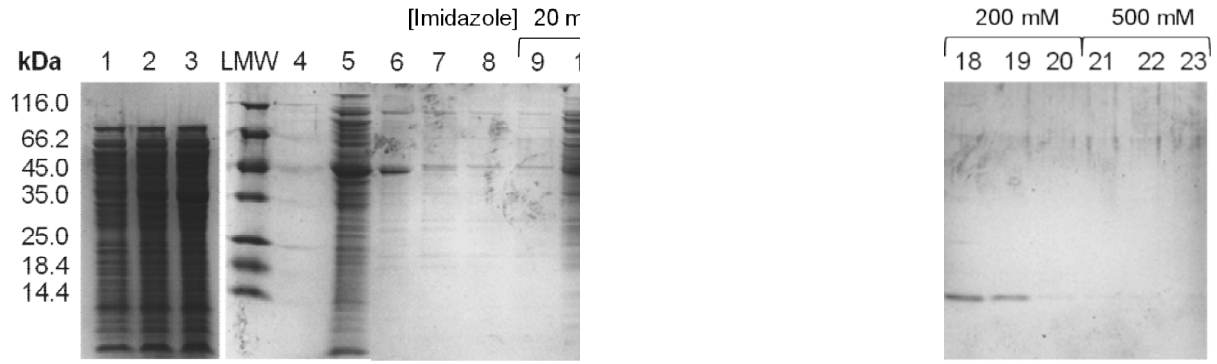


Figure 3.62 Purification of CopR_NHis$_6$ fusion protein by Ni^{2+}-sepharose affinity chromatography. Legend: **Lanes 1 to 3**: Column flow-through during sample application; **LMW**: Low molecular weight marker; **Lanes 4 to 8**: Samples of fractions after elution with 10, 20, 30, 40 and 50 mL column washing with 20 mM Tris-HCl (pH 7.6), 500 mM NaCl; **Lanes 9 and 10**: Samples of fractions after elution with 5 and 10 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 20 mM Imidazole; **Lanes 11 to 13**: Samples of fractions after elution with 5, 10 and 15 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 50 mM Imidazole; **Lanes 14 to 17**: Samples of fractions after elution with 5, 10, 15, and 20 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 100 mM Imidazole; **Lanes 18 to 20**: Samples of fractions after elution with 5, 10, and 15 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 200 mM Imidazole; **Lanes 21 to 23**: Samples of fractions after elution with 5, 10 and 15 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 500 mM Imidazole. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

Once more it is evident from analysis of lanes 1 – 3 of Figure 3.62 that the column was saturated with the desired protein of 14.5 kDa. For CopR_NHis$_6$ two fractions were formed, the first of higher purity by SDS-PAGE analysis, Fraction A (Figure 3.63 B, Lane A) which was obtained by pooling fractions 14 – 19 (protein eluted with 100 -200 mM imidazole), while fractions 13 and 14 (protein eluted with 50 – 100 mM imidazole) were pooled to form Fraction B (Figure 3.63 B, Lane B) of lower purity.

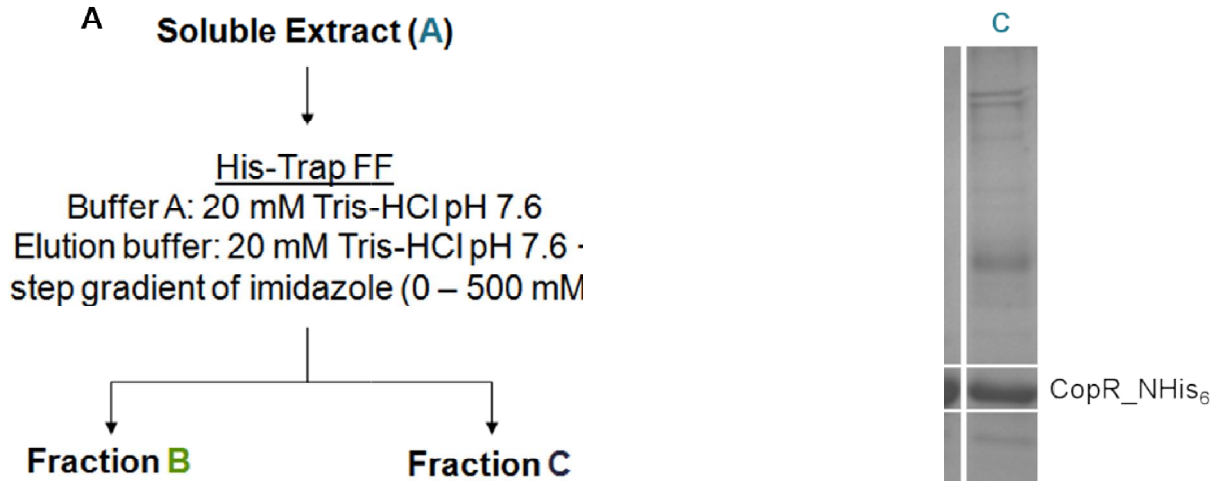


Figure 3.63 A) Flow-chart of the purification of the CopR_NHis$_6$ fusion protein. **B)** SDS-PAGE of the fractions obtained using the strategy in the flow-chart. **Lanes A and B**: Samples of the fractions identified in A. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

B. Purification of CopR_CHis$_6$

As previously, the soluble extract was loaded onto a Ni^{2+}-Sepharose column, and the protein was eluted similarly with increasing imidazole concentrations. The SDS-PAGE analysis of all the fractions collected was required to identify which contained the protein of interest (Figure 3.64).

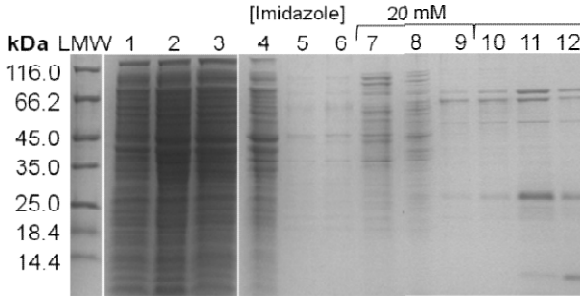


Figure 3.64 Purification of CopR_CHis$_6$ fusion protein by Ni^{2+}-sepharose affinity chromatography. Legend: **LMW**: Low molecular weight markers; **Lanes 1 to 3**: Column flow-through during sample application; **Lanes 4 to 6**: Samples of fractions after elution with 10, 20, 30 and 40 mL column washing with 20 mM Tris-HCl (pH 7.6), 500 mM NaCl; **Lanes 7 to 9**: Samples of fractions after elution with 10, 20 and 30 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 20 mM Imidazole; **Lanes 10 to 15**: Samples of fractions after elution with 5, 10, 15, 20, 25 and 30 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 50 mM Imidazole; **Lanes 16 to 21**: Samples of fractions after elution with 5, 10, 15, 20, 25 and 30 mL of 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 100 mM Imidazole; **Lanes 22 to 25**: Samples of fractions after elution with 5, 10, 15 and 20 mL 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 200 mM Imidazole; **Lanes 26 and 27**: Samples of fractions after elution with 7.5 and 15 mL 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 500 mM Imidazole. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

For the CopR_CHis$_6$ protein, fractions 13 - 20 were pooled to form the CopR_CHis$_6$ fraction (Figure 3.65 B, Lane A), which indicates that the CopR_CHis$_6$ protein was eluted with 50 - 100 mM imidazole.

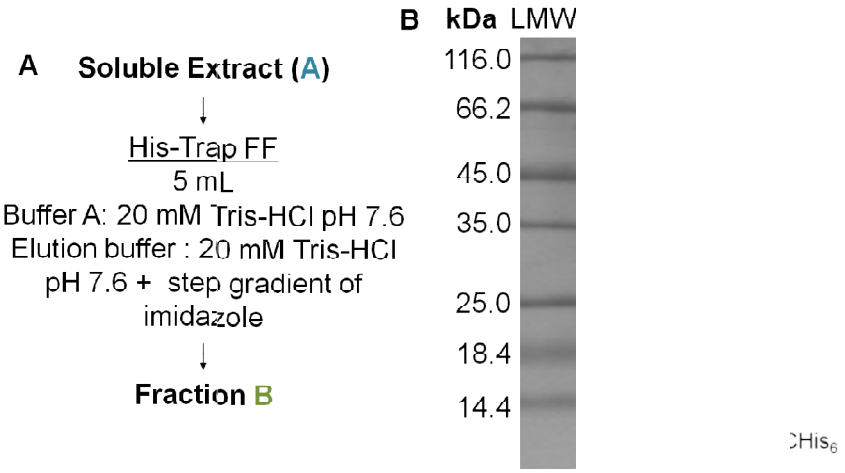


Figure 3.65 A) Flow-chart of the purification of CopR_CHis$_6$ fusion protein. **B)** SDS-PAGE of the fractions obtained using the strategy in the flow-chart in A. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

3.3.4 Biochemical Characterization of CopR_NHis$_6$ and CopR_CHis$_6$

3.3.4.1 Reducing and Non-reducing SDS-PAGE

Two cysteines exist in the full-length CopR protein, one located in each of the defined domains. One can hypothesize that when these domains are obtained independently, they may be more prone to forming intermolecular disulfide bonds.

Figure 3.66 Analysis of intermolecular disulfide bond presence and oligomeric association of CopR_CHis$_6$ (12.6 kDa) and CopR_NHis$_6$ (14.5 kDa) *via* SDS-PAGE, in reducing and non-reducing conditions. Legend: **LMW**: Low molecular weight marker; **Lane 1**: CopR_CHis$_6$ sample in the presence of SDS and β-MEtOH; **Lane 2**: CopR_NHis$_6$ sample in the presence of SDS and β-MEtOH; **Lane 3**: CopR_CHis$_6$ sample in the presence of SDS; **Lane 4**: CopR_NHis$_6$ sample in the presence of SDS. Black arrows identify the presumed protein domain monomers. SDS-PAGE prepared in Tris-Glycine buffer (10%, 150 V, 1 h) and stained with Coomassie-blue.

The bands which correspond to the reduced and denatured forms of CopR_NHis$_6$ and CopR_CHis$_6$ (Figure 3.66, Lanes 2 and 1, respectively) appear at approximately 14.4 kDa and below this value, respectively. Both of these are coherent with the expected electrophoretic mobilities for fragments with molecular masses of 14.5 kDa, CopR_NHis$_6$, and 12.6 kDa, CopR_CHis$_6$.

Analysis of Figure 3.66, lanes 3 and 4, a SDS-PAGE performed under reducing and non-reducing conditions. so that intermolecular disulfide bond presence may be identified, revealed the formation of disulfide bonds between CopR_NHis$_6$ domains, as well as the absence of these in the CopR_CHis$_6$ domain. In fact, in the case of CopR_NHis$_6$ a multitude of additional bands appear under these conditions (Figure 3.66, Lane 4), which can be interpreted as oligomers of the CopR_NHis$_6$ domain since there is an equal sized spacing between them, supporting the hypothesis that each of these bands corresponds to the addition of a CopR_NHis$_6$ monomer to the previous oligomer.

These results lead to the conclusion that the CopR_NHis$_6$ domain is obtained as a range of complexes, whilst the CopR_CHis$_6$ domain is not. As such, from this point forward the CopR_NHis$_6$ buffers will also contain DTT to a final concentration of 1 mM to avoid the formation of these bonds.

3.3.4.2 Stability Studies

Previously, precipitation of the unstable CopR protein was aggravated by concentration after purification. For the individual domains, this was not verified to the same extent, possibly due to the lower concentrations of these obtained after purification from the soluble cellular extract (CopRHis$_6$ between 3.7 and 6.4 mg/mL compared with CopR_NHis$_6$ at 0.4 mg/mL and CopR_CHis$_6$ at 0.16 mg/mL).

The obtained fractions from the respective purifications were concentrated down to 3 mL and divided into equal samples of 1.5 mL, which were then buffer exchanged into the buffers listed in Table 3.7.

Table 3.7 Buffer systems used for assaying the stability of CopR_NHis$_6$ and CopR_CHis$_6$. Fraction A and B of CopR_NHis$_6$ correspond to the two samples of CopR_NHis$_6$, the first with higher purity index and the second with a lower purity index as estimated from SDS-PAGE analysis.

	CopR_NHis$_6$	
	Fraction A	**Fraction B**
Buffer 1	20 mM Tris-HCl (pH 7.6)	20 mM Tris-HCl (pH 7.6)
Buffer 2	20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 10% glycerol	20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 10% glycerol
Buffer 3	20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 1 mM DTT, 10% glycerol	-
	CopR_CHis$_6$	
Buffer 1	20 mM Tris-HCl (pH 7.6)	
Buffer 2	20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 10% glycerol	

These samples were then maintained at 4°C, whereupon they were observed for precipitation after 24 h, after which they were maintained at -20°C and again observed for precipitation after 3 months. After 24 h and 3 months the samples were centrifuged at 14.000 rpm for 30 minutes and the quantity of pellet observed was compared between the buffers, Table 3.8.

Table 3.8 Observation of preferred buffer systems for the domains of CopR, CopR_NHis$_6$ and CopR_CHis$_6$. The choice of buffer was made upon comparing the pellet of each sample after centrifugation at 14.000 rpm for 30 min. CopR_NHis$_6$ A and CopR_NHis$_6$ B correspond to the two samples of CopR_NHis$_6$, the first with higher purity index and the second with a lower purity index.

	24 h	3 months
CopR_NHis$_6$ A	Buffer 2	Buffer 3
CopR_NHis$_6$ B	Buffer 1	Buffer 1
CopR_CHis$_6$	Buffer 1	Buffer 1

For long term storage of the domains it is recommended that Fraction A of CopR_NHis$_6$ be maintained in 20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 10% glycerol, 1 mM DTT; Fraction B of CopR_NHis$_6$ in 20 mM Tris-HCl (pH 7.6) and CopR_CHis$_6$ in 20 mM Tris-HCl (pH 7.6).

The marked difference in complexities of the CopR_NHis$_6$ and CopR_CHis$_6$ domain buffers shows that the instability associated with the full-length CopR protein is most probably due to instability of the receiver domain and not the effector domain.

3.3.4.3 Determination of CopR_NHis$_6$ and CopR_CHis$_6$ Concentrations

Only once was CopR_NHis$_6$ and CopR_CHis$_6$ isolated from LB growth medium in this work. Determination of protein concentrations were performed twice, with an interval of approximately 3 months, to facilitate the analysis of stability in each of the tested buffers.

The obtained concentrations of CopR_NHis$_6$ and CopR_CHis$_6$ from the purified soluble cellular extract were 0.28 and 0.06 mg/L of PB medium, respectively. Each of the samples was diluted three times and determination of the sample concentration was performed using the BCA method and a BSA calibration curve which was obtained as described in Chapter 2 Materials and Methods – Section 2.2.6.5.

These samples were maintained at -20°C for approximately three months, upon which their concentrations were recalculated, yielding the results summarized in Table 3.9.

Table 3.9 Concentration of the CopR_NHis$_6$ and CopR_CHis$_6$ protein domains as determined 24 h and 3 months upon purification. Protein concentration was determined using the BCA method.

	Concentration (mg/L)	Concentration after 3 months (mg/L)	Yield (%)
CopR_NHis$_6$ – Higher purity	0.42	0.10	23.8
CopR_NHis$_6$ – Lower purity	1.45	0.13	9.0
CopR_CHis$_6$	0.16	Bellow 0.06	-

For the CopR_NHis$_6$ protein domain, only the yield of the higher purity fraction is reliable. Due to the low purity of the CopR_NHis$_6$ protein domain, concentration for the lower purity sample is biased for actual CopR protein content. In the case of the higher purity sample,

24% pure protein was retrieved after purification from the LB extract and 3 months at -20°C. For the CopR_CHis$_6$ protein, sample determination of protein yield was inconclusive as determination of the protein concentration after 3 months at -20°C was not possible. In either case, the recovered protein is quite reduced, as was expected from the protein growths (Figure 3.63, lane 1 of Panel A and B).

3.3.4.4 Apparent Molecular Mass of CopR_NHis$_6$ and CopR_CHis$_6$

Determination of the apparent molecular mass for each domain was performed as described in Chapter 2 Materials and Methods – Section 3.1. This was performed for the purer fraction of CopR_NHis$_6$ and the CopR_CHis$_6$ fraction.

The obtained chromatograms are shown in Figure 3.67 and the calibration curve which was used for determination of the apparent molecular mass is presented in the insert of Figure 3.67.

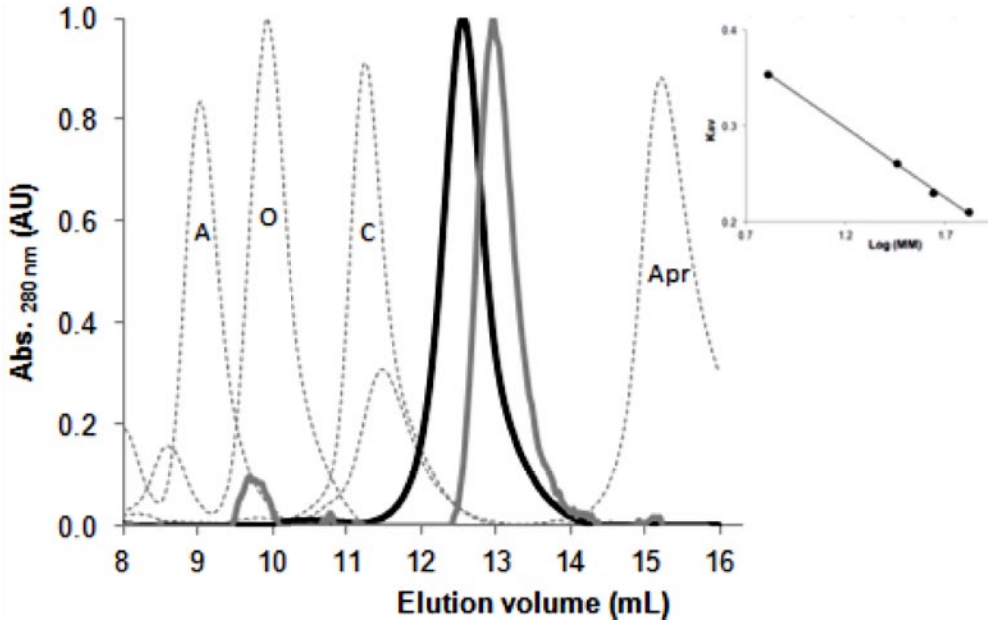


Figure 3.67 Elution profiles of calibration curve proteins on a Superdex 75 10/300 GL. Proteins were eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: **A**: Albumin, 66.8 kDa; **O**: Ovalbumin, 43.0 kDa; **C**: Chymotrypsinogen A, 25.0 kDa; **R**: Ribonuclease A, 13.7 kDa; **Apr**: Aprotinin, 6.5 kDa). Insert: Calibration curve for the molecular mass determination using LMW Gel Filtration Calibration Kit in Superdex 75 10/300 GL. Trend line equation obtained Kav = -0.14 log (MW) + 0.47, with an R^2 of 1.00. The profiles correspond to individual injections. Elution profiles of heterologously expressed and purified CopR_NHis$_6$ and CopR_CHis$_6$ protein domains in a Superdex 75 10/300 GL at 280 nm. Proteins were eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl at 4°C. Legend: Elution profile of CopR_NHis$_6$ (grey line) and of CopR_CHis$_6$ (black line).

The CopR_NHis$_6$ domain, of 128 amino acids and theoretical molecular mass of 14.5 kDa was eluted with an elution volume correspondent to a protein of 17.5 kDa. CopR_CHis$_6$ was eluted with a elution volume correspondent to a protein of 15.1 kDa, whereas the

theoretical molecular mass is of 12.6 kDa (Elution volumes presented in Appendix C.13.4, Table 6.15). The values obtained for the apparent molecular mass indicate that the proteins are eluted as monomers.

3.3.4.5 Circular Dichroism Spectroscopy

Circular dichroism spectroscopy for CopR_NHis$_6$ (Figure 3.68), the fraction of higher purity, and CopR_CHis$_6$ (Figure 3.69) samples were performed in order to verify the retention of secondary structure for each of these proteins. This analysis was hindered by the low concentrations of the proteins, 70 ng/mL for CopR_NHis$_6$ and 0.06 mg/mL for CopR_CHis$_6$, neither of which is within the optimal working conditions for CD (0.1 to 0.1 mg/mL). This is reflected in the spectra being measured from 200 nm and not 180 nm, which represents a loss of data necessary for optimal secondary structure prediction.

However, the analysis of the spectra was performed with the neural network program K_2D_2, which rendered theoretical curves very different from the experimental data. Nevertheless, one can verify the features of secondary structure, more pronounced in the case of CopR_NHis$_6$ (Figure 3.68). This is possible due to the characteristic peaks observed for each of these structures: 192, 209 and 222 nm for α-helix; 195 and 212 nm for random coil, and 186 and 211 nm for β-sheets. These peaks are really the transitions that are seen at these wavelengths, and in Appendix C.10, Figure 6.14, is an image of theoretical curves expected for pure secondary structures in CD.

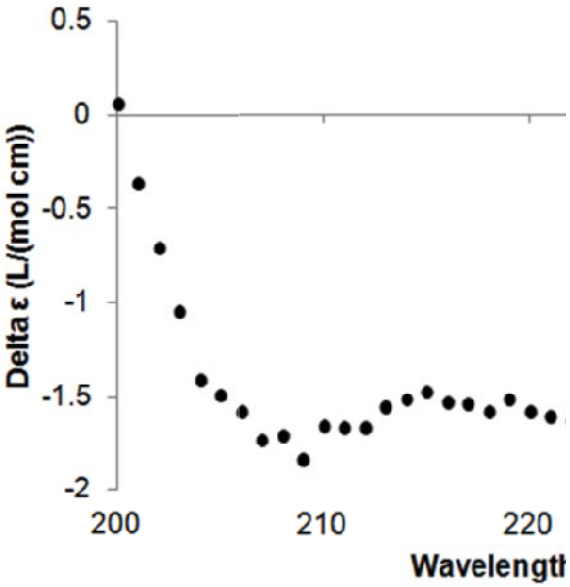


Figure 3.68 Experimental (· · · · ·) CD spectra of the CopR_NHis$_6$ protein in 5 mM phosphate buffer (pH 7.4), 250 mM NaCl, 10% glycerol. This spectrum shows the characteristic peaks at 209, 212 and 222 nm which denotes the presence of α-helices and random coil.

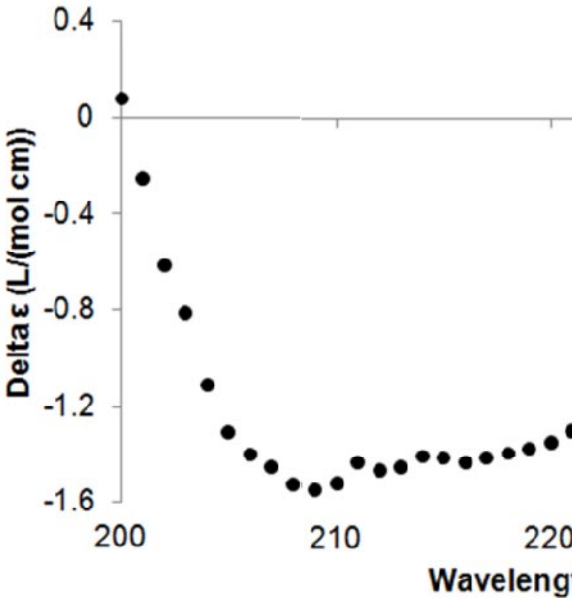


Figure 3.69 Experimental (· · · · ·) CD spectra of the CopR_CHis$_6$ protein in 5 mM phosphate buffer (pH 7.4), 250 mM NaCl, 10% glycerol. This spectrum shows the characteristic peaks at 209 and 212 nm which denotes the presence of α-helices and random coil.

Chapter 3 – Results and Discussion
Histidine Kinase CopS

3.4 C-terminal Domain of the Histidine Kinase - CopS_C

3.4.1 Bioinformatic Analysis

3.4.1.1 Full-length CopS Protein

The predicted translated protein, CopS, from the maqu_0123 gene, hereby denoted *copS*, of *M.aquaeolei* VT8 (Accession no. YP_957417.1) is annotated as an integral membrane sensor signal transduction histidine kinase. This family of proteins transfers phosphate groups to the conserved aspartate in receiver domains of RR upon stimulation of the periplasmic domain.

Similarly as was done for the *copR* gene, the *M. aquaeolei* VT8 CopS protein sequence was obtained by using the ExPASy bioinformatic tool [94], Translate (reading frames in Appendix C.2.2, Figure 6.6). The *copS* gene consists of 1329 bp, with a GC content of 57%, which upon translation yields the CopS protein sequence of 442 amino acids, as shown in Figure 3.70.

```
ATGCCCAACGTTAAAAGGCCAGCATCCGTCAAAGGCATGCTGCTGGTGTTATTGCTGCCCGCCGGCATCGCT
 M  P  N  V  K  R  P  A  S  V  K  G  M  L  L  V  L  L  L  P  A  G  I  A
CTGATGGGCGTAGCCTGGTTTGTCCACGGCCTGCTGCTGGATCGAATGTCCCGGGAATTCCTTGAAAGCCGC
 L  M  G  V  A  W  F  V  H  G  L  L  L  D  R  M  S  R  E  F  L  E  S  R
CTTAAAGACGAAGCTGCCTTTCTGGAGCACCAGATTCGCGAGGTCAGAGGTCAAGTCGAAACCCTGCAGACC
 L  K  D  E  A  A  F  L  E  H  Q  I  R  E  V  R  G  Q  V  E  T  L  Q  T
GGTGATTACTTCCAGGACGTCTTCCATCATGCCTTCGCCATACGCACGCCCGATCGAACCATCATATCGCCA
 G  D  Y  F  Q  D  V  F  H  H  A  F  A  I  R  T  P  D  R  T  I  I  S  P
AAGGCCTGGGAACCCTTGCTGGCACCACTGATCAACCATGAGCAGAATGGCACGCTTCGTCTTGAAGGCCGT
 K  A  W  E  P  L  L  A  P  L  I  N  H  E  Q  N  G  T  L  R  L  E  G  R
CAGGCGCCGGACAGTCCGTCCGATATCCTGGCATACCGTCACTCTTTTCAGGTGAACGGGTCACCGATCGTC
 Q  A  P  D  S  P  S  D  I  L  A  Y  R  H  S  F  Q  V  N  G  S  P  I  V
GTGGTTGTATCCGAGGATCTGGAGGCCCTGAAACGCAGTCAGGCCGAGCTGCACGCCTGGACAGCCGTGGTG
 V  V  V  S  E  D  L  E  A  L  K  R  S  Q  A  E  L  H  A  W  T  A  V  V
TCGGTGCTGTTGATCGCACTTCTGGTTGCCGTCATCTGGTTTGGCATCACGCTATCGCTGCGGCCGGTGGTG
 S  V  L  L  I  A  L  L  V  A  V  I  W  F  G  I  T  L  S  L  R  P  V  V
ACACTGAAAGCGGCGTTAAAACGATTGCAGGATGGAGAGATTTCCCGGATCAATGCACCATCACCCGAAGAA
 T  L  K  A  A  L  K  R  L  Q  D  G  E  I  S  R  I  N  A  P  S  P  E  E
TTTCAACCGCTCGTGATGCAGTTAAACCAGTTGCTTGACTCCCTCGACAAGCGACTGGTGCGTTCCCGGGAT
 F  Q  P  L  V  M  Q  L  N  Q  L  L  D  S  L  D  K  R  L  V  R  S  R  D
GCGCTCGCAAATCTGTCACACAGTGTCAAAACGCCCATCGCAGCCGTCCGGCAGATACTGGAGGATATGGAT
 A  L  A  N  L  S  H  S  V  K  T  P  I  A  A  V  R  Q  I  L  E  D  M  D
CGCCCATTGCCTAGTGATCTAAGAATCCAAATGGCCGCCCGTCTCAGTGACATCGACAGACAATTGGAAGCG
 R  P  L  P  S  D  L  R  I  Q  M  A  A  R  L  S  D  I  D  R  Q  L  E  A
GAAATGCGCCCGCAGCCGCTTTGCCGGGCCCCAGGTCGGGAAAAGCGCTTATCCCGTCAAACAGGCGCGGGAT
 E  M  R  R  S  R  F  A  G  P  Q  V  G  K  S  A  Y  P  V  K  Q  A  R  D
CTTTTGTGGATGCTGGGGCGGCTGTATCCGGAAAAATCCTTCGAACTATCAAGCTCACTGCCGGAAGACACC
 L  L  W  M  L  G  R  L  Y  P  E  K  S  F  E  L  S  S  S  L  P  E  D  T
CGCTGGCCGATAGAGGAGCATGACTTGAATGAAGTGCTAGGCAACCTGCTCGATAATGCCGGCAAATGGTCG
 R  W  P  I  E  E  H  D  L  N  E  V  L  G  N  L  L  D  N  A  G  K  W  S
TCGCGGTGCGTAGAACTCTCGCTGAAACAAGACAACAACAGCAGACAAATCGTTGTTTCTGATGATGGTCCT
 S  R  C  V  E  L  S  L  K  Q  D  N  N  S  R  Q  I  V  V  S  D  D  G  P
GGAGTCAATGGCGACGACTTGTCCAGTCTGGGGCAACGAGGGCTGCGACTTGACGAGCAGACCCCTGGCCAC
 G  V  N  G  D  D  L  S  S  L  G  Q  R  G  L  R  L  D  E  Q  T  P  G  H
GGTCTGGGCCTCGCGATTGTCCGGGAGATCGTCGCTCGCTATGAGGGCAACATTAGTTTTTCGACGGGGCCC
 G  L  G  L  A  I  V  R  E  I  V  A  R  Y  E  G  N  I  S  F  S  T  G  P
GGTAGCGGCTTGCGCGTAACCATAGAGTTTTAA
 G  S  G  L  R  V  T  I  E  F  -
```

Figure 3.70 DNA (1329 bp) and translated primary sequence (442 amino acids) of the maqu_0123 ORF from *M. aquaeolei* VT8. Highlighted in red are the start (ATG) and stop (TAA) codons.

The PSORTb (version 3.0) bioinformatic tool [97], which was further corroborated by other similar programs (for details see Appendix C.4, Table 6.6), predicted that CopS is a trans-

membrane protein. This is consistent with the expected localization of this protein, with a domain in the periplasm which senses an alteration in copper concentration and a cytoplasmatic domain which transmits this information to the RR protein via phosphorylation.

The MONSTER program [139] has the option for transmembrane helix prediction, and Figure 3.71 shows the secondary structure (obtained using Porter [101]) and transmembrane helix predictions (obtained using MONSTER).

```
MPNVKRPASV KGMLLVLLLP AGIALMGVAW FVHGLLLDRM SREFLESRLK DEAAFLEHQI REVRGQVETL 70
QTGDYFQDVF HHAFAIRTPD RTIISPKAWE PLLAPLINHE QNGTLRLEGR QAPDSPSDIL AYRHSFQVNG 140
SPIVVVVSED LEALKRSQAE LHAWTAVVSV LLIALLVAVI WFGITLSLRP VVTLKAALKR LQDGEISRIN 210
APSPEEFQPL VMQLNQLLDS LDKRLVRSRD ALANLSHSVK TPIAAVRQIL EDMDRPLPSD LRIQMAARLS 280
DIDRQLEAEM RRSRFAGPQV GKSAYPVKQA RDLLWMLGRL YPEKSFELSS SLPEDTRWPI EEHDLNEVLG 350
NLLDNAGKWS SRCVELSLKQ DNNSRQIVVS DDGPGVNGDD LSSLGQRGLR LDEQTPGHGL GLAIVREIVA 420
RYEGNISFST GPGSGLRVTI EF
```

Figure 3.71 Secondary structure prediction for CopS based on its primary sequence, performed by the Porter prediction server [101]. In blue and red are the predicted β-sheets and α-helices, respectively. Transmembrane helix prediction from the MONSTER prediction software [139] for the full-length CopS protein of *M.aquaeolei* VT8. Legend: Signal peptide in italic and underlined are the predicted transmembrane regions.

The SMART analysis showed that there are four domains, in addition to a signal peptide identified between residues 1 and 34.

1. The sensor domain, residues 37 to 164, which shares a low sequence homology with putative metal sensors, Figure 3.72, highlighted in purple.
2. A transmembrane helix is predicted between residues 162 and 186 (Figure 3.72), separating the sensor domain from the HAMP domain, which lies between residues 165 and 233 (Region in light grey in Figure 3.72).
3. The dimerization and phosphoacceptor domain is found between residues 237 and 298 (dark grey, Figure 3.72); and
4. The HK-like ATPase domain between residues 340 and 442, highlighted in mustard in Figure 3.72.

A protein BLAST search was performed and from the alignment in Figure 3.72, not only is the modularity of CopS evident, but also that the sensor domain is quite variable, even between proteins of sequence homology above 50%.

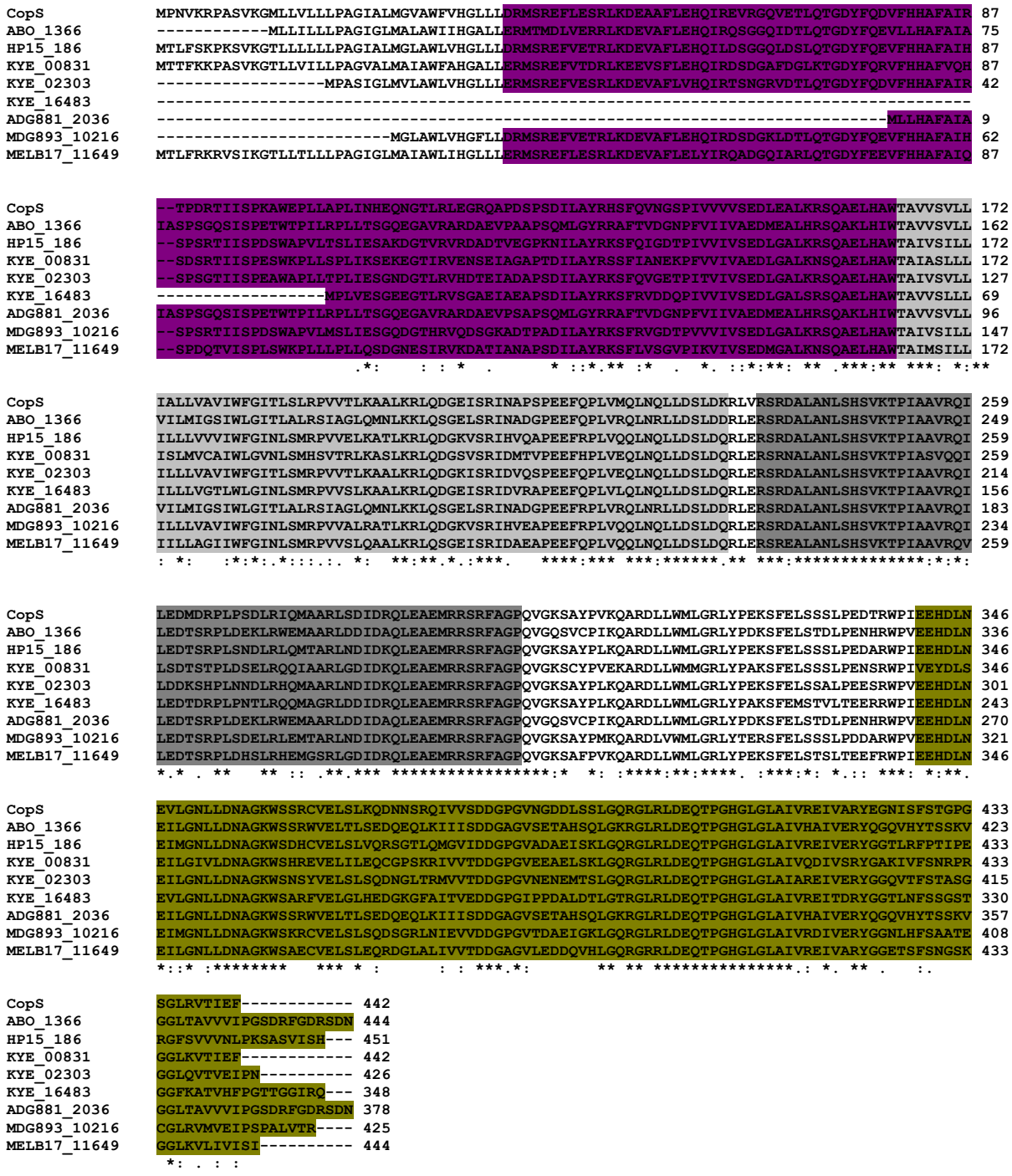


Figure 3.72 ClustalW multiple sequence alignment of proteins which share above 50% identity with the primary sequence of the histidine kinase, CopS (442 amino acids), from *M.aquaeolei* VT8. Legend: HP15_186 from – *M. adhaerens* HP15, ABO_1366 from *Alcanivorax borkumensis* SK2, MELB17_11649 from *M.* sp. ELB17, MDG893_10216 from *M.* algicola DG893, ADG881_2036 from *Alcanivorax* sp. DG881, KYE_02303, KYE_00831 and KYE_16483 from *M.* sp. MnI7-9.Asterisks (*), colons (:) and stops (.) below the sequence indicate identity, high conservation or conservation of the amino acids, respectively. Boxed regions correspond to the sensor domain (purple), HAMP domain (light grey), Dimerization and phosphoacceptor domain (dark grey) and HK-like ATPase domain (mustard).

3.4.1.2 Cytoplasmatic Domain of the CopS Protein

As part of a two component system, CopS is also modular in nature. The sensor domain varies considerably in terms of secondary structure fold and conserved motifs due to the wide range of stimuli with which these domains interact. However, given that the objective of this work is to analyze the cytoplasmic domain of CopS, CopS_C, the conservation should be evolutionarily maintained for dimerization and phosphotransfer of the phosphoryl group to the RR protein, which is clear from Figure 3.72.

Assuming that the MONSTER prediction for transmembrane helices is correct, the cytoplasmic domain of the protein is located between residues 187 and 442. However, the analysis of Figure 3.71, raises the question that disruption of a α-helix might compromise the stability of the protein, though the cloned DNA fragment contained the information for expression of the protein between residues 195 and 442 (Figure 3.73), when ideally it should have started at residue 205.

```
GCGGCGTTAAAACGATTGCAGGATGGAGAGATTTCCCGGATCAATGCACCATCACCCGAAGAATTTCAA
 A  A  L  K  R  L  Q  D  G  E  I  S  R  I  N  A  P  S  P  E  E  F  Q
CCGCTCGTGATGCAGTTAAACCAGTTGCTTGACTCCCTCGACAAGCGACTGGTGCGTTCCCGGGATGCG
 P  L  V  M  Q  L  N  Q  L  L  D  S  L  D  K  R  L  V  R  S  R  D  A
CTCGCAAATCTGTCACACAGTGTCAAAACGCCCATCGCAGCCGTCCGGCAGATACTGGAGGATATGGAT
 L  A  N  L  S  H  S  V  K  T  P  I  A  A  V  R  Q  I  L  E  D  M  D
CGCCCATTGCCTAGTGATCTAAGAATCCAAATGGCCGCCCGTCTCAGTGACATCGACAGACAATTGGAA
 R  P  L  P  S  D  L  R  I  Q  M  A  A  R  L  S  D  I  D  R  Q  L  E
GCGGAAATGCGCCGCAGCCGCTTTGCCGGGCCCCAGGTCGGGAAAAGCGCTTATCCCGTCAAACAGGCG
 A  E  M  R  R  S  R  F  A  G  P  Q  V  G  K  S  A  Y  P  V  K  Q  A
CGGGATCTTTTGTGGATGCTGGGGCGGCTGTATCCGGAAAAATCCTTCGAACTATCAAGCTCACTGCCG
 R  D  L  L  W  M  L  G  R  L  Y  P  E  K  S  F  E  L  S  S  S  L  P
GAAGACACCCGCTGGCCGATAGAGGAGCATGACTTGAATGAAGTGCTAGGCAACCTGCTCGATAATGCC
 E  D  T  R  W  P  I  E  E  H  D  L  N  E  V  L  G  N  L  L  D  N  A
GGCAAATGGTCGTCGCGGTGCGTAGAACTCTCGCTGAAACAAGACAACAACAGCAGACAAATCGTTGTT
 G  K  W  S  S  R  C  Q  I  V  V  S  D  D  G  P  G  V  N  G  D  D  L
TCTGATGATGGTCCTGGAGTCAATGGCGACGACTTGTCCAGTCTGGGGCAACGAGGGCTGCGACTTGAC
 S  S  L  G  V  E  L  S  L  K  Q  D  N  N  S  R  Q  R  G  L  R  L  D
GAGCAGACCCCTGGCCACGGTCTGGGCCTCGCGATTGTCCGGGAGATCGTCGCTCGCTATGAGGGCAAC
 E  Q  T  P  G  H  G  L  G  L  A  I  V  R  E  I  V  A  R  Y  E  G  N
ATTAGTTTTTCGACGGGGCCCGGTAGCGGCTTGCGCGTAACCATAGAGTTTTAA
 I  S  F  S  T  G  P  G  S  G  L  R  V  T  I  E  F  -
```

Figure 3.73 DNA and primary sequence of the Maqu_0123 ORF from *M. aquaeolei* VT8. Highlighted in red is the stop (TAA) codon.

The ProtParam tool from the ExPASy database[94] provided the information that the expected molecular mass of the translated DNA domain from this 744 bp fragment is 27.5 kDa, which has a theoretical isoelectric point of approximately 5.4 and an amino acid content, detailed in Table 6.5 of Appendix C.3.

There is an inherent difficulty in working with membrane bound proteins due to the difficulty in purifying the native proteins and heterologously express them. Therefore, the number of structures for full-length HKs of this type is very small. On the other hand, structures

of the C-terminal domain should be very similar for proteins of this type, given the degree of conservation necessary for this common phosphorylation scheme.

As shown in Figure 3.72 the cytoplasmatic region of the CopS protein, namely the HisKA and HK-like ATPase domains are conserved to a great degree in proteins that share identity above 50%. However, when a structure of a homologous protein is searched for, the resulting proteins share less than 20% homology, rendering any predicted structural model unsuitable for analysis.

3.4.2 Construction of the Heterologous Expression Vector for CopS_C

In order to amplify the DNA fragment which encodes the desired CopS fragment, a PCR was performed using the primers listed in Chapter 2 Materials and Methods – Section 2.2, Table 2.1. These primers were designed to clone the DNA fragment, using a Ligation Independent cloning methodology (LIC), into two expression vectors: pET-30 EK/LIC and pET-41 EK/LIC.

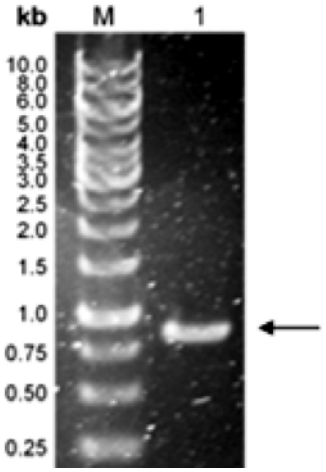


Figure 3.74 Purification of the *copS_C* DNA fragment (756 bp) resulting from PCR amplification using the primers and PCR program described in Chapter 2 Materials and Methods – Section 2.2. Legend: **M**: DNA ladder of 1 kb; **Lane 1**: Amplified PCR fragment, *copS*_C of approximately 800 bp, identified by the black arrow. 1% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

The obtained DNA fragment of approximately 800 bp (Figure 3.74) was cloned into the two plasmids: pET-30 EK/LIC and pET-41 EK/LIC (vector maps in Appendices B.2, B.3 – Figures 6.3 and 6.4), and constitute the expression vectors schematically represented in Figure 3. 75 A and B, respectively.

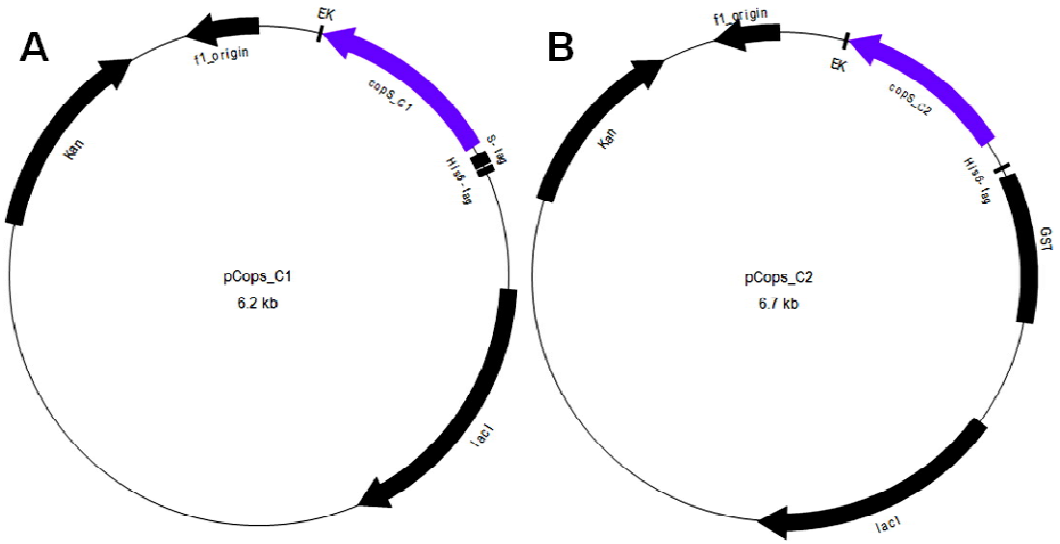


Figure 3.75 Cloning vectors for expression of the CopS_C protein **A.** pCopS_C1 (6.2 kb) and **B.** pCopS_C2 (6.7 kb). Legend: *copS_C* DNA fragment with 756 bp in purple. Cloning at the LIC sites into pET-30 EK/LIC and pET 41 EK/LIC. Image created using BVTech plasmid, version 3.1 (BV Tech Inc., Bellevue, WA).

The resulting plasmids, pCopS_C1 and pCopS_C2 of 6.2 and 6.7 kb, respectively, were transformed into a non-expression host, *E.coli* Giga Blue, and 4 of the colonies obtained from each vector were used to inoculate LB medium supplemented with 25 µg/mL kanamycin for plasmid amplification and isolation (Figure 3.76).

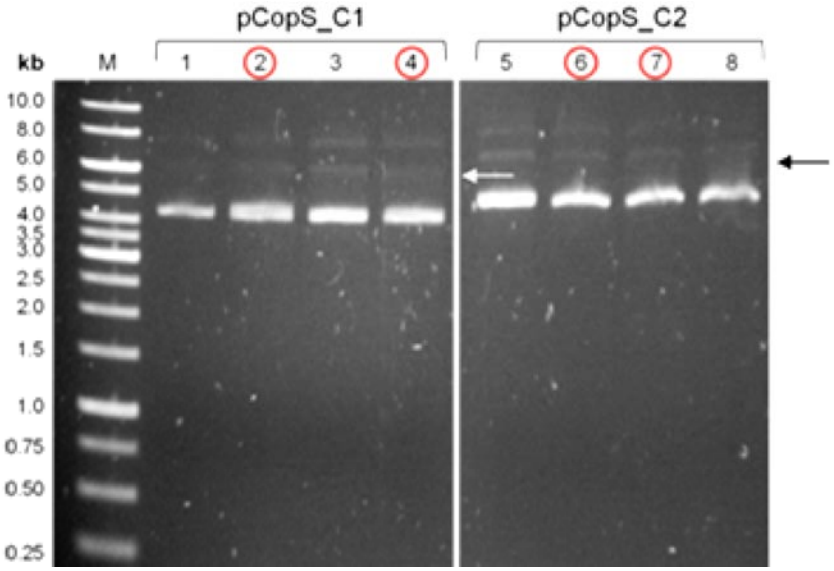


Figure 3.76 Clone screening. Legend: **M**: DNA ladder of 1 kb; **Lanes 1 to 4:** Plasmid DNA isolated from the 4 clones resulting from the transformation of *E.coli* Giga Blue with the pCopS_C1 plasmid; **Lanes 5 to 8**: Plasmid DNA from cells transformed with pCopS_C2; Arrows indicate the linear form of the plasmid DNA which migrate with approximate molecular masses of 6 and 7 kb. The plasmid DNA of the identified lanes with the circles were digested to check insertion of the desired DNA fragment. 0.8% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

The linear form of both plasmids, pCopS_C1 and pCopS_C2, migrate as fragments of approximately 6 and 7 kb, respectively. The supercoiled and nicked circle forms of the pCopS_C1 and pCopS_C2 plasmids migrate as fragments of approximately 4.5 and 5 kb, and approximately 5 and 8 kb, respectively.

The indicated clones (clones 2 and 4 of pCopS_C1 and clones 2 and 3 of pCopS_C2) were subjected to enzymatic digestion with the NdeI and XhoI enzymes for confirmation of insertion of the *copS_C* gene fragment (Figure 3.77).

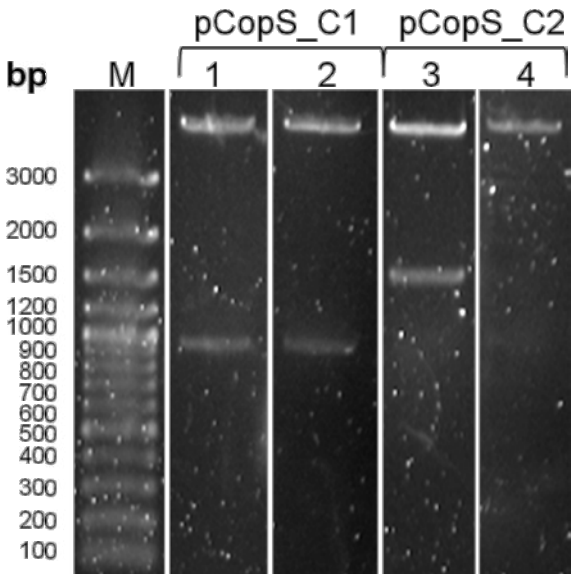


Figure 3.77 Digestion of the plasmid DNA retrieved from cells transformed with pCopS_C1 or pCopS_C2, with NdeI and XhoI restriction enzymes. Legend: **M**: 100 bp molecular marker; **Lane 1**: Clone 2 plasmid DNA from cells transformed with pCopS_C1; **Lane 2**: Clone 4 plasmid DNA from cells transformed with pCopS_C2;.**Lane 3**: Clone 2 plasmid DNA from cells transformed with pCopS_C2; **Lane 4**: Clone 3 plasmid DNA from cells transformed with pCopS_C2. 0.8% agarose gel, electrophoresis was run at 100 V for approximately 20 min. Gel was stained with SybrSafe and visualized under UV light.

Upon analysis of Figure 3.77 it seems that the pCopS_C2 3 plasmid (Figure 3.77, Lane 4) does not contain the expected DNA fragment. However, the remaining DNA fragments obtained from digestion with NdeI and XhoI seem to be coherent with the expected lengths. Due to the cloning method the fragments resulting from digestion of pCopS_C1 should have 1005 bp and those resulting from pCopS_C2 would have 1734 bp.

In order to confirm the sequence of these clones, clone 2 of pCopS_C1 and clone 1 of pCopS_C2 were sent for DNA sequencing.

```
CopS_C          AALKRLQDGEISRINAPSPEEFQPLVMQLNQLLDSLDKRLVRSRDALANLSHSVKTPIAA 60
Maqu_0123       AALKRLQDGEISRINAPSPEEFQPLVMQLNQLLDSLDKRLVRSRDALANLSHSVKTPIAA 60
                ************************************************************

CopS_C          VRQILEDMDRPLPSDLRIQMAARLSDIDRQLEAEMRRSRFAGPQVGKSAYPVKQARDLLW 120
Maqu_0123       VRQILEDMDRPLPSDLRIQMAARLSDIDRQLEAEMRRSRFAGPQVGKSAYPVKQARDLLW 120
                ************************************************************

CopS_C          MLGRLYPEKSFELSSSLPEDTRWPIEEHDLNEVLGNLLDNAGKWSSRCQIVVSDDGPGVN 180
Maqu_0123       MLGRLYPEKSFELSSSLPEDTRWPIEEHDLNEVLGNLLDNAGKWSSRCQIVVSDDGPGVN 180
                ************************************************************

CopS_C          GDDLSSLGVELSLKQDNNSRQRGLRLDEQTPGHGLGLAIVREIVARYEGNISFSTGPGSG 240
Maqu_0123       GDDLSSLGVELSLKQDNNSRQRGLRLDEQTPGHGLGLAIVREIVARYEGNISFSTGPGSG 240
                ************************************************************

CopS_C          LRVTIEF 247
Maqu_0123       LRVTIEF 247
                *******
```

Figure 3.78 Alignment of the translated protein sequence obtained by sequencing the selected vectors, against the translate of the reference gene of *M. aquaeolei* VT8, Maqu_0123. The obtained gene fragments were identical for *copS_C1* and *copS_C2*. Legend: Asterisks (*) and stops (.) below the sequence indicate identity or conservation of the amino acids, respectively. Alignment was performed using ClustalW (version 2.0) [96].

Examination of the DNA sequences, obtained by sequencing the chosen expression vectors, confirmed that the inserted DNA fragment shares 100% identity with Maqu_0123 from *M. aquaeolei* VT8, and the translated protein sequences, which also share 100% identity, when an alignment was performed with ClustalW (Figure 3.78).

The sequenced expression vectors were then used to transform the expression host cells, *E.coli* BL21(DE3), as detailed in Chapter 2 Materials and Methods for heterologous expression.

3.4.3 Heterologous Expression and Purification of CopS_C

There are two constructs for the CopS_C protein. The distinguishing features of the resulting expressed proteins from these vectors are that the fusion CopS_C proteins resulting from cells transformed with pCopS_C1 will have a S-tag and a His_6-tag at the N-terminus, while fusion proteins resulting from cells transformed with pCopS_C2 will have, in addition to the His_6 tag, a GST-tag also at the N-terminus.

3.4.3.1 Heterologous Test Expressions and Solubility Assays

The conditions tested for maximum soluble protein production were concentration of IPTG used for induction (0.1 and 1 mM), length of induction (up to 16 h) and temperature of growth (16 and 37°C). Small scale, 50 mL, cell growth was performed as described in Chapter 2 Material and Methods – Section 2.2.5.2, with a typical expression profile in LB media, as exemplified by Figures 3.79 and 3.80.

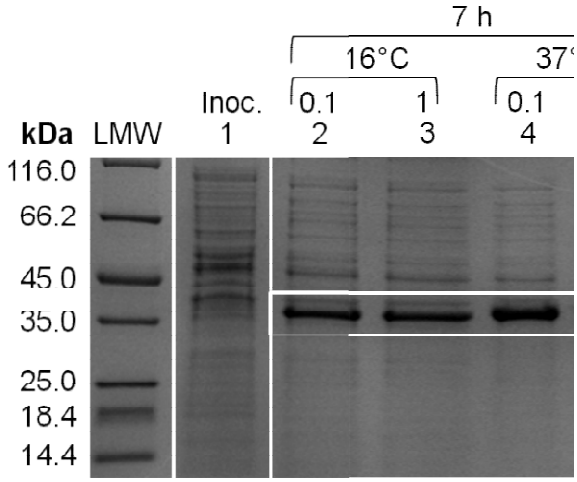


Figure 3.79 SDS-PAGE analysis of the expression profile of cells transformed with pCopS_C1 in LB growth medium, at 16 and 37°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Inoculum for cells transformed with pCopS_C1; **Lane 2**: 7 h growth of cells at 16°C, after induction with 0.1 mM IPTG; **Lane 3**: 7 h growth of cells at 16°C, after induction with 1 mM IPTG; **Lane 4**: 7h growth of cells at 37°C, after induction with 0.1 mM IPTG; **Lane 5**: 7 h growth of cells at 37°C, after induction with 1 mM IPTG; **Lane 6**: 16 h growth of cells at 16°C, after induction with 0.1 mM IPTG; **Lane 7**: 16 h growth of cells at 16°C, after induction with 1 mM IPTG; **Lane 8**: 16 h growth of cells at 37°C, after induction with 0.1 mM IPTG; **Lane 9**: 16 h growth of cells at 37°C, after induction with 1 mM IPTG. The boxed region indicates the protein expressed with the expected molecular mass of the expressed CopS_C1 (32 kDa). SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

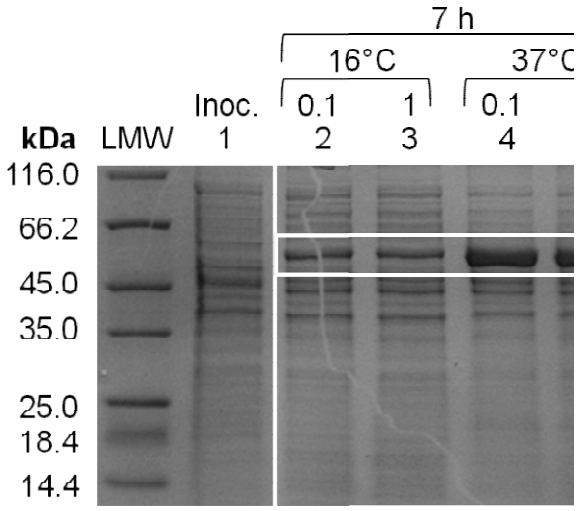


Figure 3.80 SDS-PAGE analysis of the expression profile of cells transformed with pCopS_C2 in LB growth medium, at 16 and 37°C. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Inoculum for cells transformed with pCopS_C2; **Lane 2**: 7 h growth of cells at 16°C, after induction with 0.1 mM IPTG; **Lane 3**: 7 h growth of cells at 16°C, after induction with 1 mM IPTG; **Lane 4**: 7 h growth of cells at 37°C, after induction with 0.1 mM IPTG; **Lane 5**: 7 h growth of cells at 37°C, after induction with 1 mM IPTG; **Lane 6**: 16 h growth of cells at 16°C, after induction with 0.1 mM IPTG; **Lane 7**: 16 h growth of cells at 16°C, after induction with 1 mM IPTG; **Lane 8**:16 h growth of cells at 37°C, after induction with 0.1 mM IPTG; **Lane 9**: 16 h growth of cells at 37°C, after induction with 1 mM IPTG. The boxed region indicates the protein expressed with the expected molecular mass of the expressed CopS_C2 (61 kDa). SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

For interpretation of the previous gels it is necessary to determine the expected molecular masses of the desired proteins. The CopS_C1 protein is expected to migrate in a

SDS-PAGE gel as a protein of approximately 32 kDa, while CopS_C2 as a protein of approximately 61 kDa (due to the addition of tags with 4.8 and 33.3 kDa, respectively, to the CopS_C protein of 27.5 kDa).

The amount of cells loaded onto each of the lanes in each gel was normalized, so it is possible to compare the results. First of all, the desired proteins are only expressed upon induction. The temperature of induction also has a marked influence on the quantity of protein produced, with this difference being more evident in the case of CopS_C2 (compare CopS_C2 bands in Lanes 2 and 4 or 3 and 5 of Figure 3.80, for example), and showing that heterologous expression is considerably increased at 37ºC. In terms of the effect of length of cell growth and protein expression, the effects are also evident, though much less significant. However, the impact of concentration of IPTG seems to be irrelevant for the expression of both clones and thus will be maintained at 0.1 mM IPTG.

The optimal conditions for protein expression, before the degree of solubility is determined, are:

1. For cells transformed with pCopS_C1, growth at 16°C after induction with 0.1 mM IPTG;
2. For cells transformed with pCopS_C2, growth at 16 or 37°C after induction with 0.1 mM IPTG.

In order to conclude which of the growth conditions chosen corresponds to the best conditions to obtain a high yield of soluble protein, solubility tests must be conducted. In this case, 1 mL of each of these growths, after 7 h induction, was subjected to the lysis buffer using the BugBuster protocol, as described in Chapter 2 Materials and Methods – Section 2.2.5.2 B.

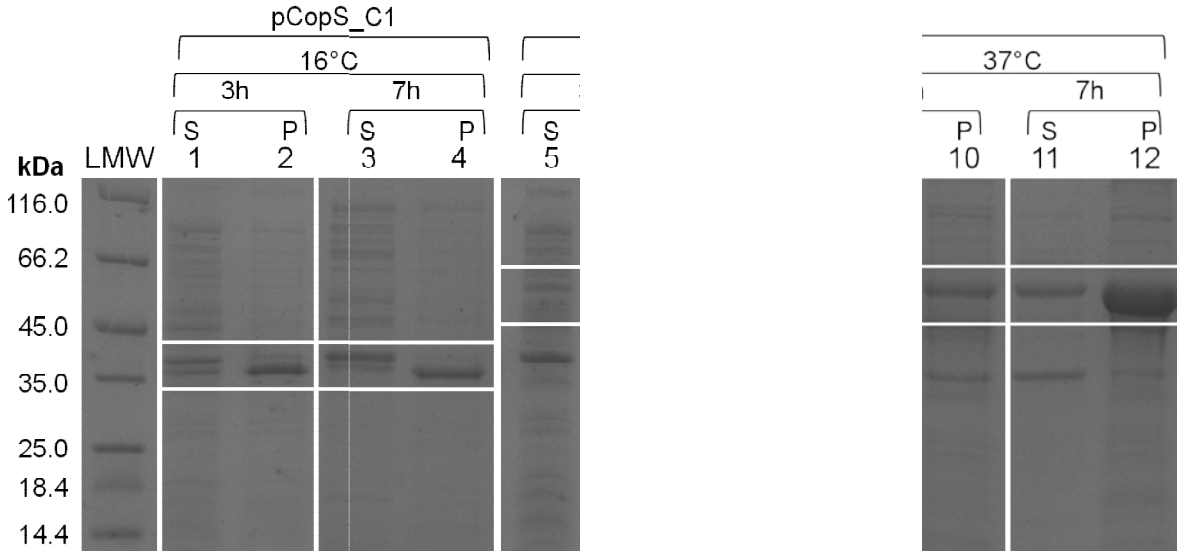


Figure 3.81 Solubility of the CopS_C1/2 (32/61 kDa) protein grown for 3 and 7 h in LB medium under various conditions after 0.1 mM IPTG induction. Legend: **LMW:** Low molecular weight marker; **Lane 1:** Soluble extract of cells transformed with pCopS_C1 and grown at 16°C for 3h; **Lane 2:** Pellet of cells transformed with pCopS_C1 and grown at 16°C for 3 h; **Lane 3:** Soluble extract of cells transformed with pCopS_C2 and grown at 16°C for 7 h; **Lane 4:** Pellet of cells transformed with pCopS_C2 and grown at 16°C for 7 h; **Lane 5:** Soluble extract of cells transformed with pCopS_C2 and grown at 37°C for 3h; **Lane 6:** Pellet of cells transformed with pCopS_C2 and grown at 37°C for 3 h; **Lane 7:** Soluble extract of cells transformed with pCopS_C1 grown at 16°C for 7 h; **Lane 8:** Pellet of cells transformed with pCopS_C1 grown at 16°C for 7 h; **Lane 9:** Soluble extract of cells transformed with pCopS_C2 grown at 16°C for 3 h; **Lane 10:** Pellet of cells transformed with pCopS_C2 grown at 16°C for 3 h; **Lane 11:** Soluble extract of cells transformed with pCopS_C2 grown at 37°C for 7 h; **Lane 12:** Pellet of cells transformed with pCopS_C2 grown at 37°C for 7 h. The target proteins are highlighted by the white boxes. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In all the conditions tested at 7 h after induction, it seems that the target proteins are mainly in the form of inclusion bodies, *i.e.* the desired proteins are present in the pellet of the cell extract (Figure 3.81). When expression is carried out for shorter periods of time, namely 3 h, the cells transformed with pCopS_C1 produce CopS_C1 still mainly in the form of an insoluble protein. Whereas, while cells transformed with pCopS_C2 also produce a high quantity of CopS_C2 in the insoluble fraction, an appreciable amount remains in the soluble extract of cells grown for 3 h at 37°C. Although the amount of protein present in the soluble extract is low, the fact that CopS_C2 is cloned with a solubility tag, which can also be used to purify the protein using affinity chromatography, overcomes the difficulty in purifying this protein using other chromatographic techniques.

Taking into account the previously obtained results, the plasmid used for large scale growth was pCopS_C2. Large scale growth for the CopS_C protein was optimally determined to be induction of cells transformed with pCopS_C2 at 0.6 OD_{600nm}, with 0.1 mM IPTG and grown at 37ºC for 3 h (Figure 3.82).

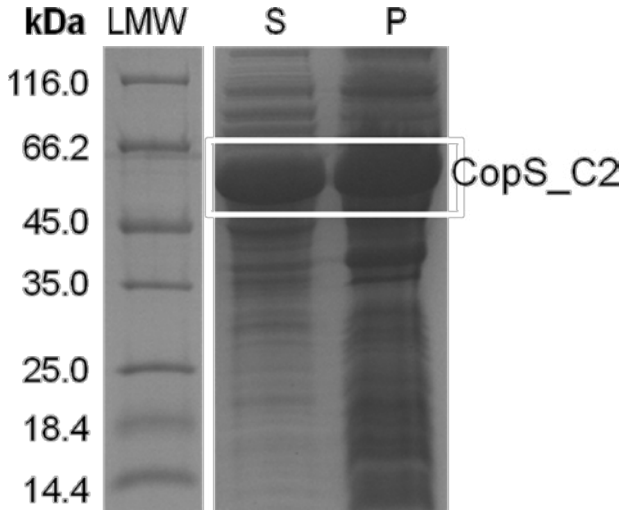


Figure 3.82 Solubility of CopS_C2 (61 kDa) in LB media upon induction with 0.1 mM IPTG at 37°C for 3 h. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Soluble fraction of cells; **Lane 2**: Pellet of cells. The arrow indicates the theoretical electrophoretic mobility expected from a protein of approximately 61 kDa. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

3.4.4 Purification of CopS_C

As mentioned, CopS_C2 was cloned as a fusion protein with a His_6 and a GST-tag at the N-terminus. The presence of the GST-tag enables purification using GST affinity chromatography, using a glutathione sepharose matrix (GSTrap).

After loading the soluble extract, the unbound proteins are washed, and the remaining proteins are eluted in steps of reduced glutathione (5-20 mM). All the fractions that were collected were analyzed by SDS-PAGE to determine which contained significant amounts of CopS_C2, Figure 3.83.

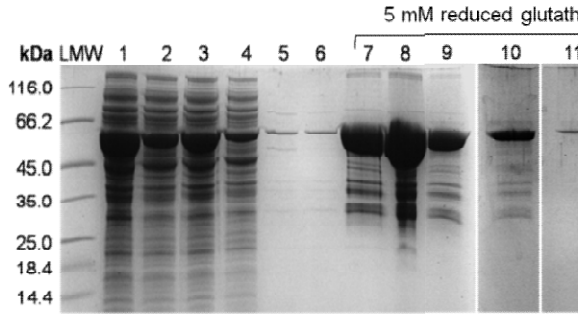


Figure 3.83 Purification of CopS_C2 fusion protein (61 kDa) by GST affinity chromatography. Legend: **LMW**: Low molecular weight marker; **Lane 1**: Sample of the injected soluble extract; **Lanes 2 and 3**: Flow-through upon sample injection; **Lanes 4 to 6**: Wash-through samples after 4, 20 and 40 mL 10 mM phosphate buffer (pH 7.4), 150 mM NaCl; **Lanes 7 to 12**: Samples of fractions after 4, 8, 12, 16, 20 and 24 mL of 50 mM Tris-HCl (pH 7.6), 5 mM reduced glutathione; **Lanes 13 to 18**: Samples of fractions after 4, 8, 12, 16, 20 and 24 mL of 50 mM Tris-HCl (pH 7.6), 10 mM reduced glutathione; **Lanes 19 to 24**: Samples of fractions after 4, 8, 12, 16, 20 and 24 mL 50 mM Tris-HCl (pH 7.6), 20 mM reduced glutathione. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

In this case, the pooled fractions were those corresponding to the samples loaded from lanes 7 through to 10, indicating that the CopS_C2 protein was eluted with 5 mM reduced glutathione. This sample was then concentrated, buffer exchanged and prepared for enzymatic digestion to remove the GST-tag.

3.4.4.1 CopS_C tag Cleavage

A. Tag cleavage and Purification of CopS_C

Due to cloning the CopS_C protein contains two N-terminal expression tags, which should be integrally removed by enzymatic cleavage with the HRV3C protease. This is possible as by design the expressed protein contains a protease cleavage sequence C-terminal to the tags and N-terminal to the desired protein.

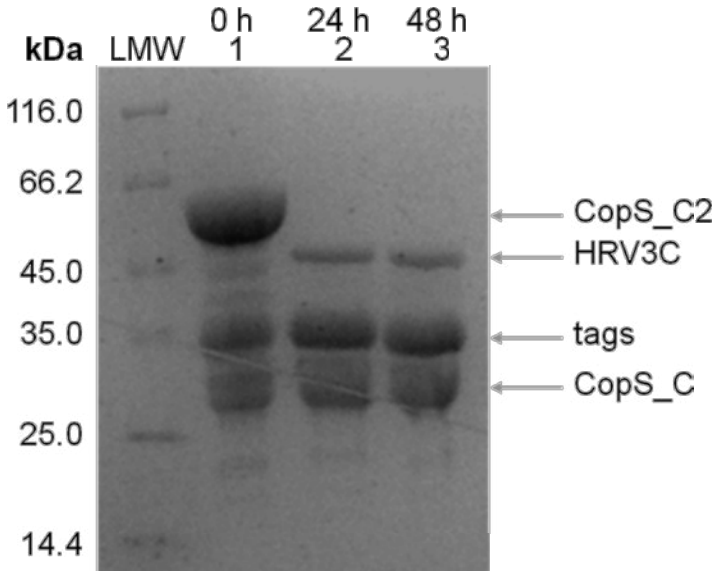


Figure 3.84 Enzymatic tag cleavage of the CopS_C2 protein (61 kDa) with 0.4 mg/mL HRV3C protease (54 kDa) in 20 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: **LMW:** Low molecular weight marker; **Lane 1**: Sample of CopS_C2 prior to addition of the protease; **Lane 2**: Sample of the enzymatic digestion after 24 h; **Lane 3**: Sample of the enzymatic digestion after 48 h. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

Enzymatic cleavage of the CopS_C2 protein (61 kDa) was efficient (Figure 3.84, Lane 2), with completion of digestion upon 24 h at 4°C.

The HRV3C protease, as mentioned previously, is also expressed as a fusion protein, with the GST-tag at the C-terminal of the protein. Therefore, since the uncut CopS_C and tag itself have a GST-tag, a GSTrap FF column was used to separate the CopS_C from these proteins (Figure 3.85).

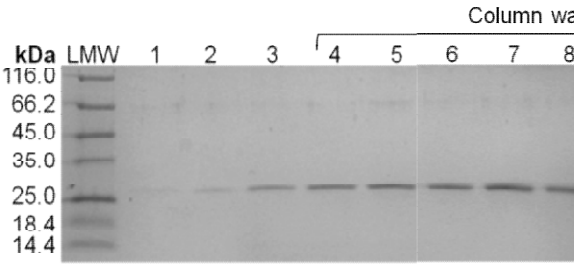


Figure 3.85 Purification of the CopS_C2 (28 kDa) protein after enzymatic cleavage. Legend: **LMW**: Low molecular weight marker; **Lanes 1 to 3**: Samples of fractions obtained upon sample injection; **Lanes 4 to 11**: Sample of fractions after elution with 5, 10, 15, 20, 25, 30, 35 and 40 mL column washing with 10 mM phosphate buffer (pH 7.4); **Lanes 12 to 15**: Samples of fractions after elution with 5, 10, 15, 20 and 25 mL of 50 mM Tris-HCl (pH 7.6), 500 mM NaCl, 10 mM reduced glutathione; **Lanes 16 to 18**: Samples of fractions after elution with 5, 10, and 15 mL of 50 mM Tris-HCl (pH 7.6), 500 mM NaCl, 300 mM Imidazole. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

Fraction 1 to 12 were then combined and buffer exchanged to 20 mM Tris-HCl (pH 7.6) and stored at -20°C for posterior use. Figure 3.86 depicts the progression of CopS_C purification from the soluble extract, including tag cleavage. The lanes identified in Figure 3.86 are also identified in the flow-chart of the purification of CopS_C2 fusion protein (61 kDa) on the next page, Figure 3.87.

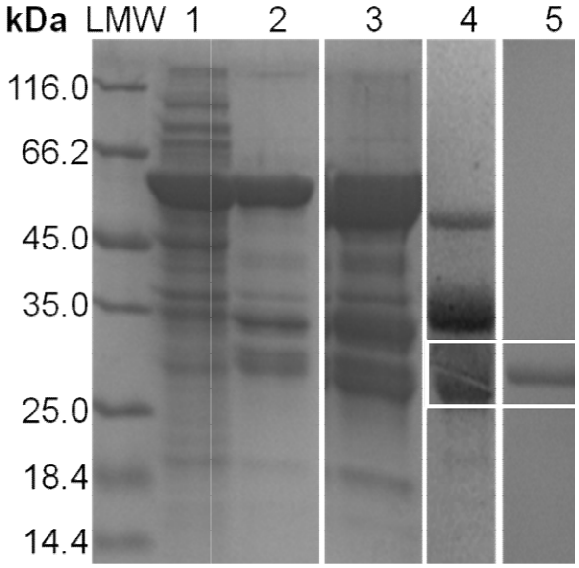


Figure 3.86 SDS-PAGE of the fractions obtained using the strategy in the flow-chart presented on the next page (Figure 3.87). **Lanes 1, 2, 3, 4,5 and 6** from the gel correspond to samples of the fractions identified in Figure 3.87. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

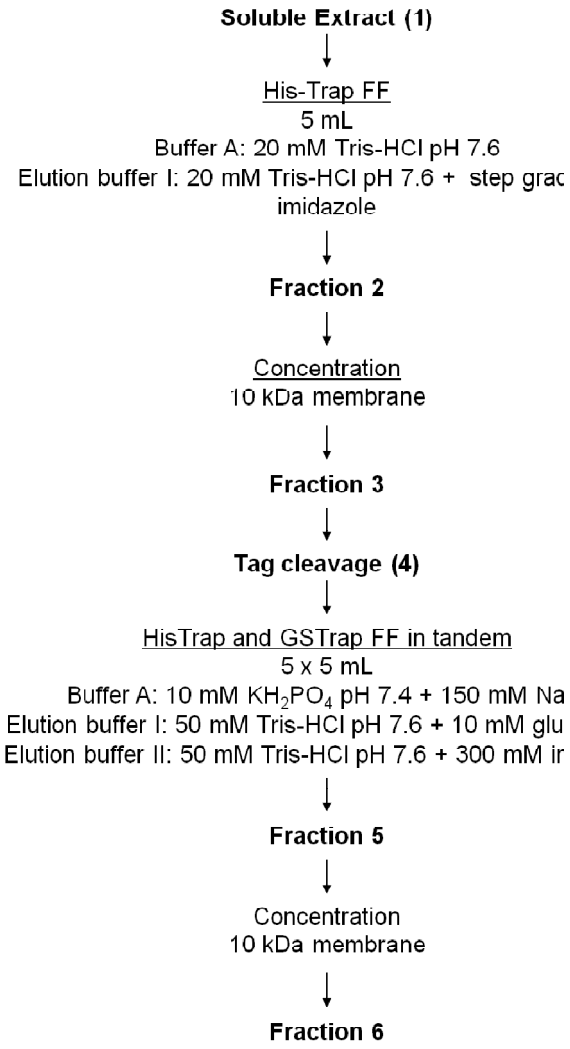


Figure 3.87 Purification flow-chart for the CopS_C fusion protein from the soluble cell extract of cells transformed with the pCopS_C2 plasmid and grown upon induction with 0.1 mM IPTG at 37°C for 3 h.

B. Determination of CopS_C Concentration

In order to calculate the efficiency of the entire purification process protein quantification has to be performed before and after cleavage of the tag from CopS_C2.

These concentrations were determined using the BCA method described in Chapter 2 Materials and Methods – Section 2.2.6.5. The concentration of CopS_C2 obtained from the purified soluble cellular extract was 7.7 mg/L of LB medium. Upon enzymatic cleavage, purification and consequent steps, CopS_C concentration was reduced to about 4.2 mg/L. In Figure 3.85, it is possible to observe that only a portion of the cleaved CopS_C protein was eluted with the addition of reduced glutathione to the buffer (lanes 13 and 14), which also eluted the fragments/protein with the tag. However, one must also take into account that the expressed and cleaved proteins have very different molecular masses, 61 kDa and 28 kDa respectively. If the results are viewed in terms of moles there seems to have occurred a loss of 40%. Also, it is

important to keep in mind that these results are obtained after various steps (two purification columns, two concentration processes and buffer exchanges), so a loss of 40% of the desired protein is not unexpected.

The global yield of the purification process is 4.2 mg of CopS_C per L of growth medium.

3.4.5 Biochemical Characterization of CopS_C

3.4.5.1 Molecular Mass of CopS_C

A. Reducing and Non-reducing SDS-PAGE

The CopS_C protein contains, as detailed in Appendix C.3, Table 6.5, only one cysteine residue throughout the length of the protein. This means that intramolecular disulfide bonds are not present, though intermolecular disulfide bonds may exist. However the analysis of the SDS-PAGE under reducing and non-reducing conditions (Figure 3.88) indicate that there is also no formation of intermolecular disulfide bonds.

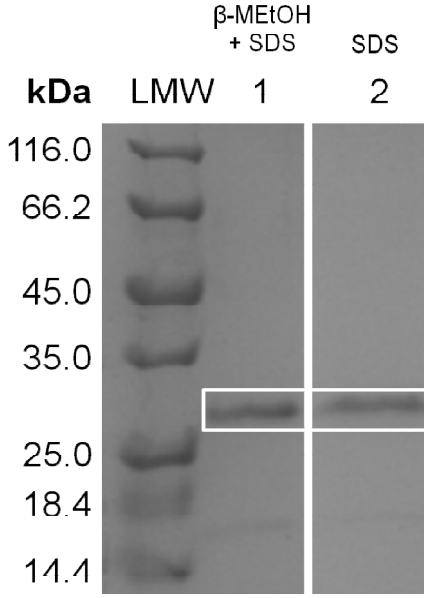


Figure 3.88 Analysis of intermolecular disulfide bond presence and oligomeric association in the CopS_C protein *via* SDS PAGE in reducing and non-reducing conditions. Legend: **LMW**: Low molecular weight marker; **Lane 1**: CopS_C sample in the presence of SDS and β-MEtOH; **Lane 2**: CopS_C sample in the presence of SDS. Black arrow identifies the presumed protein domain monomer of approximately 28 kDa. SDS-PAGE prepared in Tris-Glycine buffer (10%, 150 V, 1 h) and stained with Coomassie-blue.

B. Apparent molecular mass of CopS_C

Determination of the apparent molecular mass of the CopS_C protein in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl and further assessment of the formation of intermolecular disulfide bonds was performed in a pre-packed Superdex 75 10/300 GL, as described in Chapter 2 Materials and Methods – Section 2.3.1. The chromatogram obtained is shown in Figure 3.89,

together with the calibration curve, which was used for determination of the apparent molecular (insert Figure 3.89).

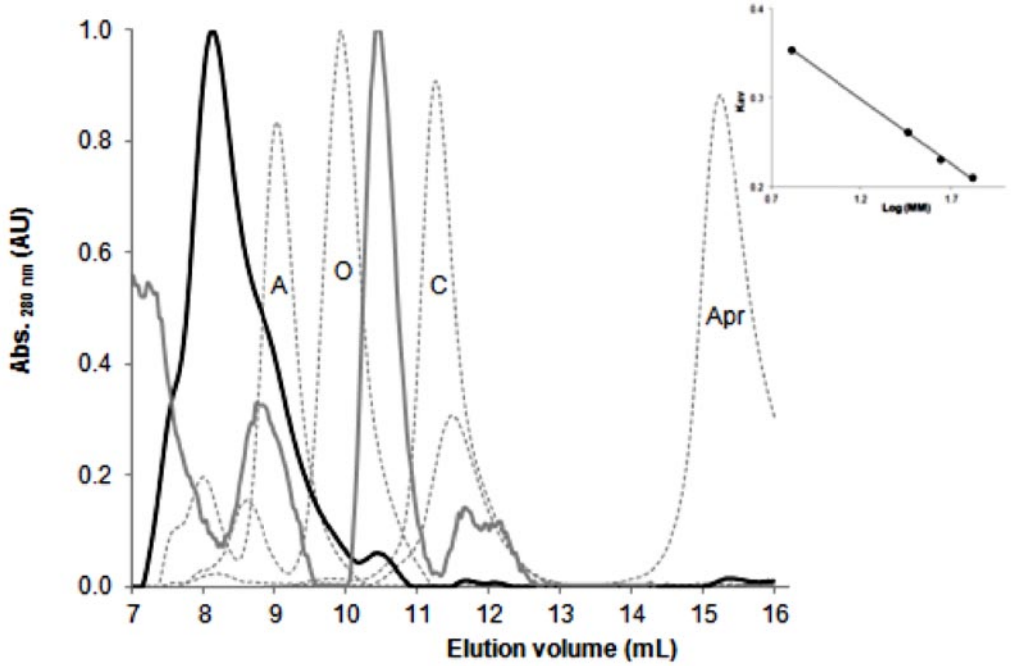


Figure 3.89 Elution profiles of calibration curve proteins (dotted curves), CopS_C2 (black line) and CopS_C (grey line) in a Superdex 75 10/300 GL. Proteins were eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl. Legend: **A**: Albumin, 66.0 kDa; **O**: Ovalbumin, 44.0 kDa; **C**: Carbonic Anhydrase, 29.0 kDa; **R**: Ribonuclease A, 13.7 kDa; **Apr**: Aprotinin, 6.5 kDa. Insert: Calibration curve for the molecular mass determination using LMW Gel Filtration Calibration Kit in Superdex 75 10/300 GL. Trend line equation obtained Kav = -0.14 log (MW) + 0.47, with an R of 1.00. The profiles correspond to individual injections. Elution profiles of CopS_CHis$_6$GST protein domain (black) and the CopS_C (grey) in a Superdex 75 10/300 GL at 280 nm.

The molecular masses of the CopS_C2 and CopS_C determined based on their primary sequence are 60.9 and 27.5 kDa respectively. In the elution profile of CopS_C (Figure 3.89, grey curve), there are four peaks. The main peak has an elution volume of approximately 10.5 mL, which corresponds to a protein with a molecular mass of 38.1 kDa, which should be CopS_C in a monomer dimer equilibrium (dimer 55 kDa and monomer 27.5 kDa). The other minor peaks observed in the chromatogram eluted as proteins of 20.4 kDa and 24.2 kDa, and a protein of molecular mass above 66.0 kDa, which might be small impurities, as can be observed in the SDS-PAGE in Figure 3.88.

In the case of CopS_C2, before tag cleavage, the elution profile shows some heterogeneity for the peak that elutes around 8.1 mL. In fact this protein elutes with a higher molecular mass than albumin (66 kDa), and may indicate that it is being eluted as a dimer, which could be confirmed by determining its apparent molecular mass using a Superdex 200 (which has a higher exclusion limit).

3.4.5.2 Circular Dichroism Spectroscopy

Given the high concentration of the CopS_C (2.2 mg/mL), circular dichroism (CD) spectroscopy can be used to determine with relative certainty the contribution of each of the secondary structure elements to the overall fold of the protein domain.

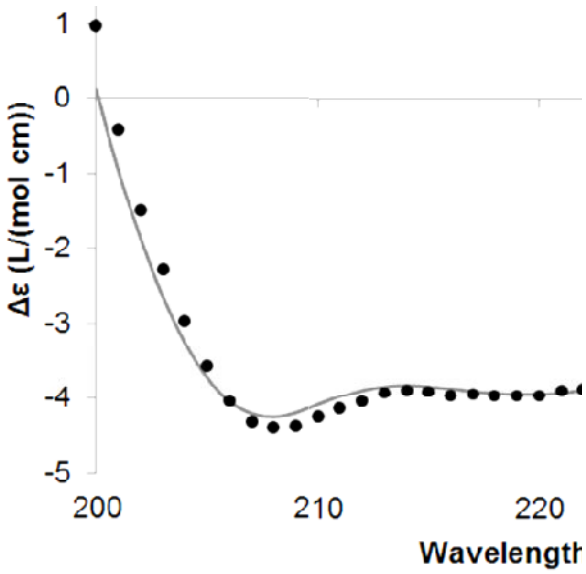


Figure 3.90 Theoretical (-----) and experimental (· · · · ·) CD spectra of the CopS_C protein in 20 mM phosphate buffer, pH 7.4.

CD spectra were collected from 200 to 240 nm and analyzed using a neural network analysis, K_2D_2, Figure 3.90. The obtained data is of good quality, and can be fitted with a theoretical curve obtained from the program. The composition of each structural element in this protein was determined as being: 40% α-helices, 10% β-sheets, and 50% random structure, Table 3.10, which agrees well with the data obtained from the predicted composition from Figure 3.71

Table 3.10 Secondary structure composition of the CopS_C protein determined by prediction servers and deconvolution of the experimental data obtained by CD.

	Prediction (%)	Deconvolution (%)
α-helix	47	40
β-sheet	20	10
Random coil	33	50

In spite of the inexistence of an approximate structural model, it is known, from previously determined structures, eg. EnvZ from *E.coli*, that the HK catalytic domain has an α/β fold consisting of five strands and three helices, which is a novel kinase fold unique to HKs.[25]

CHAPTER 4 – CONCLUSIONS AND FUTURE PERSPECTIVES

4. Conclusions and Future Perspectives

The system which was undertaken as the focus of this thesis is quite complex, as it involves two-components which interact not only between themselves but also with the environment and DNA. The first step in comprehending such a system has been achieved: The expression plasmids for the proteins, as well as their domains, have been constructed and a preliminary analysis of these proteins was performed.

4.1 CopR

CopR, the RR protein of this system, was expressed using expression plasmid pCopR2 which contained the gene of 666 bp cloned into the pET-21 c (+) vector. The resulting protein of 221 amino acids (25 kDa) and a N-terminal His_6-tag (6 kDa) was optimally expressed upon induction with 0.1 mM IPTG and growth for 16 h at 16°C, with an orbital rotation of 210 rpm. Of the buffers tested, those which are high in salt concentration and contain glycerol somewhat stabilize the protein. Therefore, CopR was maintained in 50 mM Tris-HCl (pH 7.6), 250 mM NaCl and 10% glycerol, at 4°C. In order to improve the stability of the protein, thermal denaturation screening of alternative buffer systems and additives is recommended, as precipitation continues an issue.

The adopted purification strategy for extraction of $CopRHis_6$ from the soluble cellular extract was adequate, as a reasonable amount of pure protein was recovered, between 3.7 and 6.4 mg/L of LB growth. His_6-tag cleavage, on the other hand, is time consuming, ~80 h at 4°C, and could be further optimized, by possibly augmenting the concentration of enzyme from the 0.4 mg/mL used in the assays. The overall yield of the purification from the soluble cellular extract in the form of $CopRHis_6$ to pure CopR sample (0.14 to 0.23 mg) is approximately 6%. This low yield is quite a hindrance to functional assays, as for example, EMSAs should be performed in different protein:*pro*DNA ratios, which in these conditions are not feasible.

Determination of the apparent molecular mass of the $CopRHis_6$ and CopR monomers (in a buffer containing DTT to a final concentration of 1 mM) was performed in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl. These domains were empirically determined by molecular exclusion chromatography to present molecular masses of 36 and 31 kDa vs. the theoretically determined molecular masses of 31 and 25 kDa. SDS-PAGE and PAGE assays in reducing and non-reducing conditions determined that the protein forms intermolecular bonds and as such should be maintained in a buffer containing DTT.

CD data confirmed the existence of some secondary structure, determined as 40% α-helix and 12% β-sheet. Conversely, due to the reduced concentration of the sample, this information is merely informative and not quantitative.

Once issues surrounding stability are resolved, the attempted assays, phosphorylation and EMSA, should be re-visited. The negative results which were obtained are not disheartening, in the sense that most of these assays are optimized by trial and error, and now some of the conditions which are not favourable to complex formation have been identified.

One avenue to pursue for phosphorylation would be to utilize another phosphodonor molecule instead of acetyl phosphate, *eg.* potassium phosphoramidate. If there is an alteration of the intrinsic fluorescence of the protein upon phosphorylation, stopped-flow fluorescence can be used to analyze the kinetics of the reaction and the effects of pH, ionic strength, *etc*. The EMSA still require optimization, as well as, the composition of the ligation buffer (data from gel filtration indicates the formation of the complex in the absence of BSA), which should promote interactions between the response regulator, CopR, and the DNA. The promoter sequence should also be extended to include other regions of the operon, mainly the fragment identified in Chapter 3 Results and Discussion – Section 3.2.3.5.

Again, once these conditions are optimized, DNase I footprinting assays could be used to determine the CopR binding sequence, and mutational assays could help determine the specificity. Surface plasmon resonance (SPR) could be used to monitor CopR binding from solution to surfaces coated with DNA, for example, or the CopR-CopS interaction, quantitatively, in real time and without labels. This technology is highly suitable for measurement of association and dissociation kinetics, and while the method has been traditionally limited to observations of only one binding reaction at a time, it can be adapted to arrays by using SPR imaging, which, in principle, would allow parallel kinetic measurements of RR binding to a multitude of different targets. Isothermal Titration Calorimetry (ITC) could, in conjunction with NMR, be used to determine the thermodynamic parameters of this interaction in solution. NMR is also valuable in determining the three-dimensional structure of the inactive, phosphorylated and active states so that one can understand the mechanistic aspects of the RR:DNA interaction.

To evaluate the specificity of the system response to copper and to analyze the increase of *copA, copX, copR* and *copS* transcript levels under copper ion induced stress RT-PCR could be used. This technique would also serve to elucidate the influence of other heavy ions, such as cadmium, cobalt, zinc and silver, on the system. In addition, reverse transcription can be used to confirm or deny that *copXAB* are transcribed as an operon, and elucidate the hypothesis that the *copXAB* and *copSR* operon are truly under CopR regulation. There is the possibility that, as in the previous work basal expression of CopR was observed, the *copRS* promoter may be constitutive like the *Pseudomonas copRS* and the *pcoRS* promoter[131].

4.2 CopR_N and CopR_C

Expression plasmids pCopR_N2 and pCopR_C2 encode the genes for the fusion proteins CopR_N, 360 bp, and CopR_C, 297 bp, which are expressed with a His_6-tag on the C-terminus. The individual domains, of 120 and 99 amino acids, were expressed after induction with 0.1 mM IPTG, orbital rotation of 210 rpm and growth for 3 h at 16°C. Though the growth conditions were sufficient to produce a small quantity of CopR_NHis_6 and CopR_CHis_6 in a soluble form, the majority of the expressed protein was insoluble/unstable.

The fusion proteins were obtained and purified by Ni^{2+}-sepharose affinity chromatography. The yield of each domain obtained upon purification from the soluble cellular extract was 0.42 mg for CopR_NHis_6 and 0.16 mg for CopR_CHis_6 per liter of LB medium. If upon optimization a larger fraction of the expressed protein is present in the soluble cellular extract, the losses through purification, or loss by precipitation, would be less damaging to the ability to perform certain assays.

The stability of the individual domains was examined and, once more, the rate of precipitation was quite high, although CopR_CHis_6 was more stable than CopR_NHis_6 and both were more soluble than CopR. The buffers in which the proteins were more stable were, 20 mM Tris-HCl (pH 7.6), 250 mM NaCl, 10% glycerol, 1 mM DTT for CopR_NHis_6 and 20 mM Tris-HCl, pH 7.6 for CopR_CHis_6. These buffers were used after protein purification from the soluble cellular extract. As a measure for the stability of the proteins in these buffers, the samples were maintained at -20°C for ~ 3 months, after which the amount of soluble CopR_NHis_6 was reduced by 76% (0.42 mg in the beginning and 0.10 mg at the end) and in the case of CopR_CHis_6 the amount of soluble protein was less than 0.06 mg, corresponding to a reduction of more than 49%. Once more, thermal denaturation screening of alternative buffers and additives is necessary, as precipitation is evidently still an issue.

SDS-PAGE analysis in reducing and non-reducing conditions confirmed that the CopR_CHis_6 domain is expressed and obtained from purification as a monomer, but that the CopR_NHis_6 domain is obtained, upon expression and purification, mainly as a monomer but forming oligomers which are formed *via* disulfide bonds. The apparent molecular masses of the CopR_NHis_6 and CopR_CHis_6 monomers, as determined by molecular exclusion chromatography in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl, were 17.5 and 15.1 kDa *vs.* the theoretically determined molecular masses of 14.5 and 12.6 kDa.

The reason for expressing these domains separately was to contour the issues surrounding CopR full-length protein stability. Due to the modularity of the CopR protein functional assays involving the independently obtained domains can be performed and the information obtained should, in principle, be transferable to the full-length protein.

So, all previously stated assays for phosphorylation of the receiver domain by small molecule phosphodonors or the histidine kinase can also be performed for the CopR_NHis$_6$ domain. Also, site-directed mutagenesis in the CopR_N protein domain could serve to study the specificity between CopS and CopR in contrast to the general features which allow phosphorylation of the RR. EMSAs and DNA I footprinting assays in which the hypothesis that the CopR protein might bind as a monomer can be examined using the CopR_C assay. The technique which would most benefit from this strategy is NMR which, given the smaller size of each of the domains, would be easier to interpret, though the low yield of protein might preclude this approach and X-ray crystallography might be of choice.

4.3 CopS_C

The full-length CopS protein is the HK component of this system, of which the cytoplasmatic region of the protein, CopS_C, the kinase domain, was expressed in this work. This domain was obtained using the pCopS_C2 plasmid in which the 756 bp fragment containing the gene for *copS_C* was cloned using the LIC cloning methodology into the pET-41 EK/LIC vector.

This plasmid was expressed in cells grown in LB medium upon 0.1 mM IPTG induction, 210 rpm orbital rotation for 3 h at 37°C. Of all the expressed proteins in this work, the CopS_C domain appears to be the most stable, with precipitation being largely nonexistent when maintained in 50 mM Tris-HCl (pH 7.6).

Purification of the CopS_C domain from the soluble cellular extract was performed by GST-tag affinity, from which 143.87 μM of protein was obtained. Tag cleavage of this protein was also far more efficient than that which occurred for the CopRHis$_6$ protein. In this case the protein is completely cleaved after 24 h at 4°C with 0.4 mg/mL of the protease. The overall yield of protein obtained after subsequent steps was 4.2 mg per liter of LB, an efficiency of 60%.

The CopS_C apparent molecular mass in 50 mM Tris-HCl (pH 7.6), 150 mM NaCl was determined by molecular exclusion chromatography as 38.2 kDa, whereas the theoretical molecular mass was determined as 28.5 kDa. It also appears that the protein may form a dimer, an observation which should be analyzed further as HKs auto-phosphorylate to form dimers *in vivo*. Analysis of SDS-PAGE in reducing and non-reducing conditions however showed that the protein was expressed and purified as a monomer.

CD data retrieved from the measurement of the CopS_C domain revealed that the protein retained some secondary structure (40% α-helix and 10% β-sheet).

All conditions are thereby reunited for the domain tertiary structure to be determined by NMR, also solid state NMR methods are available for studying membrane proteins immobilized in lipid bilayers and other membrane mimics. The position of the second transmembrane helix of

chemoreceptors relative to the membrane is a crucial factor in transmembrane signaling. Moving the positions of key residues can change the relative orientation of the helix and thereby alter the signaling state of the receptor in a predictable manner. However, this mutational approach would need to be optimized and determined if it can be applied to manipulate the signaling state of sensor kinases. Autophosphorylation assays of the CopS_C domain and consequent dimerization would be interesting to monitor, as well as the kinetics of CopS and CopR interaction by the detergent/lipid vesicle approach, which was applied to analyze the activity of all of the membrane sensor kinases from *Entercoccus faecalis* [140]. The FRET method can also be applied *in vitro* by analyzing purified fusion proteins, as shown for the OmpR/ PhoB subfamily and for the interaction between the HK EnvZ and RR OmpR within spheroblasts.

CHAPTER 5 – REFERENCES

1. Osman, D. & Cavet, J.S. Copper Homeostasis in Bacteria. *Advances in Applied Microbiology* **65**, 217-47 (2008).
2. Barton, L.L. *et al.* The Bacterial Metallome: Composition and Stability with Specific Reference to the Anaerobic Bacterium Desulfovibrio desulfuricans. *Biometals* **20**, 291-302 (2007).
3. Outten, C.E. & O'Halloran, T.V. Femtomolar Sensitivity of Metalloregulatory Proteins Controlling Zinc Homeostasis. *Science* **292**, 2488-92 (2001).
4. Cánovas, D., Cases, I. & Lorenzo, V. Heavy Metal Tolerance and Metal Homeostasis in Pseudomonas putida as Revealed by Complete Genome Analysis. *Environmental Microbiology* **5**, 1242-56 (2003).
5. Brown, N.L. & Lee, B.T.O. Bacterial Transport of and Resistance to Copper. *Metal Ions in Biological Systems* **30**, 405-34 (1994).
6. Boal, A.K. & Rosenzweig, A.C. Structural Biology of Copper Trafficking. *Chemical Reviews* **109**, 4760-79 (2009).
7. Stern, B.R. *et al.* Copper and Human Health: Biochemistry, Genetics, and Strategies for Modelling Dose-Response Relationships. *Journal of Toxicology and Environmental Health. Part B* **10**, 157-222 (2007).
8. Decker, H. & Terwilliger, N. Cops and Robbers: Putative Evolution of Copper Oxygen-Binding Proteins. *Journal of Experimental Biology* **203**, 1777-82 (2000).
9. Fraústo da Silva, J.J.R. & Williams, R.J.P. The Inorganic Chemistry of Life. *The Biological Chemistry of the Elements* (2001).
10. Xue, Y. *et al.* Cu (I) Recognition Via Cation-pi and Methionine Interactions in CusF. *Nature Chemical Biology* **4**, 107-9 (2007).
11. Crichton, R.R. & Pierre, J.-L. Old Iron, Young Copper: from Mars to Venus. *BioMetals* **14**, 99-112 (2001).
12. Arredondo, M. & Núñez, M.T. Iron and Copper Metabolism. *Molecular Aspects of Medicine* **26**, 313-27 (2005).
13. Mercer, J.F.B. The Molecular Basis of Copper-transport Diseases. *Trends in Molecular Medicine* **7**, 64-9 (2001).
14. Outten, F.W., Huffman, D.L., Hale, J.A. & O'Halloran, T.V. The Independent cue and cus Systems Confer Copper Tolerance during Aerobic and Anaerobic Growth in Escherichia coli. *The Journal of Biological Chemistry* **276**, 30670-7 (2001).
15. Tottey, S., Harvie, D.R. & Robinson, N.J. Understanding How Cells Allocate Metals using Metal Sensors and Metallochaperones. *Accounts of Chemical Research* **38**, 775-83 (2005).
16. Barakat, M. *et al.* P2CS: A Two-Component System Resource for Prokaryotic Signal Transduction Research. *BMC Genomics* **10**, 315-25 (2009).

17. West, A.H. & Stock, A.M. Histidine Kinases and Response Regulator Proteins in Two-Component Signaling Systems. *TRENDS in Biochemical Sciences* **26**, 369-76 (2001).
18. Attwood, P.V., Piggott, M.J., Zu, X.L., Besant, P.G. & G., B.P. Focus on Phosphohistidine. *Amino Acids* **32**, 145-56 (2007).
19. Gao, R. & Stock, A.M. Biological Insights from Structures of Two-Component Proteins. *Annual Review of Microbiology* **63**, 133-54 (2009).
20. Hoch, J.A. & Varughese, K.I. Keeping Signals Straight in Phosphorelay Signal Transduction. *Journal of Bacteriology* **183**, 4941-4949 (2001).
21. Barakat, M., Ortet, P. & Whitworth, D.E. P2CS: A Database of Prokaryotic Two-Component Systems. *Nucleic Acids Research* **39**, D771-6 (2011).
22. Finn, R.D. *et al.* The Pfam Protein Families Database. *Nucleic Acids Research* **38**, D211-22 (2010).
23. Stock, A.M., Robinson, V.L. & Goudreau, P.N. Two-Component Signal Transduction. *Annual Review of Biochemistry* **69**, 183-215 (2000).
24. Cheung, J. & Hendrickson, W.A. Sensor Domains of Two-Component Regulatory Systems. *Current Opinion in Microbiology* **13**, 116-23 (2010).
25. Robinson, V.L., Buckler, D.R. & Stock, A.M. A Tale of Two Components: A Novel Kinase and a Regulatory Switch. *Nature Structural Biology* **7**, 626-33 (2000).
26. Nies, D.H. Microbial Heavy-metal Resistance. *Applied Microbiology and Biotechnology* **51**, 730-50 (1999).
27. Lawrance, G.A. *Introduction to Coordination Chemistry*. *Reading* (John Wiley and Sons, Ltd.: 2010).
28. Cheung, J., Le-Khac, M. & Hendrickson, W.A. Crystal Structure of a Histidine Kinase Sensor Domain with Similarity to Periplasmic Binding Proteins. *Proteins* **77**, 235-41 (2009).
29. Cheung, J., Bingman, C.A., Reyngold, M., Hendrickson, W.A. & Waldburger, C.D. Crystal Structure of a Functional Dimer of the PhoQ Sensor Domain. *The Journal of Biological Chemistry* **283**, 13762-70 (2008).
30. Egger, L.A., Park, H. & Inouye, M. Signal Transduction Via the Histidyl-Aspartyl Phosphorelay. *Genes to Cells* **2**, 167-84 (1997).
31. Stock, J. Signal Transduction: Gyrating Protein Kinases. *Current Biology* **9**, R364-7 (1999).
32. Parkinson, J.S. & Kofoid, E.C. Communication Modules in Bacterial Signaling Proteins. *Annual Review of Genetics* **26**, 71–112 (1992).
33. Stock, J.B., Ninfa, A.J. & Stock, A.M. Protein Phosphorylation and Regulation of Adaptive Responses in Bacteria. *Microbiological Reviews* **53**, 450-90 (1989).
34. Hakenbeck, R. & Stock, J.B. Analysis of Two-component Signal Transduction Systems Involved in Transcriptional Regulation. *Methods in Enzymology* **273**, 281-300 (1996).

35. Casino, P., Rubio, V. & Marina, A. Structural Insight into Partner Specificity and Phosphoryl Transfer in Two-Component Signal Transduction. *Cell* **139**, 325-36 (2009).

36. Yang, Y. & Inouye, M. Intermolecular Complementation between Two Defective Mutant Signal-transducing Receptors of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **88**, 11057-61 (1991).

37. Laub, M.T. & Goulian, M. Specificity in Two-Component Signal Transduction Pathways. *Annual Review of Genetics* **41**, 121-45 (2007).

38. Alves, R. & Savageau, M.A. Comparative Analysis of Prototype Two-Component Systems With Either Bifunctional or Monofunctional Sensors: Differences in Molecular Structure and Physiological Function. *Molecular Microbiology* **48**, 25-51 (2003).

39. Letunic, I., Doerks, T. & Bork, P. SMART 6: Recent Updates and New Developments. *Nucleic Acids Research* **37**, D229-32 (2009).

40. Finn, R.D. *et al.* The Pfam Protein Families Database. *Nucleic Acids Research* **38**, D211-22 (2010).

41. Gao, R., Mack, T.R. & Stock, A.M. Bacterial Response Regulators: Versatile Regulatory Strategies from Common Domains. *Trends in Biochemical Sciences* **32**, 225-34 (2007).

42. Bourret, R.B. Receiver Domain Structure and Function in Response Regulator Proteins. *Current Opinion in Microbiology* **13**, 142-9 (2010).

43. Feher, V.A. *et al.* 1H, 15N, and 13C Backbone Chemical Shift Assignments, Secondary Structure, and Magnesium-binding Characteristics of the Bacillus subtilis Response Regulator, SpoOF, Determined by Heteronuclear High-resolution NMR. *Protein Science* **4**, 1801-14 (1995).

44. McCleary, W.R. & Stock, J.B. Acetyl Phosphate and the Activation of Two-Component Response Regulators. *The Journal of Biological Chemistry* **269**, 31567-72 (1994).

45. Lukat, G.S., McCleary, W.R., Stock, A.M. & Stock, J.B. Phosphorylation of Bacterial Response Regulator Proteins by Low Molecular Weight Phospho-donors. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 718-22 (1992).

46. Mayover, T.L., Halkides, C.J. & Stewart, R.C. Kinetic Characterization of CheY Phosphorylation Reactions: Comparison of P-CheA and Small-Molecule Phosphodonors. *Biochemistry* **38**, 2259-71 (1999).

47. Zapf, J.W., Hoch, J. a & Whiteley, J.M. A Phosphotransferase Activity of the Bacillus subtilis Sporulation Protein Spo0F that Employs Phosphoramidate Substrates. *Biochemistry* **35**, 2926-33 (1996).

48. Igo, M.M., Ninfa, A.J. & Silhavy, T.J. A Bacterial Environmental Sensor that Functions as a Protein Kinase and Stimulates Transcriptional Activation. *Genes & Development* **3**, 598-605 (1989).

49. Weiss, V. & Magasanik, B. Phosphorylation of Nitrogen Regulator I (NRI) of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 5088-92 (1992).
50. Wemmer, D.E. & Kern, D. Beryllofluoride Binding Mimics Phosphorylation of Aspartate in Response Regulators. *Journal of Bacteriology* **187**, 8229-8230 (2005).
51. Yan, D. *et al.* Beryllofluoride Mimics Phosphorylation of NtrC and Other Bacterial Response Regulators. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 14789-94 (1999).
52. Cho, H.S. *et al.* NMR Structure of Activated CheY. *Journal of Molecular Biology* **297**, 543-51 (2000).
53. Lee, S.Y. *et al.* Crystal Structure of an Activated Response Regulator Bound to its Target. *Nature Structural Biology* **8**, 52–6 (2001).
54. Varughese, K.I. Rebuttal: Beryllofluoride Binding Mimics Phosphorylation of Aspartate in Response Regulators. *Journal of Bacteriology* **187**, 8231 (2005).
55. Halkides, C.J. *et al.* The 1.9 A Resolution Crystal Structure of Phosphono-CheY, an Analogue of the Active Form of the Response Regulator, CheY. *Biochemistry* **39**, 5280-6 (2000).
56. Kern, D. *et al.* Structure of a Transiently Phosphorylated Switch in Bacterial Signal Transduction. *Nature* **402**, 894-8 (1999).
57. Birck, C. *et al.* Conformational Changes Induced by Phosphorylation of the FixJ Receiver Domain. *Structure* **7**, 1505-15 (1999).
58. Lewis, R.J., Brannigan, J.A., Muchova, K., Barak, I. & Wilkinson, A.J. Phosphorylated Aspartate in the Structure of a Response Regulator Protein. *Structure* 9-15 (1999).
59. Robinson, V.L., Wu, T. & Stock, A.M. Structural Analysis of the Domain Interface in DrrB , a Response Regulator of the OmpR / PhoB Subfamily. *Journal of Bacteriology* **185**, 4186-94 (2003).
60. King-Scott, J. *et al.* The Structure of a Full-Length Response Regulator from Mycobacterium tuberculosis in a Stabilized Three-Dimensional Domain-Swapped, Activated State. *The Journal of Biological Chemistry* **282**, 37717-29 (2007).
61. Martínez-Hackert, E. & Stock, A.M. Structural Relationships in the OmpR Family of Winged-Helix Transcription Factors. *Journal of Molecular Biology* **269**, 301-12 (1997).
62. Pelton, J.G., Kustu, S. & Wemmer, D.E. Solution Structure of the DNA-Binding Domain of NtrC with Three Alanine Substitutions. *Journal of Molecular Biology* **292**, 1095-110 (1999).
63. Baikalov, I. *et al.* NarL Dimerization? Suggestive Evidence from a New Crystal Form. *Biochemistry* **37**, 3665-76 (1998).
64. Martínez-Hackert, E. & Stock, A.M. Structural Relationships in the OmpR Family of Winged-Helix Transcription Factors. *Journal of Molecular Biology* **269**, 301-12 (1997).

65. Martínez-hackert, E. & Stock, A.M. The DNA-Binding Domain of OmpR: Crystal Structures of a Winged Helix Transcription Factor. *Structure* **5**, 109-24 (1997).

66. Buckler, D.R., Zhou, Y. & Stock, A.M. Evidence of Intradomain and Interdomain Flexibility in an OmpR/PhoB Homolog from Thermotoga maritima. *Structure* **10**, 153-64 (2002).

67. Brennan, R.G. The Winged-Helix DNA-Binding Motif: Another Helix-Turn-Helix Takeoff. *Cell* **74**, 773-6 (1993).

68. Djordjevic, S., Goudreau, P.N., Xu, Q., Stock, A.M. & West, A.H. Structural Basis for Methylesterase CheB Regulation by a Phosphorylation-Activated Domain. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 1381-6 (1998).

69. Nikolskaya, A.N. & Galperin, M.Y. A Novel Type of Conserved DNA-Binding Domain in the Transcriptional Regulators of the AlgR/AgrA/LytR Family. *Nucleic Acids Research* **30**, 2453-9 (2002).

70. Laguri, C., Phillips-Jones, M.K. & Williamson, M.P. Solution Structure and DNA Binding of the Effector Domain from the Global Regulator PrrA (RegA) from Rhodobacter sphaeroides: Insights into DNA Binding Specificity. *Nucleic Acids Research* **31**, 6778-87 (2003).

71. Stock, A.M. & Guhaniyogi, J. A New Perspective on Response Regulator Activation. *Journal of Bacteriology* **188**, 7328-30 (2006).

72. Rhee, J.E. *et al.* Amino Acids Important for DNA Recognition by the Response Regulator OmpR. *The Journal of Biological Chemistry* **283**, 8664-77 (2008).

73. Harrison-McMonagle, P., Denissova, N., Martínez-Hackert, E., Ebright, R.H. & Stock, A.M. Orientation of OmpR Monomers within an OmpR:DNA Complex Determined by DNA Affinity Cleaving. *Journal of Molecular Biology* **285**, 555-66 (1999).

74. Mattison, K. & Kenney, L.J. Phosphorylation Alters the Interaction of the Response Regulator OmpR with its Sensor Kinase EnvZ. *The Journal of Biological Chemistry* **277**, 11143-8 (2002).

75. Harrison-McMonagle, P., Denissova, N., Martínez-Hackert, E., Ebright, R.H. & Stock, A.M. Orientation of OmpR Monomers within an OmpR: DNA Complex Determined by DNA Affinity Cleaving. *Journal of Molecular Biology* **285**, 555-66 (1999).

76. Latchman, D.S. *Transcription factors: a practical approach*. *Molecular Biology* **201**, (Oxford University Press, USA: 1999).

77. Outten, F.W., Outten, C.E., Hale, J. & O'Halloran, T.V. Transcriptional Activation of an Escherichia coli Copper Efflux Regulon by the Chromosomal MerR Homologue, CueR. *The Journal of Biological Chemistry* **275**, 31024 -9 (2000).

78. Stoyanov, J.V., Hobman, J.L. & Brown, N.L. CueR (YbbI) of Escherichia coli is a MerR family regulator controlling expression of the copper exporter CopA. *Molecular Microbiology* **39**, 502-11 (2001).

79. Gaballa, A., Cao, M. & Helmann, J.D. Two MerR homologues that affect copper induction of the Bacillus subtilis copZA operon. *Microbiology* **149**, 3413-21 (2003).

80. Julian, D.J., Kershaw, C.J., Brown, N.L. & Hobman, J.L. Transcriptional activation of MerR family promoters in Cupriavidus metallidurans CH34. *Antonie van Leeuwenhoek* **96**, 149-59 (2009).

81. Thaden, J.T., Lory, S. & Gardner, T.S. Quorum-Sensing Regulation of a Copper Toxicity System in Pseudomonas aeruginosa. *Journal of Bacteriology* **192**, 2557-68 (2010).

82. Reeve, W.G., Tiwari, R.P., Kale, N.B., Dilworth, M.J. & Glenn, A.R. ActP controls copper homeostasis in Rhizobium leguminosarum bv. viciae and Sinorhizobium meliloti preventing low pH-induced copper toxicity. *Molecular Microbiology* **43**, 981-91 (2002).

83. Magnani, D., Barré, O., Gerber, S.D. & Solioz, M. Characterization of the CopR regulon of Lactococcus lactis IL1403. *Journal of bacteriology* **190**, 536-45 (2008).

84. Portmann, R. *et al.* Interaction kinetics of the copper-responsive CopY repressor with the cop promoter of Enterococcus hirae. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry* **9**, 396-402 (2004).

85. Magnani, D. & Solioz, M. How Bacteria Handle Copper. *Bacterial Transition Metal Homeostasis* 259–85 (2007).at <http://www.springerlink.com/index/hu258j72ph85212w.pdf>

86. Ward, S.K., Hoye, E. a & Talaat, A.M. The global responses of Mycobacterium tuberculosis to physiological levels of copper. *Journal of bacteriology* **190**, 2939-46 (2008).

87. Schelder, S., Zaade, D., Litsanov, B., Bott, M. & Brocker, M. The two-component signal transduction system CopRS of Corynebacterium glutamicum is required for adaptation to copper-excess stress. *PloS one* **6**, e22143 (2011).

88. Rouch, D. a & Brown, N.L. Copper-inducible transcriptional regulation at two promoters in the Escherichia coli copper resistance determinant pco. *Microbiology (Reading, England)* **143 (Pt 4**, 1191-202 (1997).

89. Munson, G.P., Lam, D.L., Outten, F.W. & Halloran, T.V.O. Identification of a Copper-Responsive Two-Component System on the Chromosome of Escherichia coli K-12 Identification of a Copper-Responsive Two-Component System on the Chromosome of Escherichia coli K-12. *Society* (2000).doi:10.1128/JB.182.20.5864-5871.2000.Updated

90. Mills, S.D., Jasalavich, C. a & Cooksey, D. a A two-component regulatory system required for copper-inducible expression of the copper resistance operon of Pseudomonas syringae. *Journal of bacteriology* **175**, 1656-64 (1993).

91. Márquez, M.C. & Ventosa, A. Marinobacter hydrocarbonoclasticus Gauthier et al. 1992 and Marinobacter aquaeolei Nguyen et al. 1999 are Heterotypic Synonyms. *International Journal of Systematic and Evolutionary Microbiology* **55**, 1349-51 (2005).

92. Huu, N.B., Denner, E., Ha, D.T.C., Wanner, G. & Stan-Lotter, H. Marinobacter aquaeolei sp. nov., a halophilic bacterium isolated from a Vietnamese oil-producing well. *International Journal of Systematic and Evolutionary Microbiology* **49**, 367-75 (1999).

93. Kenney, L.J., Bauer, M.D. & Silhavy, T.J. Phosphorylation-Dependent Conformational Changes in OmpR, an Osmoregulatory DNA-Binding Protein of Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 8866-70 (1995).

94. Gasteiger, E. *et al.* ExPASy: The Proteomics Server for in-depth Protein Knowledge and Analysis. *Nucleic Acids Research* **31**, 3784-8 (2003).

95. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-10 (1990).

96. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947-8 (2007).

97. Yu, N.Y. *et al.* PSORTb 3.0: Improved Protein Subcellular Localization Prediction with Refined Localization Subcategories and Predictive Capabilities for all Prokaryotes. *Bioinformatics* **26**, 1608-15 (2010).

98. Yu, C.S., Chen, Y.C., Lu, C.-H. & Hwang, J.K. Prediction of Protein Subcellular Localization. *Proteins: Structure, Function, and Bioinformatics* **64**, 643–51 (2006).

99. Shen, H.-B. & Chou, K.-C. Gneg-mPLoc: A top-down Strategy to Enhance the Quality of Predicting Subcellular Localization of Gram-negative Bacterial Proteins. *Journal of Theoretical Biology* **264**, 326-33 (2010).

100. Marchler-Bauer, A. *et al.* CDD: a Conserved Domain Database for the Functional Annotation of Proteins. *Nucleic Acids Research* **39**, D225-9 (2011).

101. Pollastri, G. & McLysaght, A. Porter: A New, Accurate Server for Protein Secondary Structure Prediction. *Structural Bioinformatics* **21**, 1719-20 (2005).

102. Jones, D.T. Protein Secondary Structure Prediction Based on Position-specific Scoring Matrices. *Journal of Molecular Biology* **292**, 195-202 (1999).

103. Petersen, B., Petersen, T.N., Andersen, P., Nielsen, M. & Lundegaard, C. A Generic Method for Assignment of Reliability Scores Applied to Solvent Accessibility Predictions. *BMC Structural Biology* **9**, 51-61 (2009).

104. Zhang Lab Protein Secondary Structure PREDiction - PSSPRED. at <http://zhanglab.ccmb.med.umich.edu/PSSpred >

105. Karypis, G. YASSPP: Better Kernels and Coding Schemes Lead to Improvements in Protein Secondary Structure Prediction. *Proteins: Structure, Function, and Bioinformatics* **586**, 575-86 (2006).

106. Cole, C., Barber, J.D. & Barton, G.J. The Jpred 3 Secondary Structure Prediction Server. *Nucleic Acids Research* **36**, W197-201 (2008).

107. Kiefer, F., Arnold, K., Künzli, M., Bordoli, L. & Schwede, T. The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research* **37**, D387-92 (2009).

108. Pettersen, E.F. *et al.* UCSF Chimera - A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* **25**, 1605-12 (2004).

109. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *Journal of Applied Crystallography* **26**, 283-91 (1993).

110. Wiederstein, M. & Sippl, M.J. ProSA-web: Interactive Web Service for the Recognition of Errors in Three-dimensional Structures of Proteins. *Nucleic Acids Research* **35**, W407-10 (2007).

111. Tortoli, E. *et al.* Performance Assessment of new Multiplex Probe Assay for Identification of Mycobacteria. *Journal of Clinical Microbiology* **39**, 1079-84 (2001).

112. Sambrook, J. & Russell, D.W. *Molecular Cloning: A Laboratory Manual*. **2**, (Cold Spring Harbor Laboratory Press: 2001).

113. Lowry, O.H., Rosebrough, N.J., Farr, A.L. & Randall, R.J. Protein Measurement with the Folin Phenol Reagent. *Journal of Biological Chemistry* **193**, 265–75 (1951).

114. Perez-Iratxeta, C. & Andrade-Navarro, M.A. K2D2: Estimation of Protein Secondary Structure from Circular Dichroism Spectra. *BMC Structural Biology* **8**, 25-30 (2008).

115. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research* **40**, 109-14 (2012).

116. Menon, S. & Wang, S. Structure of the Response Regulator PhoP from Mycobacterium tuberculosis Reveals a Dimer through the Receiver Domain. *Biochemistry* **50**, 5948-57 (2011).

117. Guex, N., Peitsch, M.C. & Schwede, T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. *Electrophoresis* **30**, S162–S173 (2009).

118. Bordoli, L. *et al.* Protein structure homology modeling using SWISS-MODEL workspace. *Nature Protocols* **4**, 1-13 (2009).

119. Friedland, N. *et al.* Domain Orientation in the Inactive Response Regulator Mycobacterium tuberculosis MtrA Provides a Barrier to Activation. *Biochemistry* **46**, 6733-43 (2007).

120. Kenney, L.J. Structure/Function Relationships in OmpR and other Winged-helix Transcription Factors. *Current Opinion in Microbiology* **5**, 135-41 (2002).

121. Healthcare, G.E. Recombinant Protein Purification Handbook.

122. Via, L.E. *et al.* Elements of Signal Transduction in Mycobacterium tuberculosis: In vitro Phosphorylation and In vivo Expression of the Response Regulator MtrA. *Journal of Bacteriology* **178**, 3314-21 (1996).

123. Da Re, S.S., Deville-Bonne, D., Tolstykh, T., V ron, M. & Stock, J.B. Kinetics of CheY Phosphorylation by Small Molecule Phosphodonors. *FEBS letters* **457**, 323-6 (1999).

124. Wolfe, A.J. Physiologically Relevant Small Phosphodonors Link Metabolism to Signal Transduction. *Current Opinion in Microbiology* **13**, 204-9 (2010).

125. Mccleary, W.R., Stock, J.B. & Ninfa, A.J. Is Acetyl Phosphate a Global Signal in Escherichia coli? *Journal of Bacteriology* **175**, 2793-8 (1993).

126. Gusa, A. a, Gao, J., Stringer, V., Churchward, G. & Scott, J.R. Phosphorylation of the Group A Streptococcal CovR Response Regulator Causes Dimerization and Promoter-specific Recruitment by RNA polymerase. *Journal of Bacteriology* **188**, 4620-6 (2006).

127. Abo-amer, A.E. *et al.* DNA Interaction and Phosphotransfer of the C4-Dicarboxylate-Responsive DcuS-DcuR Two-Component Regulatory System from Escherichia coli. *Journal of Bacteriology* **186**, 1879-89 (2004).

128. Miller, a a, Engleberg, N.C. & DiRita, V.J. Repression of Virulence Genes by Phosphorylation-dependent Oligomerization of CsrR at Target Promoters in S. pyogenes. *Molecular Microbiology* **40**, 976-90 (2001).

129. Zhang, J.H., Xiao, G., Gunsalus, R.P. & Hubbell, W.L. Phosphorylation Triggers Domain Separation in the DNA Binding Response Regulator NarL. *Biochemistry* **42**, 2552-9 (2003).

130. Lewin, B. *Genes VIII*. (Pearson Prentice Hall: 2004).

131. Laub, M.T., Biondi, E.G. & Skerker, J.M. Phosphotransfer Profiling: Systematic Mapping of Two-Component Signal Transduction Pathways and Phosphorelays. *Methods in Enzymology* **423**, 531–48 (2007).

132. Qin, L., Yoshida, T. & Inouye, M. The critical role of DNA in the equilibrium between OmpR and phosphorylated OmpR mediated by EnvZ in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 908-13 (2001).

133. Holden, N.S. & Tacon, C.E. Principles and problems of the electrophoretic mobility shift assay. *Journal of pharmacological and toxicological methods* **63**, 7-14 (2010).

134. Hughes, T. *A Handbook of Transcription Factors*. *Annals of Physics* **52**, (Springer Verlag: 2011).

135. Wang, S., Engohang-Ndong, J. & Smith, I. Structure of the DNA-binding Domain of the Response Regulator PhoP from Mycobacterium tuberculosis. *Biochemistry* **46**, 14751-61 (2007).

136. Trinh, C.-H., Liu, Y., Phillips, S.E.V. & Phillips-Jones, M.K. Structure of the Response Regulator VicR DNA-binding Domain. *Acta Crystallographica. Section D, Biological crystallography* **63**, 266-9 (2007).

137. McCleary, W.R. No Phobias about PhoB Activation. *Structure* **13**, 1238-9 (2005).

138. Browning, D.F. & Busby, S.J. The Regulation of Bacterial Transcription Initiation. *Nature Reviews. Microbiology* **2**, 57-65 (2004).

139. Rangwala, H., Kauffman, C. & Karypis, G. svmPRAT: SVM-based Protein Residue Annotation Toolkit. *BMC Bioinformatics* **12**, 1-12 (2009).

140. Ma, P. *et al.* Expression, Purification and Activities of the Entire Family of Intact Membrane Sensor Kinases from Enterococcus faecalis. *Molecular Membrane Biology* **25**, 449-73 (2008).

CHAPTER 6 – APPENDICES

6. APPENDIX

A. Solutions and Methods

A.1 Luria Broth Media

The culture medium Luria broth is constituted by 10g/L NaCl, 10g/L tryptone and 5 g/L yeast extract. When solid culture media is necessary 20 g/L of agar is added to the medium. The cultures were sterilized in an autoclave in a 20 min cycle, at 120°C and 1 bar. The antibiotic and IPTG solutions used were sterilized by flame filtration with a 0.2 μM pore filter (Millipore), which were then maintained at -20°C.

A.2 Silver Staining

Silver staining was used for EMSA gels in order to identify DNA and proteins. For fixation the gel is soaked for 120 min (or overnight) in fixing solution, thereupon this solution is discarded and sensitizing solution is added and left to react with the gel for an additional 30 min. The gel is then removed from this solution and distilled water is used to wash five times the gel for 15 min each time. The silver reaction is achieved by addition of the silver solution and an additional incubation for 20 min. Once more the gel is washed as before, and the gel is then immersed in the developing solution which is left shaking until the desired intensity of the bands is achieved. Once this point is reached the stopping solution is added. The gel may then be washed and transferred to the preserving solution for posterior drying and storage.

The composition of the solutions used throughout this protocol are listed in Table 6.1.

Table 6.1 Composition of the various solutions necessary for silver staining by the method described in the text.

Step	Component and quantity
Fixing solution	Ethanol 30% - 75 mL; Glacial acetic acid - 25 mL; Water to 250 mL
Sensitizing solution	Ethanol 75 mL; Sodium thiosulphate (5% w/v) - 10 mL;
Silver solution	Silver nitrate solution (2.5% w/v) - 25 mL; Water to 250 mL
Developing solution	Sodium carbonate - 6.25 g; Water to 250 mL. Stir vigorously to Before use: Add 0.2 mL formaldehyde (37% w/v)
Stop solution	EDTA-Na_2•$2H_2O$ (3.65 g); Water to 250 mL
Washing solution	Water
Preserving solution	Glycerol (87% w/w) 25 mL; Water to 250 mL

Note: Glutardialdehyde and formaldehyde should be added immediately before use, 250 mL solution is needed per gel and step and are as described in Table 6.1. All steps should be performed with gentle shaking of the staining tray.

A.3 Purification of the HRV3C protease

The expression vector containing the gene encoding the HRV3C protease, a GST-tag fusion protein, was a kind gift of Dr. Manolis Matzapetakis. The optimal over-expression conditions and purification were previously optimized.

When the cell culture OD_{600nm} > 0.6, it was induced with 0.1 mM IPTG and allowed to grow for an additional 1.5 h at 37°C with an orbital rotation of 210 rpm. Once the *E.coli* cells had been harvested, lysed and ultracentrifuged, the soluble cellular extract was loaded onto a GST-Trap FF column for purification.

Figure 6.1 shows a typical purification of the HRV3C protease. Typically the yield of pure HRV3C obtained per L of growth medium was 6.3 mg.

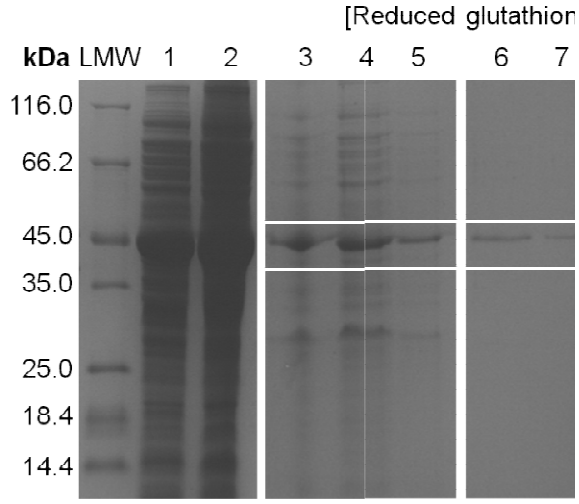


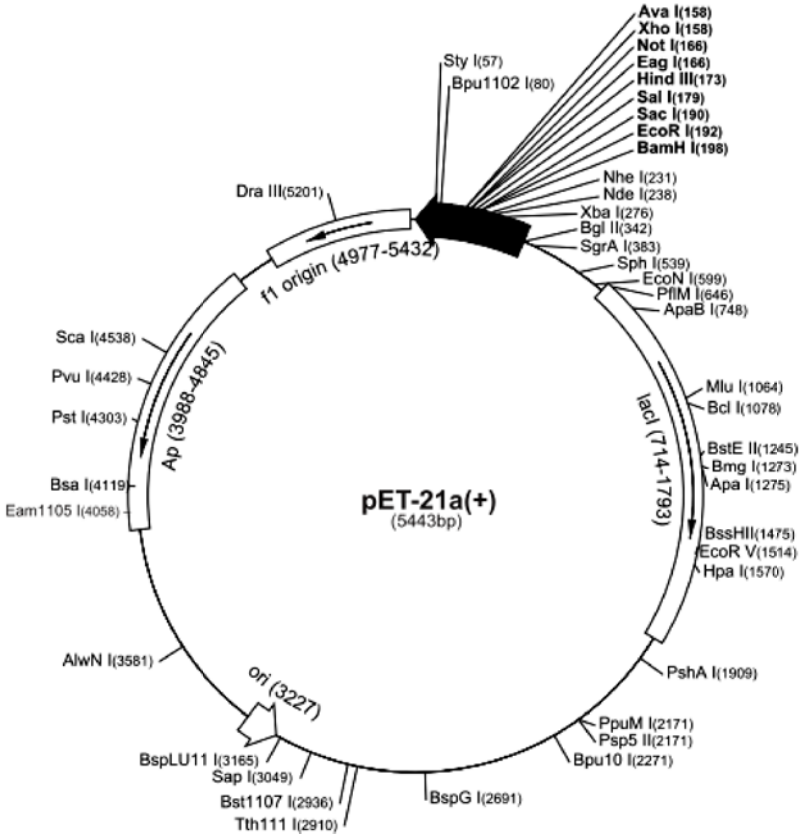
Figure 6.1 Purification of the HRV3C protease (54 kDa). Legend: **LMW**: Low molecular weight marker; **Lane1**: Injected sample; **Lane 2**: Sample of the flow-through; **Lanes 3 to 7**: 4, 12, 20, 28, 36 mL of 10 mM imidazole, 150 mM NaCl. **Lanes 8 to 11**: 2.5, 5, 7.5 and 10 mL of 50 mM Tris-HCl (pH 7.6), 5 mM glutathione; **Lanes 12 to 16**: 2.5, 5, 10, 15, 20 mL of 50 mM Tris-HCl (pH 7.6), 10 mM reduced glutathione; **Lanes 17 and 18**: 2.5 and 7.5 mL 50 mM Tris-HCl (7.6), 20 mM reduced glutathione. The arrow indicates the expected electrophoretic mobility of a protein of approximately 54 kDa, which is the MW of the HRV3C protease. SDS-PAGE prepared in Tris-Tricine buffer (12,5%, 150 V, 1 h) and stained with Coomassie-blue.

B. Vectors and Strains

The pET – 21 c (+), pET-30 EK/LIC and pET-41 EK/LIC expression plasmids were utilized throughout this work in the construction of various expression vectors. In all cases, the primers utilized for DNA sequencing are indicated in the cloning region of the sequences.

B.1 Vector pET-21 c (+)

A.



B.

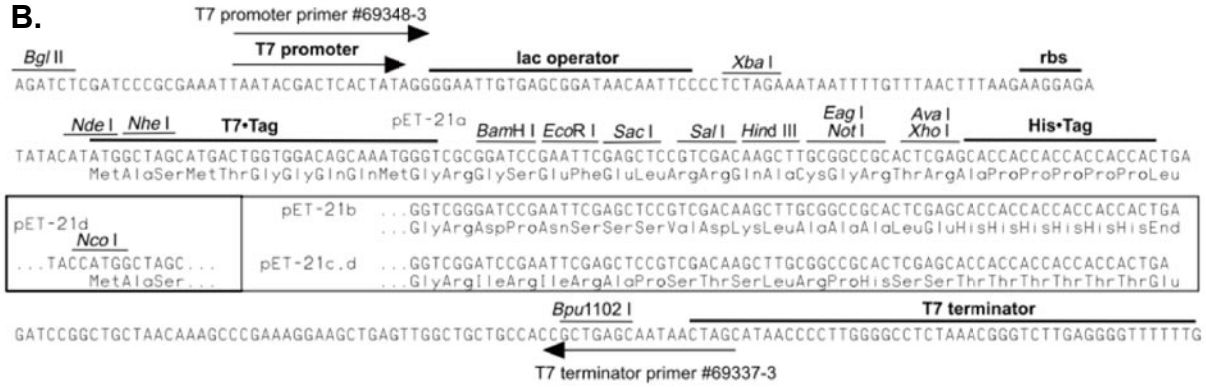
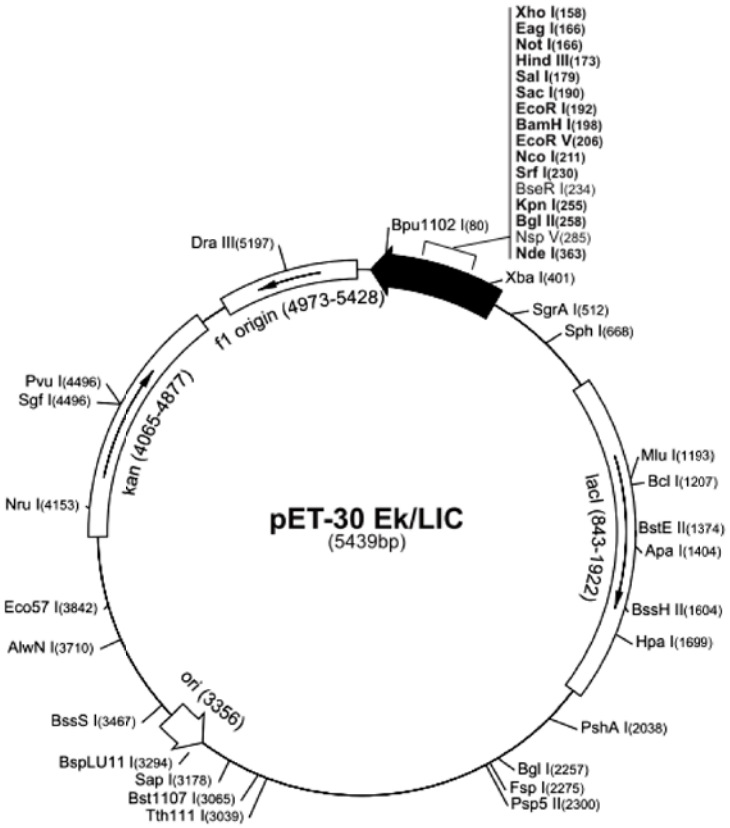


Figure 6.2 A. Genetic map of the pET-21 c (+) expression plasmid. **B.** Cloning/expression region on the coding strand transcribed by the T7 RNA polymerase. Images reproduced from the information available from Novagen on the pET-21 vector.

B.2 Vector pET30 EK/LIC

A.



B.

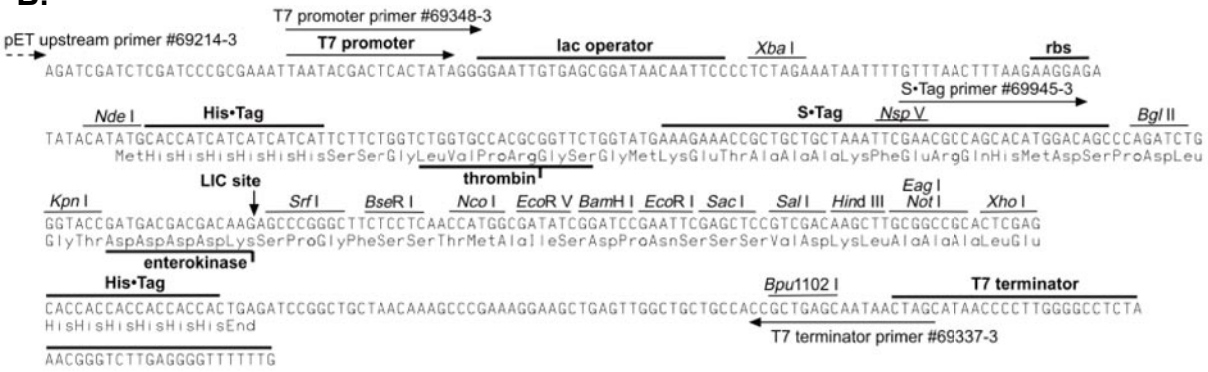
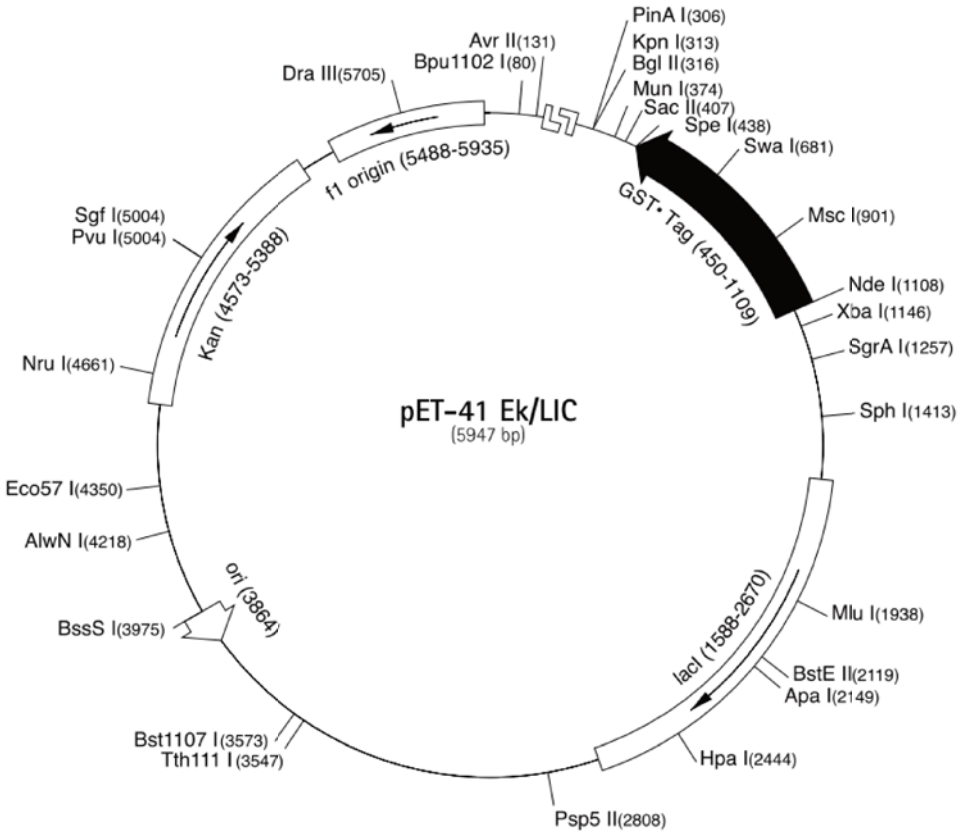


Figure 6.3 A. Genetic map of the pET-30 EK/LIC expression plasmid. **B.** Cloning/expression region on the coding strand transcribed by the T7 RNA polymerase. Images reproduced from the information available from Novagen on the pET-30 EK/LIC vector.

B.3 Vector pET41 EK/LIC

A.



B.

Figure 6.4 A. Genetic map of the pET-41 EK/LIC expression plasmid. **B.** Cloning/expression region on the coding strand transcribed by the T7 RNA polymerase. Images reproduced from the information available from Novagen on the pET-41 EK/LIC vector.

B.4 Bacterial Strains

For the homology alignments, as well as template screening, several proteins from a range of bacteria were used, their identifiers are presented in Table 6.2 in case verification is necessary.

Table 6.2 List of bacterial strains from which protein sequences, indicated by the RefSeq identifier or PDB ID were utilized for homology alignments or template screening.

Protein	Strain	RefSeq/PDB
CopR	*M.aquaeolei* VT8	YP_957418.1
CopS	*M.aquaeolei* VT8	YP_957417.1
-	*M. algicola* DG893	ZP_01892187.1
-	*Alcanivorax borkumensis* SK2	YP_693085.1
-	*Allochromatium vinosum* DSM 180	YP_003442015.1
-	*Methylophaga aminisulfidivorans* MP	ZP_08536950.1
-	*Pseudomonas stutzeri* A1501	YP_001171390.1
-	*Magnetococcus sp.* MC-1	YP_864245.1
-	*M. sp.* ELB17	ZP_01739820.1
-	*M.* sp. MnI7-9	ZP_09158578.1
-	*M. adhaerens* HP15	ADP95950.1
-	*M.* sp. ELB17	ZP_01740005.1
-	*M.* algicola DG893	ZP_01892188.1
-	*M.* sp. MnI7-9	ZP_09161362.1
-	*M.* sp. ELB17	ZP_01739819.1
-	*M.* sp. MnI7-9	ZP_09158294.1
-	*Alcanivorax borkumensis* SK2	YP_693086.1
-	*Alcanivorax* sp. DG881	ZP_05042513.1
DrrD	*Thermotoga maritima*	1KGS
DrrB	*Thermotoga maritima*	1P2F
PrrA	*Mycobacterium tuberculosis*	1YS6
RegX3	*Mycobacterium tuberculosis*	2OQR
MtrA	*Mycobacterium tuberculosis*	2GWR
PhoP	*Mycobacterium tuberculosis*	3R0J

C. Bioinformatic Analysis

C.1. *M. aquaeolei* VT8

Construction of Figures 3.1 A and B in Chapter 3 for overall analysis of two-component protein class distribution in *M.aquaeolei* VT8, was based on the data retrieved from the protein two-component system, P2CS[21], presented in Tables 6.2 and 6.3.

Table 6.3 List of all HK from *M. aquaeolei* VT8 with a brief description of the HK family and domains. Data obtained from P2CS.

Gene	P2CS description
Maqu_1972	HK, CheA contains 1 Hpt,1 H-kinase_dim,1 HATPase_c,1 CheW
Maqu_3772	HK, CheA contains 7 Hpt,1 H-kinase_dim,1 HATPase_c,1 CheW,1 REC
Maqu_0123	HK, Classic contains 1 HAMP, 1 HisKA_MA, 1 HATPase_c,
Maqu_0249	HK, Classic contains 2 PAS,1 HisKA,1 HATPase_c
Maqu_0285	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_0300	HK, Classic contains 1 HisKA_MA, 1 HATPase_c
Maqu_0489	HK, Classic contains 1 His_kinase,1 HATPase_c
Maqu_0500	HK, Classic contains 2 PBPb,1 HisKA,1 HATPase_c
Maqu_0671	HK, Classic contains 1 PAS,1 HisKA,1 HATPase_c
Maqu_0767	HK, Classic contains 1 PAS_4,1 HisKA,1 HATPase_c
Maqu_0875	HK, Classic contains 1 PAS,1 HisKA,1 HATPase_c
Maqu_0896	HK, Classic contains 1 HisKA_MA, 1 HATPase_c
Maqu_1008	HK, Classic contains 1 HisKA,1 HATPase_c
Maqu_1286	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_1618	HK, Classic contains 1 HisKA_MA, 1 HATPase_c
Maqu_2000	HK, Classic contains 1 HisKA,1 HATPase_c
Maqu_2094	HK, Classic contains 1 PAS_4,1 HisKA,1 HATPase_c
Maqu_2222	HK, Classic contains 1 HDOD,1 HisKA,1 HATPase_c
Maqu_2320	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_2348	HK, Classic contains 1 HisKA,1 HATPase_c
Maqu_2532	HK, Classic contains 1 CHASE2,1 PAS,1 HisKA,1 HATPase_c
Maqu_2827	HK, Classic contains 1 HisKA,1 HATPase_c
Maqu_2868	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_2898	HK, Classic contains 1 PAS, 1 HisKA_MA, 1 HATPase_c,
Maqu_3059	HK, Classic contains 1 2CSK_N, 1 HAMP, 1 HisKA_MA, 1 HATPase_c,
Maqu_3088	HK, Classic contains 1 HAMP,1 GAF,1 HisKA_3,1 HATPase_c
Maqu_3218	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_3533	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_3597	HK, Classic contains 1 HisKA,1 HATPase_c
Maqu_3671	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_3672	HK, Classic contains 2 GAF,1 PAS_3,1 HisKA,1 HATPase_c
Maqu_3673	HK, Classic contains 1 CHASE,1 HisKA,1 HATPase_c
Maqu_3686	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_3739	HK, Classic contains 1 HAMP,1 HisKA,1 HATPase_c
Maqu_0898	HK, Hybrid contains 1 PAS,1 HisKA,1 HATPase_c,1 Response_reg
Maqu_1360	HK, Hybrid contains 1 HAMP,1 HisKA,1 HATPase_c,1 REC
Maqu_1769	HK, Hybrid contains 1 HisKA,1 HATPase_c,1 REC
Maqu_2687	HK, Hybrid contains 1 HAMP,1 HisKA,1 HATPase_c,1 REC,1 Response_reg
Maqu_3868	HK, Hybrid contains 1 HisKA,1 HATPase_c,1 REC
Maqu_0100	HK, Unorthodox contains 1 PAS_3,1 HisKA,1 HATPase_c,2 REC,1 Hpt
Maqu_2241	HK, Unorthodox contains 1 HisKA,1 HATPase_c,1 Response_reg,1 Hpt
Maqu_3675	HK, Unorthodox contains 3 MHYT,1 PAS,1 HisKA,1 HATPase_c,1 REC,1 Hpt

Table 6.4 List of all RR from *M. aquaeolei* VT8 with a brief description of the RR family and domains. Data obtained from P2CS.

Gene	P2CS description
Maqu_1971	Response regulator, CheB family contains 1 REC,1 CheB_methyle
Maqu_1176	Response regulator, CheV family contains 1 CheW,1 REC
Maqu_3396	Response regulator, CheV family contains 1 CheW,1 REC
Maqu_0102	Response regulator, CheY family contains 1 REC
Maqu_0248	Response regulator, CheY family contains 1 REC
Maqu_1974	Response regulator, CheY family contains 1 REC
Maqu_3668	Response regulator, CheY family contains 1 Response_reg
Maqu_3767	Response regulator, CheY family contains 1 REC
Maqu_3768	Response regulator, CheY family contains 1 REC
Maqu_0488	Response regulator, LytTR family contains 1 REC,1 LytTR
Maqu_1191	Response regulator, NarL family contains 1 REC,1 GerE
Maqu_2095	Response regulator, NarL family contains 1 Response_reg,1 GerE
Maqu_2218	Response regulator, NarL family contains 1 Response_reg,1 HTH_LUXR
Maqu_3089	Response regulator, NarL family contains 1 REC,1 HTH_LUXR
Maqu_3869	Response regulator, NarL family contains 1 REC,1 HTH_LUXR
Maqu_0766	Response regulator, NtrC family contains 1 Response_reg,1 AAA_2,1 HTH_8
Maqu_0874	Response regulator, NtrC family contains 1 Response_reg,1 AAA,1 HTH_8
Maqu_1619	Response regulator, NtrC family contains 1 Response_reg,1 AAA_3,1 HTH_8
Maqu_1999	Response regulator, NtrC family contains 1 Response_reg,1 AAA_5,1 HTH_8
Maqu_2347	Response regulator, NtrC family contains 1 Response_reg,1 AAA_2,1 HTH_8
Maqu_2826	Response regulator, NtrC family contains 1 REC,1 AAA_3,1 HTH_8
Maqu_0124	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_0286	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_1007	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_1287	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_2321	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_2530	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_2869	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_3060	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_3217	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_3534	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_3596	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_3685	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_3740	Response regulator, OmpR family contains 1 REC,1 Trans_reg_C
Maqu_2586	Response regulator, PleD family contains 1 REC,1 GGDEF
Maqu_3442	Response regulator, PleD family contains 1 Response_reg,1 REC,1 GGDEF
Maqu_3670	Response regulator, PleD family contains 1 Hpt,1 Response_reg,1 REC,1 GGDEF
Maqu_0247	Response regulator, PleD_VieA family contains 1 REC,1 PAS,1 GGDEF,1 EAL
Maqu_0463	Response regulator, PleD_VieA family contains 1 REC,1 GGDEF,1 EAL
Maqu_2223	Response regulator, PleD_VieA family contains 1 REC,1 PAS_4,1 GGDEF,1 EAL
Maqu_2785	Response regulator, PleD_VieA family contains 1 REC,1 PAS_4,1 GGDEF,1 EAL
Maqu_0299	Response regulator, PrrA family contains 1 Response_reg,1 HTH_8
Maqu_3669	Response regulator, RpfG family contains 1 Response_reg,1 HDOD
Maqu_3676	Response regulator, RpfG family contains 1 REC,1 HD
Maqu_1479	Response regulator, RsbU family contains 1 Response_reg,1 SpoIIE
Maqu_1991	Response regulator, RsbU family contains 1 REC,1 SpoIIE
Maqu_0670	Response regulator, unclassified contains 1 Response_reg,1 AAA
Maqu_0895	Response regulator, unclassified contains 1 Response_reg
Maqu_1851	Response regulator, unclassified contains 1 REC
Maqu_1888	Response regulator, unclassified contains 1 Response_reg
Maqu_2606	Response regulator, unclassified contains 1 REC
Maqu_2899	Response regulator, unclassified contains 1 REC
Maqu_3598	Response regulator, unclassified contains 1 REC
Maqu_0897	Response regulator, VieB family contains 1 REC,1 TPR_1

C.2. Translation of DNA Sequences

The maqu_0123 and maqu_124 ORFs (*copS* and *copR* genes) were inserted into the ExPASy server tool translate in order to obtain the corresponding protein sequences.

C.2.1 Frames of Translation for RR - CopR

5'3' Frame 1

```
MRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAREAQILGEQEDYRAAVL
DLGLPDGNGLEVLKRWRSKLVECPVLVLTARGDWQDKVNGLKAGADDYLA
KPFQTEELIARINALIRRSEGRVHSQVKAGGFELDENRQSLRTEEGAEHA
LTGTEFRLLRCLMSRPGHIFSKEQLMEQLYNLDESPSENVIEAYIRRLRK
LVGNETITTRRGQGYMFNAQR-
```

5'3' Frame 2

```
CVYCSLKTTVCWPRGWSDSWKKRGLASTTRPVPVRPRFWGSRRIIVPLFS
ISACRMATGWRF-NDGDRSSSSARYWSSPPEATGRTRSMD-KQGRMTIWP
NRSRQKN-SPASMHSYAAAKGECTPRLKPVASNWTKIARACGQKKEQNTP
-RVLSSACYDA--AVRATSFPRNS-WSSYTTWMRAPAKT-LRRIFGA-ES
WSATKRSPHAVARDTCSMPNV
```

5'3' Frame 3

```
AFIAR-RRPFAGRGAGQTAGKSGV-HRPHVQCP-GPDSGGAGGLSCRCSR
SRPAGWQRAGGFETMAIEARRVPGIGPHRQRRLAGQGQWTESRGG-LSGQ
TVPDRRTDRPHQCTHTPQRRASALPG-SRWLRTGRKSPEPADRRRSRTRP
DGY-VPPATMPDEPSGPHLFQGTANGAAIQPG-EPQRKRD-GVYSALEKA
GRQRNDHHTPWPGIHVQCPTL
```

3'5' Frame 1

```
LTLGIEHVSLATACGDRFVADQLSQAPNIRLNHVFAGALIQVV-LLH-LF
LGKDVARTAHQAS-QAELSTRQGVFCSFFCPQALAIFVQFEATGFNLGVH
SPFAAAYECIDAGDQFFCLERFGQIVIRPCFQSIDLVLPVASGGEDQYRA
LDELRSPSFQNLQPVAIRQAEIENSGTIILLLPQNLGLTGTGRVVDAKPR
FFQLSDQPLGQQTVVFNEQ-TH
```

3'5' Frame 2

```
-RWALNMYPWPRRVVIVSLPTSFLKRRIYASITFSLGLSSRLYSCSISCS
LEKMWPGRLIRHRSRRNSVPVRACSAPSSVRRLWRFSSSSKPPALTWECT
RPSLRRMSALMRAISSSVWNGLAR-SSAPAFSPLTLSCQSPLAVRTNTGH
STSFDRHRFKTSSPLPSGRPRSRTAAR-SSCSPRIWASRALDVWSMLNPA
FSSCLTSPSASKRSSSTSNKR
```

3'5' Frame 3

```
NVGH-TCIPGHGVW-SFRCRPAFSSAEYTPQSRFRWGSHPGCIAAPLAVP
WKRCGPDGSSGIVAGGTQYPSGRVLLLLLSAGSGDFRPVRSHRL-PGSAL
ALRCGV-VH-CGRSVLLSGTVWPDSHPPLLSVH-PCPASRLWR-GPIPGT
RRASIAIVSKPPARCHPAGRDREQRHDNPPAPPESGPHGHWTCGRC-TPL
FPAV-PAPRPANGRLQRAINA
```

Figure 6.5 Translation frames obtained upon insertion of *copR*, Maqu_0123 ORF from M.aquaeolei VT8 in the Translation tool from ExPASy. The chosen protein sequence is highlighted in grey.

C.2.2 Frames of Translation for Histidine Kinase – CopS

5'3' Frame 1

MPNVKRPASVKGMLLVLLLPAGIALMGVAWFVHGLLLDRMSREFLESRLKDEAAFLEHQI
REVRGQVETLQTGDYFQDVFHHAFAIRTPDRTIISPKAWEPLLAPLINHEQNGTLRLEGR
QAPDSPSDILAYRHSFQVNGSPIVVVVSEDLEALKRSQAELHAWTAVVSVLLIALLVAVI
WFGITLSLRPVVTLKAALKRLQDGEISRINAPSPEEFQPLVMQLNQLLDSLDKRLVRSRD
ALANLSHSVKTPIAAVRQILEDMDRPLPSDLRIQMAARLSDIDRQLEAEMRRSRFAGPQV
GKSAYPVKQARDLLWMLGRLYPEKSFELSSSLPEDTRWPIEEHDLNEVLGNLLDNAGKWS
SRCVELSLKQDNNSRQIVVSDDGPGVNGDDLSSLGQRGLRLDEQTPGHGLGLAIVREIVA
RYEGNISFSTGPGSGLRVTIEF-

5'3' Frame 2

CPTLKGQHPSKACCWCYCCPPASL-WA-PGLSTACCWIECPGNSLKAALKTKLPFWSTRF
ARSEVKSKPCRPVITSRTSSIMPSPYARPIEPSYRQRPGNPCWHH-STMSRMARFVLKAV
RRRTVRPISWHTVTLFR-TGHRSSWLYPRIWRP-NAVRPSCTPGQPWCRCC-SHFWLPSS
GLASRYRCGRW-H-KRR-NDCRMERFPGSMHHHPKNFNRS-CS-TSCLTPSTSDWCVPGM
RSQICHTVSKRPSQPSGRYWRIWIAHCLVI-ESKWPPVSVTSTDNWKRKCAAAALPGPRS
GKALIPSNRRGIFCGCWGGCIRKNPSNYQAHCRKTPAGR-RSMT-MKC-ATCSIMPANGR
RGA-NSR-NKTTTADKSLFLMMVLESMATTCPVWGNEGCDLTSRPLATVWASRLSGRSSL
AMRATLVFRRGPVAACA-P-SF

5'3' Frame 3

AQR-KASIRQRHAAGVIAARRHRSDGRSLVCPRPAAGSNVPGIP-KPP-RRSCLSGAPDS
RGQRSSRNPADR-LLPGRLPSCLRHTHARSNHHIAKGLGTLAGTTDQP-AEWHASS-RPS
GAGQSVRYPGIPSLFSGERVTDRRGCIRGSGGPETQSGRAARLDSRGVGAVDRTSGCRHL
VWHHAIAAAGGDTESGVKTIAGWRDFPDQCTITRRISTARDAVKPVA-LPRQATGAFPGC
ARKSVTQCQNAHRSRPADTGGYGSPIA--SKNPNGRPSQ-HRQTIGSGNAPQPLCRAPGR
EKRLSRQTGAGSFVDAGAAVSGKILRTIKLTAGRHPLADRGA-LE-SARQPAR-CRQMVV
AVRRTLAETRQQQQTNRCF--WSWSQWRRLVQSGATRAAT-RADPWPRSGPRDCPGDRRS
L-GQH-FFDGAR-RLARNHRVL

3'5' Frame 1

LKLYGYAQAATGPRRKTNVALIASDDLPDNREAQTVARGLLVKSQPSLPQTGQVVAIDSR
TIIRNNDLSAVVVLFQREFYAPRRPFAGIIEQVA-HFIQVMLLYRPAGVFRQ-A--FEGF
FRIQPPQHPQKIPRLFDGISAFPDLGPGKAAAAHFRFQLSVDVTETGGHLDS-ITRQWAI
HILQYLPDGCDGRFDTV-QICERIPGTHQSLVEGVKQLV-LHHERLKFFG-WCIDPGNLS
ILQSF-RRFQCHHRPQR-RDAKPDDGNQKCDQQHRHHGCPGVQLGLTAFQGLQILGYNHD
DR-PVHLKRVTVCQDIGRTVRRLTAFKTKRAILLMVDQWCQQGFPGLWRYDGSIGREYGE
GMMEDVLEVITGLQGFDLTSDLANLVLQKGSFVFKAAFKEFPGHSIQQQAVDKPGYAHQS
DAGGQQ-HQQHAFDGCWPFNVGH

3'5' Frame 2

-NSMVTRKPLPGPVEKLMLPS-RATISRTIARPRPWPGVCSSSRSPRCPRLDKSSPLTPG
PSSETTICLLLLSCFSESSTHRDDHLPALSSRLPSTSFKSCSSIGQRVSSGSELDSSKDF
SGYSRPSIHKRSRACLTG-ALFPTWGPAKRLRRISASNCLSMSLRRAAIWILRSLGNGRS
ISSSICRTAAMGVLTLCDRFASASRERTSRLSRESSNWFNCITSG-NSSGDGALIREISP
SCNRFNAAFSVTTGRSDSVMPNQMTATRSAINSTDTTAVQACSSA-LRFRASRSSDTTTT
IGDPFT-KE-RYARISDGLSGA-RPSRRSVPFCSWLISGASKGSQAFGDMMVRSGVRMAK
A-WKTSWK-SPVCRVST-PLTSRIWCSRKAASSLRRLSRNSRDIRSSSRPWTNQATPIRA
MPAGSNNTSSMPLTDAGLLTLG

3'5' Frame 3

KTLWLRASRYRAPSKN-CCPHSERRSPGQSRGPDRGQGSARQVAALVAPDWTSRRH-LQD
HHQKQRFVCCCCLVSARVLRTATTICRHYRAGCLALHSSHAPLSASGCLPAVSLIVRRIF
PDTAAPASTKDPAPV-RDKRFSRPGARQSGCGAFPLPIVCRCH-DGRPFGFLDH-AMGDP
YPPVSAGRLRWAF-HCVTDLRAHPGNAPVACRGSQATGLTASRAVEILRVMVH-SGKSLH
PAIVLTPLSVSPPAAAAIA-CQTR-RQPEVRSTAPTPRLSRRAARPDCVSGPPDPRIQPRR
SVTRSPEKSDGMPGYRTDCPAPDGLQDEACHSAHG-SVVPARVPRPLAI-WFDRACVWRR
HDGRRPGSNHRSAGFRLDL-PRESGAPERQLRL-GGFQGIPGTFDPAAGRGQTRLRPSER
CRRAAITPAACL-RMLAF-RWA

Figure 6.6 The *copS* gene was inserted into the ExPASy server tool, Translate, in order to obtain the corresponding protein sequence. The desired protein sequence is highlighted in grey.

C.3 Amino Acid Distribution

Amino acid frequency and distribution throughout the primary sequence of CopR and CopS_C from *M. aquaeolei* VT8 are presented in Table 6.5.

Table 6.5 Amino acid frequency and distribution throughout the primary sequence of CopR and CopS_C from *M. aquaeolei* VT8. Data obtained from the ExPASy bioinformatic tool ProtParam.

		CopR		CopS_C	
Amino Acid	**R Group**	**Number of residues**	**Frequency (%)**	**Number of residues**	**Frequency (%)**
Ala (A)	Small	17	7.69	16	6.2
Arg (R)	Basic	23	10.41	23	9.0
Asn (N)	Amide	8	3.62	10	3.9
Asp (D)	Acidic	12	5.43	19	7.4
Cys (C)	Nucleophilic	2	0.90	1	0.4
Gln (Q)	Amide	11	4.98	14	5.5
Glu (E)	Acidic	23	10.41	18	7.0
Gly (G)	Small	18	8.14	19	7.4
His (H)	Basic	4	1.81	3	1.2
Ile (I)	Hydrophobic	9	4.07	12	4.7
Leu (L)	Hydrophobic	31	14.03	36	14.1
Lys (K)	Basic	9	4.07	8	3.1
Met (M)	Hydrophobic	4	1.81	6	2.3
Phe (F)	Aromatic	6	2.71	6	2.3
Pro (P)	Hydrophobic	5	2.26	15	5.9
Ser (S)	Nucleophilic	11	4.98	24	9.4
Thr (T)	Nucleophilic	9	4.07	5	2.0
Trp (W)	Aromatic	2	0.90	3	1.2
Tyr (Y)	Aromatic	5	2.26	3	1.2
Val (V)	Hydrophobic	12	5.43	15	5.9

C.4 Subcellular localization

For prediction of the subcellular localization of the desired proteins, CopR and CopS, various predictors were utilized in conjunction to reduce bias. The PSORTb [97] (version 3.0.2.), sub**CEL**lular **LO**calization predictor – CELLO [98] (version 2.5.) and Gneg-mPLoc [99] (version 2.0.) predictors were used and their accuracy,as well as the obtained results are presented in Table 6.6.

Table 6.6 Predictions for CopR and CopS protein subcellular localization and the accuracies of the individual predictors used [97].

Predictor	Predictor accuracy	CopR	CopS
PSORTb	96%	Cytoplasm	Cytoplasmic Membrane
CELLO	95%	Cytoplasm	Inner Membrane
Gneg-mPLoc	96%	Cytoplasm	Cell inner membrane

C.5 Domains

For determination of the CopR and CopS protein domains, the SMART bioinformatic tool was utilized, and these results were corroborated with the CDD and Pfam predictions of protein domains (Tables 6.7 and 6.8). The differences of the predictions in the effector domain limits are due to the inclusion or not of the characteristic 4 β-sheets preceding the effector domain in the OmpR/PhoB family.

Table 6.7 Predictions for CopR protein domain confines from three independent prediction servers, SMART, CDD and Pfam.

Predictor	Prediction	
	Receiver	Effector
SMART	1-112	146-217
CDD	3-113	127-217
Pfam	3-113	147-217

Table 6.8 Predictions for CopS protein domain confines from three independent prediction servers, SMART, CDD and Pfam.

Predictor	Prediction			
	Transmembrane	**HAMP**	**HisKA**	**HATPase_c**
SMART	15-37/163-185	-	237-298	340-442
CDD	-	181-231	-	345-442
Pfam	-	165-233	237-293	341-442

C.6 Secondary Structure Prediction

For secondary structure prediction the same principle was applied as for subcellular localization and protein domain prediction. In this case the main reason for use of multiple prediction servers is due to the diverse prediction formulae and the errors associated with β-sheet prediction. The Porter [101], PSIPRED [102], NetSurfP [103], PSSpred [104] and JPred 3 [106] were used for the secondary structure predictions presented in Figures 6.7 and 6.8 for CopR and CopS respectively.

C.6.1 CopR

Porter

```
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL
VECPVLVLTA RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS
LRTEEGAEHA LTGTEFRLLR CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR
RGQGYMFNAQ R
```

PSIPRED

```
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL
VECPVLVLTA RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS
LRTEEGAEHA LTGTEFRLLR CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR
RGQGYMFNAQ R
```

NetSurfP

```
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL
VECPVLVLTA RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS
LRTEEGAEHA LTGTEFRLLR CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR
RGQGYMFNAQ R
```

PSSpred

```
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL
VECPVLVLTA RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS
LRTEEGAEHA LTGTEFRLLR CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR
RGQGYMFNAQ R
```

JPred 3

```
MRLLLVEDDR LLAEGLVRQL EKAGFSIDHT SSAREAQILG EQEDYRAAVL DLGLPDGNGL EVLKRWRSKL
VECPVLVLTA RGDWQDKVNG LKAGADDYLA KPFQTEELIA RINALIRRSE GRVHSQVKAG GFELDENRQS
LRTEEGAEHA LTGTEFRLLR CLMSRPGHIF SKEQLMEQLY NLDESPSENV IEAYIRRLRK LVGNETITTR
RGQGYMFNAQ R
```

Figure 6.7 Secondary structure predictions for the CopR protein from various bioinformatic servers, Porter, PSIPRED, NetSurfP, PSSpred, JPred 3. The α-helices and β-sheets are presented in red and blue respectively.

C.6.2 CopS

Porter

```
MPNVKRPASV KGMLLVLLLP AGIALMGVAW FVHGLLLDRM SREFLESRLK DEAAFLEHQI REVRGQVETL
QTGDYFQDVF HHAFAIRTPD RTIISPKAWE PLLAPLINHE QNGTLRLEGR QAPDSPSDIL AYRHSFQVNG
SPIVVVVSED LEALKRSQAE LHAWTAVVSV LLIALLVAVI WFGITLSLRP VVTLKAALKR LQDGEISRIN
APSPEEFQPL VMQLNQLLDS LDKRLVRSRD ALANLSHSVK TPIAAVRQIL EDMDRPLPSD LRIQMAARLS
DIDRQLEAEM RRSRFAGPQV GKSAYPVKQA RDLLWMLGRL YPEKSFELSS SLPEDTRWPI EEHDLNEVLG
NLLDNAGKWS SRCVELSLKQ DNNSRQIVVS DDGPGVNGDD LSSLGQRGLR LDEQTPGHGL GLAIVREIVA
RYEGNISFST GPGSGLRVTI EF
```

NetSurfP

```
MPNVKRPASV KGMLLVLLLP AGIALMGVAW FVHGLLLDRM SREFLESRLK DEAAFLEHQI REVRGQVETL
QTGDYFQDVF HHAFAIRTPD RTIISPKAWE PLLAPLINHE QNGTLRLEGR QAPDSPSDIL AYRHSFQVNG
SPIVVVVSED LEALKRSQAE LHAWTAVVSV LLIALLVAVI WFGITLSLRP VVTLKAALKR LQDGEISRIN
APSPEEFQPL VMQLNQLLDS LDKRLVRSRD ALANLSHSVK TPIAAVRQIL EDMDRPLPSD LRIQMAARLS
DIDRQLEAEM RRSRFAGPQV GKSAYPVKQA RDLLWMLGRL YPEKSFELSS SLPEDTRWPI EEHDLNEVLG
NLLDNAGKWS SRCVELSLKQ DNNSRQIVVS DDGPGVNGDD LSSLGQRGLR LDEQTPGHGL GLAIVREIVA
RYEGNISFST GPGSGLRVTI EF
```

JPred 3

```
MPNVKRPASV KGMLLVLLLP AGIALMGVAW FVHGLLLDRM SREFLESRLK DEAAFLEHQI REVRGQVETL
QTGDYFQDVF HHAFAIRTPD RTIISPKAWE PLLAPLINHE QNGTLRLEGR QAPDSPSDIL AYRHSFQVNG
SPIVVVVSED LEALKRSQAE LHAWTAVVSV LLIALLVAVI WFGITLSLRP VVTLKAALKR LQDGEISRIN
APSPEEFQPL VMQLNQLLDS LDKRLVRSRD ALANLSHSVK TPIAAVRQIL EDMDRPLPSD LRIQMAARLS
DIDRQLEAEM RRSRFAGPQV GKSAYPVKQA RDLLWMLGRL YPEKSFELSS SLPEDTRWPI EEHDLNEVLG
NLLDNAGKWS SRCVELSLKQ DNNSRQIVVS DDGPGVNGDD LSSLGQRGLR LDEQTPGHGL GLAIVREIVA
RYEGNISFST GPGSGLRVTI EF
```

PSIPred

```
MPNVKRPASV KGMLLVLLLP AGIALMGVAW FVHGLLLDRM SREFLESRLK DEAAFLEHQI REVRGQVETL
QTGDYFQDVF HHAFAIRTPD RTIISPKAWE PLLAPLINHE QNGTLRLEGR QAPDSPSDIL AYRHSFQVNG
SPIVVVVSED LEALKRSQAE LHAWTAVVSV LLIALLVAVI WFGITLSLRP VVTLKAALKR LQDGEISRIN
APSPEEFQPL VMQLNQLLDS LDKRLVRSRD ALANLSHSVK TPIAAVRQIL EDMDRPLPSD LRIQMAARLS
DIDRQLEAEM RRSRFAGPQV GKSAYPVKQA RDLLWMLGRL YPEKSFELSS SLPEDTRWPI EEHDLNEVLG
NLLDNAGKWS SRCVELSLKQ DNNSRQIVVS DDGPGVNGDD LSSLGQRGLR LDEQTPGHGL GLAIVREIVA
RYEGNISFST GPGSGLRVTI EF
```

Figure 6.8 Secondary structure predictions for the full-length CopS protein from various bioinformatic servers, Porter, PSIPRED, NetSurfP, PSSpred, JPred 3. The α-helices and β-sheets are presented in red and blue respectively.

C.7 Homology alignments

The bacterial strains from which protein sequences of response regulators and histidine kinases closely related to the CopR and CopS proteins were retrieved, are listed in Table 6.9 and 6.10. These were aligned using the ClustalW tool, Figures 3.6 and 3.72 from Chapter 3 Results and Discussion, obtaining the scores, also identified in the tables, which allowed identification of the sequences with over 50% identity.

Table 6.9 Bacterial strains from which RR protein sequences which share above 50% identity from those that were retrieved and aligned using the ClustalW tool.

Name	Strain	Score
MELB17_11654	*M. sp. ELB17*	79.0
MDG893_10211	*M. algicola DG893*	77.0
ABO_1365	*Alcanivorax borkumensis SK2*	73.0
Alvin_0011	*Allochromatium vinosum DSM 180*	52.0
MAMP_00174	*Methylophaga aminisulfidivorans MP*	52.0
PST_0851	*Pseudomonas stutzeri A1501*	51.0
Mmc1_0312	*Magnetococcus sp. MC-1*	51.0

Table 6.10 Bacterial strains from which HK protein sequences which share above 50% identity from those that were retrieved and aligned using the ClustalW tool.

Name	Strain	Score
KYE_02303	*M.* sp. Mnl7-9	77.0
HP15_186	*M. adhaerens* HP15	73.0
KYE_16483	*M.* sp. Mnl7-9	72.0
MDG893_10216	*M.* algicola DG893	72.0
MELB17_11649	*M.* sp. ELB17	71.0
KYE_00831	*M.* sp. Mnl7-9	68.0
ABO_1366	*Alcanivorax borkumensis* SK2	63.0
ADG881_2036	*Alcanivorax* sp. DG881	61.0

C.8 Template for homology modelling - DrrD

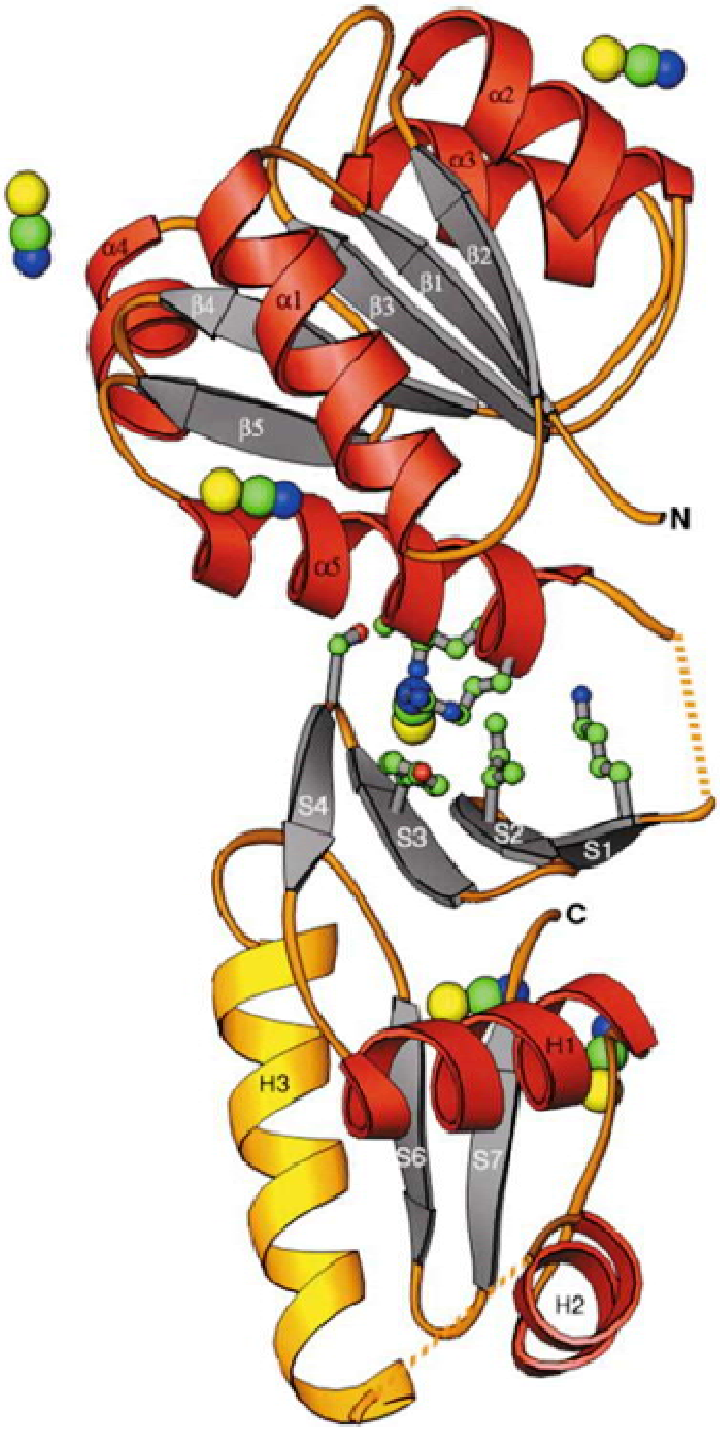
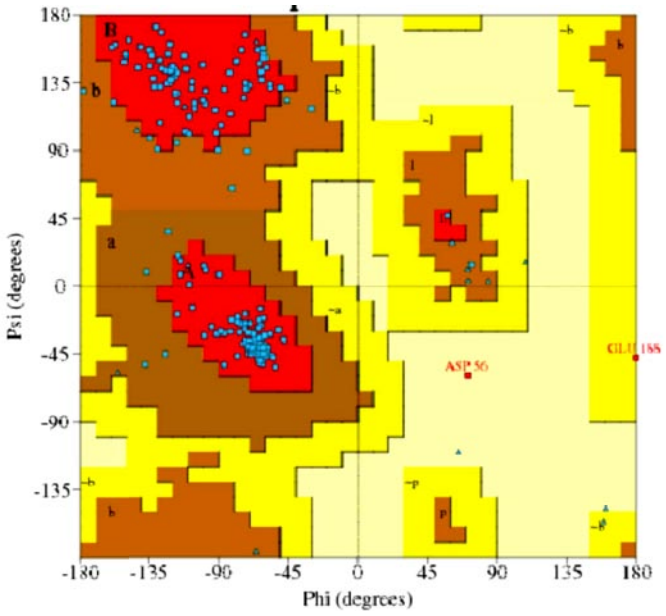


Figure 6.9 Structure of Full-Length DrrD from *T. maritima*. A ribbon representation of the DrrD structure highlighting residues at the interdomain interface (ball and stick rendering) and six thiocyanate ions (CPK rendering). For the stick and CPK renderings, carbon is in green, nitrogen is in blue, oxygen is in red, and sulphur is in yellow. Secondary structural elements are labelled α1–5 and β1–5 for the regulatory domain and H1–3 and S1–7 for the effector domain. The recognition helix, H3, in the effector domain is in gold. Two regions through which electron density was not observed, the interdomain linker and the loop connecting H2 and H3, are indicated by dashed lines. This figure was generated using Ribbons Version 3.14. Figure adapted from Buckler, D R, Yuchen Zhou, and A. M. Stock. 2002. "Evidence of intradomain and interdomain flexibility in an OmpR/PhoB homolog from *Thermotoga maritima*." Structure 10(2): 153-164.

C.9. Homology Model Validity

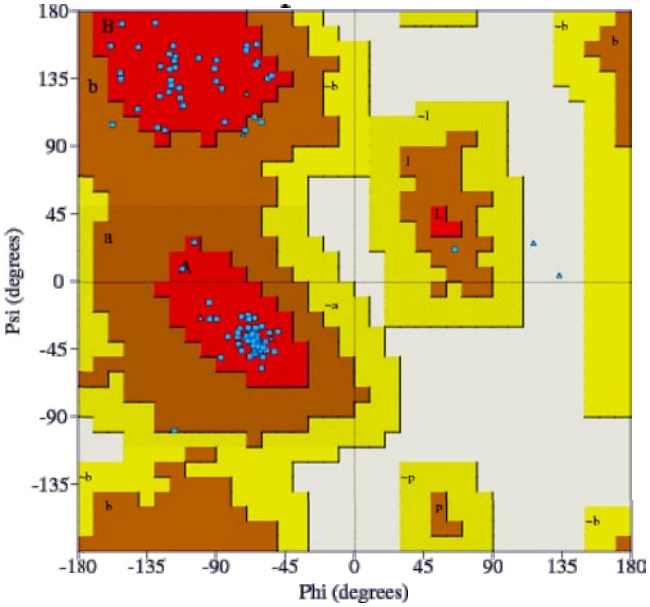
C.9.1 CopR



Plot statistics		
	No. of residues	**Percentage**
Residues in most favoured regions [A,B,L]	180	92.3%
Residues in additional allowed regions [a,b,l,p]	13	6.7%
Residues in generously allowed regions [-a,-b,-l,-p]	1	0.5%
Residues in disallowed regions [XX]	1	0.5%
Number of non-glycine and non-proline residues	195	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues	18	
Number of proline residues	5	
Total number of residues	220	

Figure 6.10 Ramachandran plot of the predicted model of *M.aquaeolei* VT8, CopR. This figure is generated by PDBsum bioinformatic system, the PROCHECK bioinformatic tool. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R-factor* no grater than 20%, a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

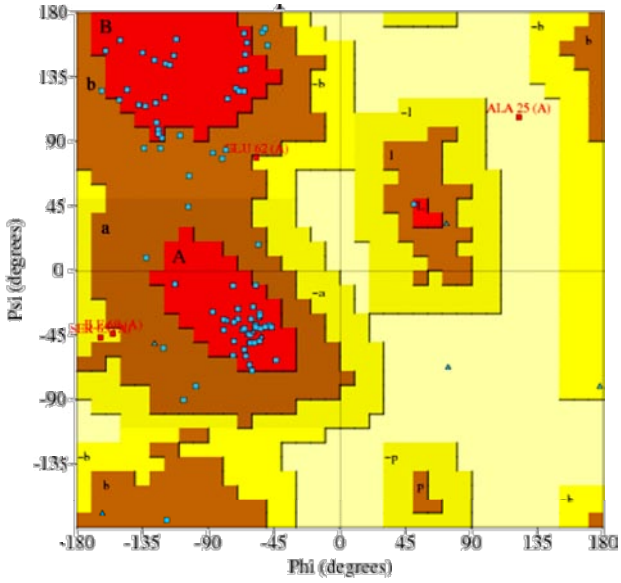
C.9.2 CopR_N



Plot statistics		
	No. of residues	**Percentage**
Residues in most favoured regions [A,B,L]	101	95.3%
Residues in additional allowed regions [a,b,l,p]	5	4.7%
Residues in generously allowed regions [-a,-b,-l,-p]	0	0.0%
Residues in disallowed regions [XX]	0	0.0%
Number of non-glycine and non-proline residues	106	100.0%
Number of end-residues (excl. Gly and Pro)	3	
Number of glycine residues	9	
Number of proline residues	3	
Total number of residues	121	

Figure 6.11 Ramachandran plot of the predicted model of *M.aquaeolei* VT8 CopR_N. This figure is generated by PDBsum bioinformatic system, the PROCHECK bioinformatic tool. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R-factor* no grater than 20%, a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

C.9.3 CopR_C



Plot statistics		
	No. of residues	**Percentage**
Residues in most favoured regions [A,B,L]	62	73.8%
Residues in additional allowed regions [a,b,l,p]	18	21.4%
Residues in generously allowed regions [-a,-b,-l,-p]	3	3.6%
Residues in disallowed regions [XX]	1	1.2%
Number of non-glycine and non-proline residues	84	100.0%
Number of end-residues (excl. Gly and Pro)	2	
Number of glycine residues	8	
Number of proline residues	2	
Total number of residues	96	

Figure 6.12 Ramachandran plot of the predicted model of *M.aquaeolei* VT8 CopR_C. This figure is generated by PDBsum bioinformatic system, the PROCHECK bioinformatic tool. The red regions in the graph indicate the most allowed regions whereas the yellow regions represent allowed regions. Based on an analysis of 118 structures of resolution of at least 2.0 Angstroms and *R-factor* no grater than 20%, a good quality model would be expected to have over 90% in the most favoured regions [A,B,L].

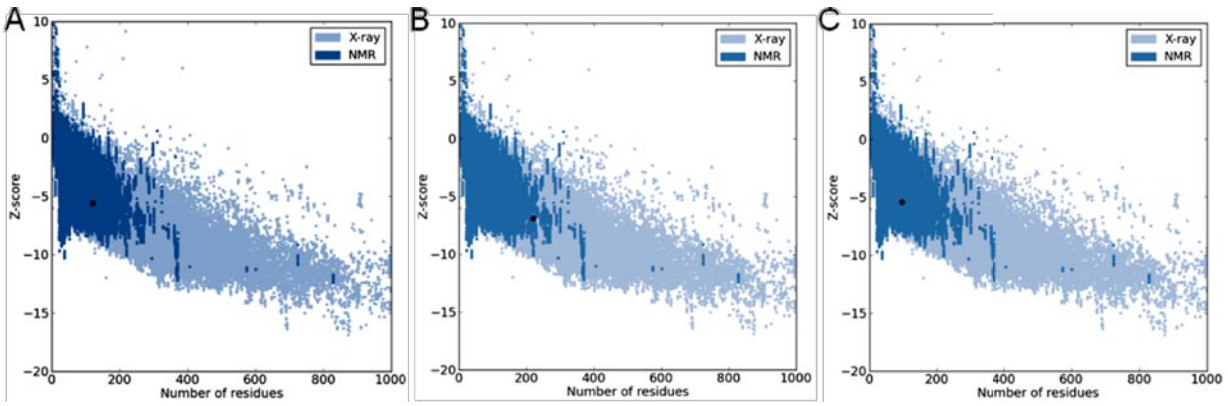


Figure 6.13 ProSA plot for the **A)** full-length CopR protein, **B)** N-terminal domain of CopR, CopR_N and **C)** C-terminal domain of CopR, CopR_C. The z-score obtained for each of these models was -6.88, -5.59 and -5.39, respectively.

C.10 Circular dichroism

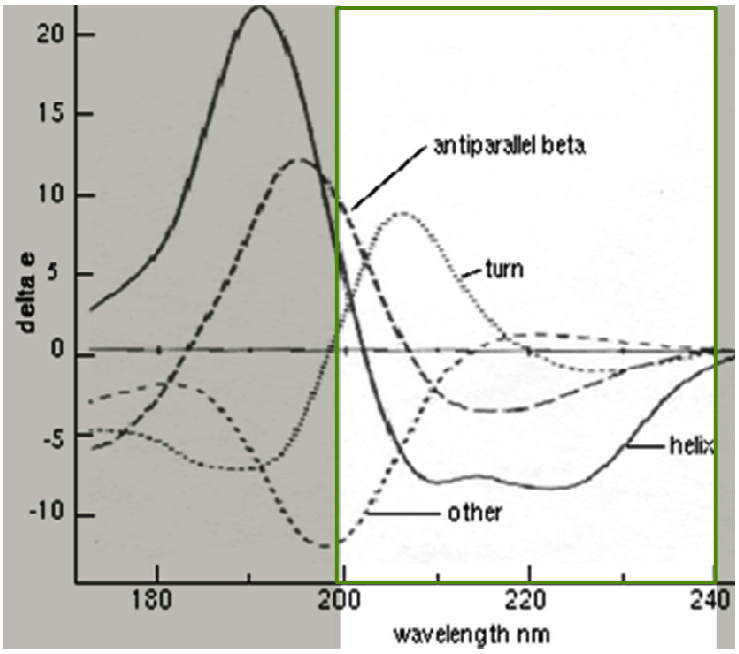


Figure 6.14 Representation of circular dichroism spectra for pure secondary structure elements. The highlighted region corresponds to the regions utilized in the assays performed in this work.

C.11. Promoter Analysis

The DNA sequence for the entire *copRSXAB* operon was retrieved from the KEGG database[115], as well as the complementary DNA strand in order to analyze this region in terms of promoter sequences. The actual analysis is presented in Figure 3.42, from Chapter 3 Results and Discussion – Section 3.2.3.5. In purple, red, green and blue are the *copS/R*, *copX*, *copA* and *copB* ORFs as delineated in this database.

C.11.1 *M.aquaeolei* VT8 DNA (150214 to 157339 nt of the genome)

TTAAAACTCTATGGTTACGCGCAAGCCGCTACCGGGCCCCGTCGAAAAACTAATGTTGCCCTCATAGCGAGCGACGATCTCCCGGA
CAATCGCGAGGCCCAGACCGTGGCCAGGGGTCTGCTCGTCAAGTCGCAGCCCTCGTTGCCCCAGACTGGACAAGTCGTCGCCATTG
ACTCCAGGACCATCATCAGAAACAACGATTTGTCTGCTGTTGTTGTCTTGTTTCAGCGAGAGTTCTACGCACCGCGACGACCATTT
GCCGGCATTATCGAGCAGGTTGCCTAGCACTTCATTCAAGTCATGCTCCTCTATCGGCCAGCGGGTGTCTTCCGGCAGTGAGCTTG
ATAGTTCGAAGGATTTTTCCGGATACAGCCGCCCCAGCATCCACAAAAGATCCCGCGCCTGTTTGACGGGATAAGCGCTTTTCCCG
ACCTGGGGCCCGGCAAAGCGGCTGCGGCGCATTTCCGCTTCCAATTGTCTGTCGATGTCACTGAGACGGGCGGCCATTTGGATTCT
TAGATCACTAGGCAATGGGCGATCCATATCCTCCAGTATCTGCCGGACGGCTGCGATGGGCGTTTTGACACTGTGTGACAGATTTG
CGAGCGCATCCCGGGAACGCACCAGTCGCTTGTCGAGGGAGTCAAGCAACTGGTTTAACTGCATCACGAGCGGTTGAAATTCTTCG
GGTGATGGTGCATTGATCCGGGAAATCTCTCCATCCTGCAATCGTTTTAACGCCGCTTTCAGTGTCACCACCGGCCGCAGCGATAG
CGTGATGCCAAACCAGATGACGGCAACCAGAAGTGCGATCAACAGCACCGACACCACGGCTGTCCAGGCGTGCAGCTCGGCCTGAC
TGCGTTTCAGGGCCTCCAGATCCTCGGATACAACCACGACGATCGGTGACCCGTTCACCTGAAAAGAGTGACGGTATGCCAGGATA
TCGGACGGACTGTCCGGCGCCTGACGGCCTTCAAGACGAAGCGTGCCATTCTGCTCATGGTTGATCAGTGGTGCCAGCAAGGGTTC
CCAGGCCTTTGGCGATATGATGGTTCGATCGGGCGTGCGTATGGCGAAGGCATGATGGAAGACGTCCTGGAAGTAATCACCGGTCT
GCAGGGTTTCGACTTGACCTCTGACCTCGCGAATCTGGTGCTCCAGAAAGGCAGCTTCGTCTTTAAGGCGGCTTTCAAGGAATTCC
CGGGACATTCGATCCAGCAGCAGGCCGTGGACAAACCAGGCTACGCCCATCAGAGCGATGCCGGCGGGCAGCAATAACACCAGCAG
CATGCCTTTGACGGATGCTGGCCTTTTAACGTTGGGCATTGAACATGTATCCCTGGCCACGGCGTGTGGTGATCGTTTCGTTGCCG
ACCAGCTTTCTCAAGCGCCGAATATACGCCTCAATCACGTTTTCGCTGGGGCTCTCATCCAGGTTGTATAGCTGCTCCATTAGCTG

TTCCTTGGAAAAGATGTGGCCCGGACGGCTCATCAGGCATCGTAGCAGGCGGAACTCAGTACCCGTCAGGGCGTGTTCTGCTCCTT
CTTCTGTCCGCAGGCTCTGGCGATTTTCGTCCAGTTCGAAGCCACCGGCTTTAACCTGGGAGTGCACTCGCCCTTCGCTGCGGCGT
ATGAGTGCATTGATGCGGGCGATCAGTTCTTCTGTCTGGAACGGTTTGGCCAGATAGTCATCCGCCCCTGCTTTCAGTCCATTGAC
CTTGTCCTGCCAGTCGCCTCTGGCGGTGAGGACCAATACCGGGCACTCGACGAGCTTCGATCGCCATCGTTTCAAAACCTCCAGCC
CGTTGCCATCCGGCAGGCCGAGATCGAGAACAGCGGCACGATAATCCTCCTGCTCCCCCAGAATCTGGGCCTCACGGGCACTGGAC
GTGTGGTCGATGCTAAACCCCGCTTTTTCCAGCTGTCTGACCAGCCCCTCGGCCAGCAAACGGTCGTCTTCAACGAGCAATAAACG
CATGTATCCGTCCTTAACGGTGCTTGTTTCAGCCGCCAGTGTCGTCGATCAACCTGAATTGTGCCTGAACGCAAGAAAGTGCCCGT
TTTTCAGAAAGAATTCAGGTTGGTGAGAGAAAGTGTCGACTGATTCTGGACATTGCTCAAGATTCAGTACGAATCAGCCGTGTTCT
CAAAACCATTATTCCACAGGAGACAAACGGATGAATGCATCAAACCTGATTGTGGCCGGCCTGGCTTTTTCGATGTCGGTATCGGC
GTTTGCGGCTGGCACCCATGGCGGTGGTCACGGCCATGGTGCTTCCATTGGCGAACCGGGAAAAGCCTCAGAAGCCAGCCGGACCA
TAACGGTTGAAATGCACGACAACTACTACGAACCCGAAGAAATCCGTGTGAAGCCGGGCGAAACCGTTCGGTTTGTGGTGCAGAAC
AAGGGCAATCTTGTGCACGAGTTCAACATCGGGACTCCGGGTATGCATGAGGCCCACCAGAAAGAAATGAGAATGATGGTCGAGCA
TGGCGTGATCCAGGGTAACAAGTTGAATCATGACATGATGAACATGGACATGGGCAACGGCCACTCGATGAAACATGACGATCCCA
ACAGCGTCCTGCTGGAGCCGGGCCAGAGCCGGGAAGTGGTCTGGACGTTTGCCAATCAGGGCAATATCGAATTCGCATGTAACGTA
CCAGGGCACTACCAGTCTGGCATGTACGGTGACGTGAATTTCGAATAAGCCGGTAGATGGCTTTAACCATGACGTTATGGGCTGAC
GCTACACCCTCCTGTCAGCCCCGATAAAGTGCGTTTGGAGCAAAACACATGAACATTCGCTTGGGTGCAGTAGTGCTGTGCCTGCT
GATGCCCGCACTGGCGTTGGCTGGCGAGTACGACCTGACGGTTGATCGCGTAGAGATAGATACCGGTGATTTTGTAAAGCAGGGCA
TCGGCTACAACGGCAAGTCTCCCGGGCCGGTACTGAGATTCAAGGAAGGTGAGACTGTCCGGATCATTGTGACGAATAACCTTGAT
GAAATGACATCCATCCACTGGCACGGGCTTATCCTGCCCTATCAACAGGATGGCGTGCCAGGTATCAGTTTCCCGGGTATCAAACC
GGGTGAGACCTTTACCTACGAGTTTCCCATTCAGCAGGCTGGGACCTACTGGTTTCACAGCCACTCGGGCTTTCAGGAGCCGGATG
GGGCCTACGGTGCGATCGTGATCGAACCTGAGGGCCGGGAACCCTTCCGGTACGACCGGGAATACGTGGTCCAACTAACGGACAAG
CATCCGCATTCCGGTGACCGCATCATGCGCAACCTGAAAATGATGCCGGACTACTACAACCGTAAACAGCAGACCGTTGGTGAGTT
TTTCTCTGATAGCTCCGAACAGGGGTTCTGGAACACCGTCCGGGATCGCCTGGCCTGGGGTGATATGCGCATGATGAAAGCGGATG
TAGAGGATCTCCAGGGGTTTACGGGCCTGATTAACGGCAAAGGCCCGGAGCAGAACTGGACGGGGTTGTTTGAGCCGGGTGAACGC
ATTCGGCTGCGGTTCATCAATTCCTCCGCCATGACCTATTTCGATGTCCGGATTCCGGGCCTGGATATGACGGTGGTGCAGGCGGA
TGGCAATAACGTGCAACCGGTGACGGTGGACGAATTCCGCATAGGCGTCGCGGAAACCTATGACGTGATCGTGCGCCCGAGAGATG
AACAGGCCTACACCATCTTCGCCGAATCCATGGGGCGTTCCGGATACGCCCGGGCAACACTGGCTCCTGAAGAGGGAATGGAAGCG
GCCGTGCCGCCACTGCGAGAACCTGCCCGCCTGACCATGGCGGATATGGGGAATATGCATGGTATGGATCATGGAAACATGGATAT
GGGCAGCTCGGACGATATGGCCGGCATGGACCACTCGTCGATGTCGGGAATGGATCACTCCGGCATGTCTGGCATGAATCACGGCT
CCATGATGATGGATAACCAGGCTCAAAATGATCCGTTTTACGCCAGAGGCAGTGGCCTTGTGCCGACCGCGGCCAATGGCGGTAAG
TTCCTCTCCTATGCCGATCTGAGGGCGCAGAATCCGTTATACGAGGAGCGTGAACCGACCCGGGAAATCGAGCTCCGCCTGACCGG
TAATATGGAACGTTACAGCTGGAGCATTAATGGCGTTAAGTATGAAGAGGCTGATCCGATTCGTCTCCAATACGGGGAGCGCGTCC
GGTTCAAGTTCGTCAATGAAACCATGATGACGCACCCCATGCATCTGCACGGCATGTGGTCGATTCTTGACGTGGGCGCCGGTCAG
TGGAACCCGATCAAACACACCATTAACGTGAACCCGGGTACCACGGTGTATATGGAAACGGAGGTTGATGAACCGGGCCAGTGGGC
ATTCCACTGCCATCTGTCCTATCACGCAGCTGCCGGTATGTTCCGAAAAGTCATCGTCGAGGGGGGCCCAGATTCGCAGCAGGCGA
AAACAGAGGCGATGGCTGAAGATGGAGGTGAGGCATGAGTGGGGTGCGATTCTCTGGGGCCGTCAGTTTGTTTGCCTTCCTTGTAT
TGCCGGCTATGGCCGCGGCACAGGCAATGATTGCAGATACCTCTGGCCGCAAGCAACCGCTGACAACCTGGGGTATTCAGTTTGAA
GAACTTGAATACCGATACAGTGATGACGATGAGGGACTGGGCGTCTGGAACGCCGATGCCTTCTACGGTACAGATGATTTCAAAGT
GCGGCTGCTCTCGACGGGCGAATACGAGATTGAGGAGCAGGCCTATGAAACGCTGGAACATCAGCTGGTAGGTCAGATCCCCGTTT
CCAGGTTCTTTGATGCCAAAGCGGGTGTTCGTTTCGATACGCCGGAGGGCCCGGACCGCACCTACGCTGTTTTCGGTGTGACTGGA
CTGGCGCCGCAGTGGTTCGAGATCGATGCCAGCCTGTATGTAAGCAAAGAGGGTGATACCTCGGCCGCGCTGGATGCAGAGTACGA
ACTGCTGTTCACCAACCACTGGATTCTGACAGCAACACTGGATGCAACCGTGGCGTTCAGCGAGGACGAGGAAATCGGTATCGGCA
AAGGACTGGTTTCTACGGAAACAGGCCTGCGGCTGAGCTATGACTTTATTGACCGCGCCGTCTCACCCTATTTCGGTGTGGTGCAT
GAACGGAAGTACGGTGATACCGCTGATCTGGCTGAGACCGCTGCTGGAAGCACGGAAGACTGGTTTGCCGTGATCGGCGCCCGGAT
CGCTTTCTGA

C.11.2 *M.aquaeolei* VT8 complement DNA (complement of 150214 to 157339 nt of the genome)

TCAGAAAGCGATCCGGGCGCCGATCACGGCAAACCAGTCTTCCGTGCTTCCAGCAGCGGTCTCAGCCAGATCAGCGGTATCACCGT
ACTTCCGTTCATGCACCACACCGAAATAGGGTGAGACGGCGCGGTCAATAAAGTCATAGCTCAGCCGCAGGCCTGTTTCCGTAGAA
ACCAGTCCTTTGCCGATACCGATTTCCTCGTCCTCGCTGAACGCCACGGTTGCATCCAGTGTTGCTGTCAGAATCCAGTGGTTGGT
GAACAGCAGTTCGTACTCTGCATCCAGCGCGGCCGAGGTATCACCCTCTTTGCTTACATACAGGCTGGCATCGATCTCGAACCACT
GCGGCGCCAGTCCAGTCACACCGAAAACAGCGTAGGTGCGGTCCGGGCCCTCCGGCGTATCGAAACGAACACCCGCTTTGGCATCA
AAGAACCTGGAAACGGGGATCTGACCTACCAGCTGATGTTCCAGCGTTTCATAGGCCTGCTCCTCAATCTCGTATTCGCCCGTCGA
GAGCAGCCGCACTTTGAAATCATCTGTACCGTAGAAGGCATCGGCGTTCCAGACGCCCAGTCCCTCATCGTCATCACTGTATCGGT
ATTCAAGTTCTTCAAACTGAATACCCCAGGTTGTCAGCGGTTGCTTGCGGCCAGAGGTATCTGCAATCATTGCCTGTGCCGCGGCC
ATAGCCGGCAATACAAGGAAGGCAAACAAACTGACGGCCCCAGAGAATCGCACCCCACTCATGCCTCACCTCCATCTTCAGCCATC
GCCTCTGTTTTCGCCTGCTGCGAATCTGGGCCCCCCTCGACGATGACTTTTCGGAACATACCGGCAGCTGCGTGATAGGACAGATG
GCAGTGGAATGCCCACTGGCCCGGTTCATCAACCTCCGTTTCCATATACACCGTGGTACCCGGGTTCACGTTAATGGTGTGTTTGA
TCGGGTTCCACTGACCGGCGCCCACGTCAAGAATCGACCACATGCCGTGCAGATGCATGGGGTGCGTCATCATGGTTTCATTGACG
AACTTGAACCGGACGCGCTCCCCGTATTGGAGACGAATCGGATCAGCCTCTTCATACTTAACGCCATTAATGCTCCAGCTGTAACG
TTCCATATTACCGGTCAGGCGGAGCTCGATTTCCCGGGTCGGTTCACGCTCCTCGTATAACGGATTCTGCGCCCTCAGATCGGCAT
AGGAGAGGAACTTACCGCCATTGGCCGCGGTCGGCACAAGGCCACTGCCTCTGGCGTAAAACGGATCATTTTGAGCCTGGTTATCC
ATCATCATGGAGCCGTGATTCATGCCAGACATGCCGGAGTGATCCATTCCCGACATCGACGAGTGGTCCATGCCGGCCATATCGTC
CGAGCTGCCCATATCCATGTTTCCATGATCCATACCATGCATATTCCCCATATCCGCCATGGTCAGGCGGGCAGGTTCTCGCAGTG
GCGGCACGGCCGCTTCCATTCCCTCTTCAGGAGCCAGTGTTGCCCGGGCGTATCCGGAACGCCCCATGGATTCGGCGAAGATGGTG
TAGGCCTGTTCATCTCTCGGGCGCACGATCACGTCATAGGTTTCCGCGACGCCTATGCGGAATTCGTCCACCGTCACCGGTTGCAC
GTTATTGCCATCCGCCTGCACCACCGTCATATCCAGGCCCGGAATCCGGACATCGAAATAGGTCATGGCGGAGGAATTGATGAACC
GCAGCCGAATGCGTTCACCCGGCTCAAACAACCCCGTCCAGTTCTGCTCCGGGCCTTTGCCGTTAATCAGGCCCGTAAACCCCTGG
AGATCCTCTACATCCGCTTTCATCATGCGCATATCACCCCAGGCCAGGCGATCCCGGACGGTGTTCCAGAACCCCTGTTCGGAGCT
ATCAGAGAAAAACTCACCAACGGTCTGCTGTTTACGGTTGTAGTAGTCCGGCATCATTTTCAGGTTGCGCATGATGCGGTCACCGG
AATGCGGATGCTTGTCCGTTAGTTGGACCACGTATTCCCGGTCGTACCGGAAGGGTTCCCGGCCCTCAGGTTCGATCACGATCGCA
CCGTAGGCCCCATCCGGCTCCTGAAAGCCCGAGTGGCTGTGAAACCAGTAGGTCCCAGCCTGCTGAATGGGAAACTCGTAGGTAAA
GGTCTCACCCGGTTTGATACCCGGGAAACTGATACCTGGCACGCCATCCTGTTGATAGGGCAGGATAAGCCCGTGCCAGTGGATGG
ATGTCATTTCATCAAGGTTATTCGTCACAATGATCCGGACAGTCTCACCTTCCTTGAATCTCAGTACCGGCCCGGGAGACTTGCCG
TTGTAGCCGATGCCCTGCTTTACAAAATCACCGGTATCTATCTCTACGCGATCAACCGTCAGGTCGTACTCGCCAGCCAACGCCAG
TGCGGGCATCAGCAGGCACAGCACTACTGCACCCAAGCGAATGTTCATGTGTTTTGCTCCAAACGCACTTTATCGGGGCTGACAGG
AGGGTGTAGCGTCAGCCCATAACGTCATGGTTAAAGCCATCTACCGGCTTATTCGAAATTCACGTCACCGTACATGCCAGACTGGT
AGTGCCCTGGTACGTTACATGCGAATTCGATATTGCCCTGATTGGCAAACGTCCAGACCACTTCCCGGCTCTGGCCCGGCTCCAGC
AGGACGCTGTTGGGATCGTCATGTTTCATCGAGTGGCCGTTGCCCATGTCCATGTTCATCATGTCATGATTCAACTTGTTACCCTG
GATCACGCCATGCTCGACCATCATTCTCATTTCTTTCTGGTGGGCCTCATGCATACCCGGAGTCCCGATGTTGAACTCGTGCACAA
GATTGCCCTTGTTCTGCACCACAAACCGAACGGTTTCGCCCGGCTTCACACGGATTTCTTCGGGTTCGTAGTAGTTGTCGTGCATT
TCAACCGTTATGGTCCGGCTGGCTTCTGAGGCTTTTCCCGGTTCGCCAATGGAAGCACCATGGCCGTGACCACCGCCATGGGTGCC
AGCCGCAAACGCCGATACCGACATCGAAAAAGCCAGGCCGGCCACAATCAGGTTTGATGCATTCATCCGTTTGTCTCCTGTGGAAT
AATGGTTTTGAGAACACGGCTGATTCGTACTGAATCTTGAGCAATGTCCAGAATCAGTCGACACTTTCTCTCACCAACCTGAATTC
TTTCTGAAAAACGGGCACTTTCTTGCGTTCAGGCACAATTCAGGTTGATCGACGACACTGGCGGCTGAAACAAGCACCGTTAAGGA
CGGATACATGCGTTTATTGCTCGTTGAAGACGACCGTTTGCTGGCCGAGGGGCTGGTCAGACAGCTGGAAAAAGCGGGGTTTAGCA
TCGACCACACGTCCAGTGCCCGTGAGGCCCAGATTCTGGGGGAGCAGGAGGATTATCGTGCCGCTGTTCTCGATCTCGGCCTGCCG
GATGGCAACGGGCTGGAGGTTTTGAAACGATGGCGATCGAAGCTCGTCGAGTGCCCGGTATTGGTCCTCACCGCCAGAGGCGACTG
GCAGGACAAGGTCAATGGACTGAAAGCAGGGGCGGATGACTATCTGGCCAAACCGTTCCAGACAGAAGAACTGATCGCCCGCATCA
ATGCACTCATACGCCGCAGCGAAGGGCGAGTGCACTCCCAGGTTAAAGCCGGTGGCTTCGAACTGGACGAAAATCGCCAGAGCCTG
CGGACAGAAGAAGGAGCAGAACACGCCCTGACGGGTACTGAGTTCCGCCTGCTACGATGCCTGATGAGCCGTCCGGGCCACATCTT
TTCCAAGGAACAGCTAATGGAGCAGCTATACAACCTGGATGAGAGCCCCAGCGAAAACGTGATTGAGGCGTATATTCGGCGCTTGA
GAAAGCTGGTCGGCAACGAAACGATCACCACACGCCGTGGCCAGGGATACATGTTCAATGCCCAACGTTAAAAGGCCAGCATCCGT
CAAAGGCATGCTGCTGGTGTTATTGCTGCCCGCCGGCATCGCTCTGATGGGCGTAGCCTGGTTTGTCCACGGCCTGCTGCTGGATC
GAATGTCCCGGGAATTCCTTGAAAGCCGCCTTAAAGACGAAGCTGCCTTTCTGGAGCACCAGATTCGCGAGGTCAGAGGTCAAGTC
GAAACCCTGCAGACCGGTGATTACTTCCAGGACGTCTTCCATCATGCCTTCGCCATACGCACGCCCGATCGAACCATCATATCGCC
AAAGGCCTGGGAACCCTTGCTGGCACCACTGATCAACCATGAGCAGAATGGCACGCTTCGTCTTGAAGGCCGTCAGGCGCCGGACA
GTCCGTCCGATATCCTGGCATACCGTCACTCTTTTCAGGTGAACGGGTCACCGATCGTCGTGGTTGTATCCGAGGATCTGGAGGCC
CTGAAACGCAGTCAGGCCGAGCTGCACGCCTGGACAGCCGTGGTGTCGGTGCTGTTGATCGCACTTCTGGTTGCCGTCATCTGGTT
TGGCATCACGCTATCGCTGCGGCCGGTGGTGACACTGAAAGCGGCGTTAAAACGATTGCAGGATGGAGAGATTTCCCGGATCAATG
CACCATCACCCGAAGAATTTCAACCGCTCGTGATGCAGTTAAACCAGTTGCTTGACTCCCTCGACAAGCGACTGGTGCGTTCCCGG
GATGCGCTCGCAAATCTGTCACACAGTGTCAAAACGCCCATCGCAGCCGTCCGGCAGATACTGGAGGATATGGATCGCCCATTGCC
TAGTGATCTAAGAATCCAAATGGCCGCCCGTCTCAGTGACATCGACAGACAATTGGAAGCGGAAATGCGCCGCAGCCGCTTTGCCG

GGCCCCAGGTCGGGAAAAGCGCTTATCCCGTCAAACAGGCGCGGGATCTTTTGTGGATGCTGGGGCGGCTGTATCCGGAAAAATCC
TTCGAACTATCAAGCTCACTGCCGGAAGACACCCGCTGGCCGATAGAGGAGCATGACTTGAATGAAGTGCTAGGCAACCTGCTCGA
TAATGCCGGCAAATGGTCGTCGCGGTGCGTAGAACTCTCGCTGAAACAAGACAACAACAGCAGACAAATCGTTGTTTCTGATGATG
GTCCTGGAGTCAATGGCGACGACTTGTCCAGTCTGGGGCAACGAGGGCTGCGACTTGACGAGCAGACCCCTGGCCACGGTCTGGGC
CTCGCGATTGTCCGGGAGATCGTCGCTCGCTATGAGGGCAACATTAGTTTTTCGACGGGGCCCGGTAGCGGCTTGCGCGTAACCAT
AGAGTTTTAA

C.11.3 OmpR

The OmpR response regulator binds selectively to several promoter regions. These were characterized by Martínez-Hackert, E. *et al.* in 1997 [65]. From mutation assays and analysis of five of these sequences, identified the consensus sequence for this transcription factor as : nTnnCnnnnn-nnnACnnnnn in which the dash separates the tandem tem base pair recognition sequence that comprises the twenty base pair site bound by two OmpR molecules.

C.12. Protein Purification CopR – Strategy 1

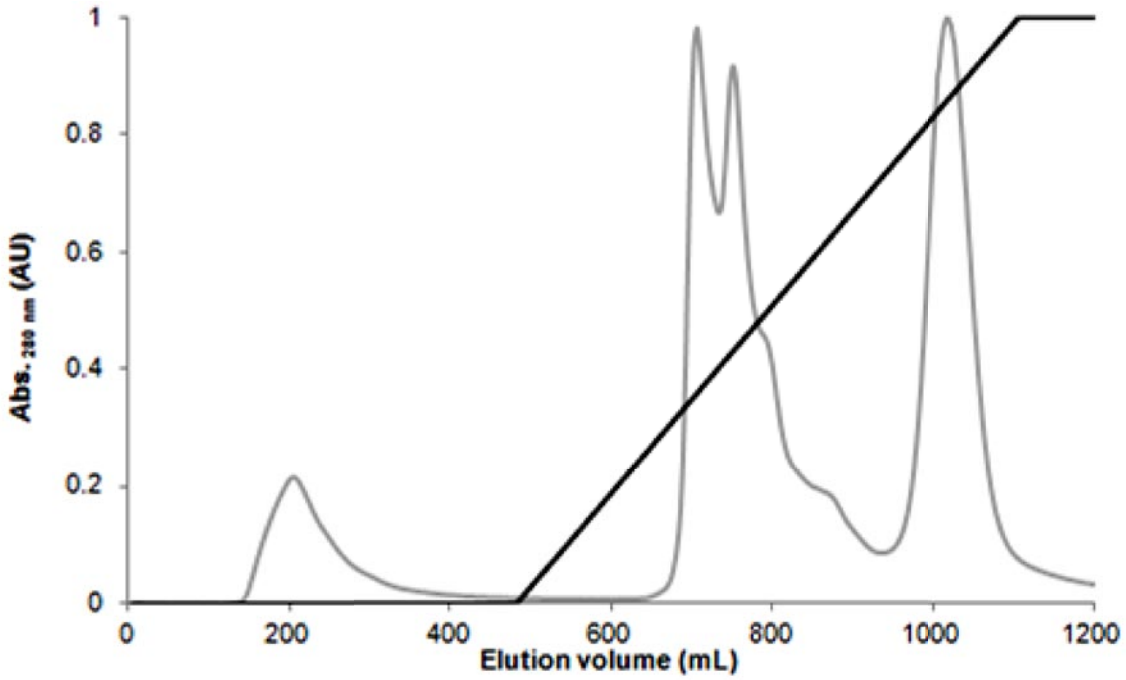


Figure 6.15 Elution profile of the CopR protein as obtained from the soluble cell extract in a DEAE-52 column. The protein was eluted in a 20 mM Tris-HCl (pH 7.6) buffer with a gradient of 0 to 150 mM NaCl (represented by the black curve).

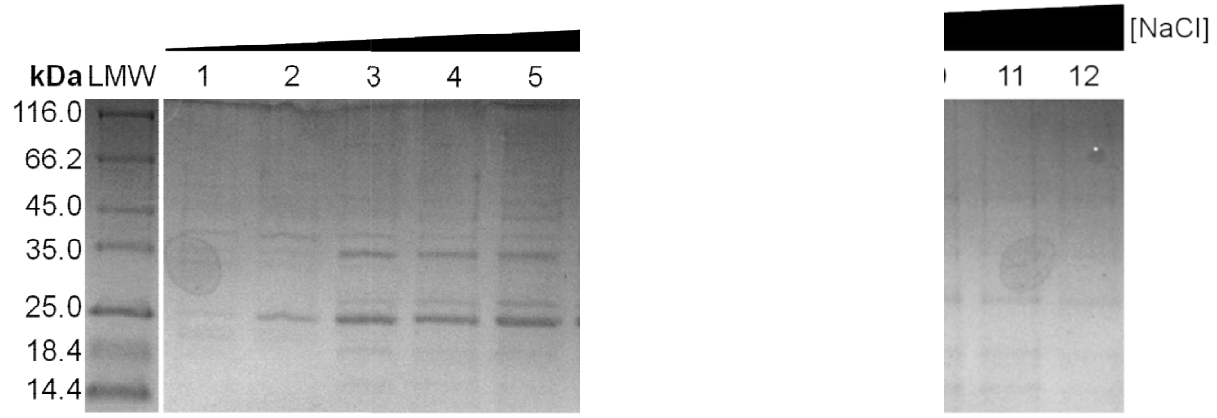


Figure 6.16 SDS-PAGE analysis of the fractions obtained from injection of the CopR protein into a DEAE52 column. **LMW:** Low molecular weight marker; **Lanes 1 to 12:** Samples of fractions obtained after elution with 5.9, 6.4, 6.8, 7.9, 8.4, 8.8, 9.3, 10.4, 19.6, 20.0, 20.5, 20.9, 21.4% elution buffer 20 mM Tris-HCl (pH 7.6) + 500 mM NaCl.

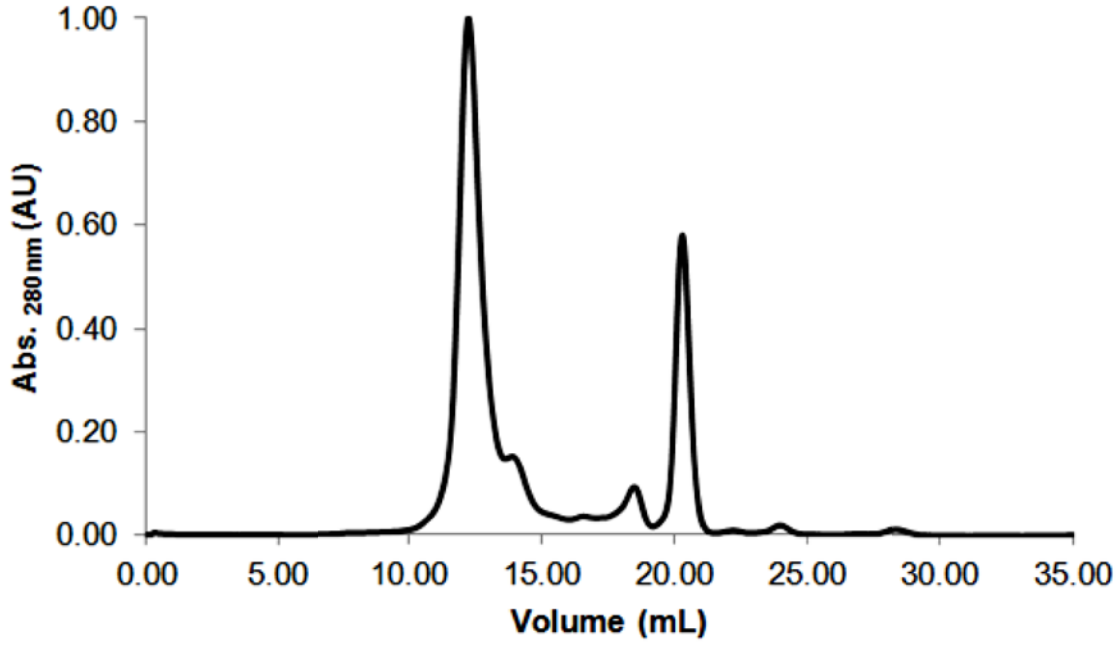


Figure 6.17 Elution profile of the CopR protein as obtained from the DEAE-52 column, on a Superdex 75 10/300 GL. The protein was eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl.

The SDS-PAGE gel performed for the identification of the CopR protein in the fractions obtained from the purification represented in Figure 6.16 was inconclusive as the protein concentration was too small to be detected. The fractions eluted at the elution volume of around 12.12 mL were concentrated.

C.13 Molecular Mass Determination

C.13.1 CopRHis$_6$

This protein construct contains, in addition to the 25.1 kDa associated with the CopR protein, an additional 1.0 kDa which encodes for the HRV3C cleavage site and 4.8 kDa that is associated with the cloning method and strategy which includes a His$_6$ tag to the N-terminus of the protein of choice. So, in total the CopRHis$_6$ protein should, upon expression, present with approximately 31 kDa, Figure 6.19.

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKMLEVLFQGPMRLLLVEDDRLLAEGLVRQLEKAGFSIDHTSSAR
EAQILGEQEDYRAAVLDLGLPDGNGLEVLKRWRSKLVECPVLVLTARGDWQDKVNGLKAGADDYLAKPFQTEELIARINALIRRSE
GRVHSQVKAGGFELDENRQSLRTEEGAEHALTGTEFRLLRCLMSRPGHIFSKEQLMEQLYNLDESPSENVIEAYIRRLRKLVGNET
ITTRRGQGYMFNAQR

Figure 6.18 Protein sequence of the CopRHis$_6$ protein. **Legend**: In *italic* is the 4.8 kDa fragment which contains the His$_6$-tag. Underlined is the HRV3C protease recognition sequence, followed by the full-length CopR protein of 25.1 kDa. The CopRHis$_6$, upon expression, should have a molecular mass of 31 kDa.

C.13.2 CopS_CHis$_6$

This protein construct, obtained from cells transformed with the pCopS_C1 vector contains, in addition to the 27.5 kDa associated with the CopS_C protein, an additional 4.8 kDa that is associated with the cloning method and strategy which includes a His$_6$ tag to the N-terminus of the protein of choice. So, in total the CopS_CHis$_6$ protein should, upon expression, present with approximately 32 kDa, Figure 6.20.

MHHHHHHSSGLVPRGSGMKETAAAKFERQHMDSPDLGTDDDDKAALKRLQDGEISRINAPSPEEFQPLVMQLNQLLDSLDKRLVRS
RDALANLSHSVKTPIAAVRQILEDMDRPLPSDLRIQMAARLSDIDRQLEAEMRRSRFAGPQVGKSAYPVKQARDLLWMLGRLYPEK
SFELSSSLPEDTRWPIEEHDLNEVLGNLLDNAGKWSSRCVELSLKQDNNSRQIVVSDDGPGVNGDDLSSLGQRGLRLDEQTPGHGL
GLAIVREIVARYEGNISFSTGPGSGLRVTIEF

Figure 6.19 Protein sequence of the CopS_CHis$_6$ protein. **Legend**: In *italic* is the 4.8 kDa fragment which contains the His$_6$-tag, followed by the C-terminal domain of CopS. The CopS_CHis$_6$, upon expression, should have a molecular mass of 32 kDa.

C.13.3 CopS_CHis$_6$GST

This protein construct, obtained from cells transformed with the pCopS_C2 vector contains, in addition to the 27.5 kDa associated with the CopS_C protein, 8 kDa that is associated with the cloning method and 25.2 kDa associated with the strategy which includes a GST-tag to the N-terminus of the protein of choice. So, in total the CopS_CHis$_6$GST protein should, upon expression, present with approximately 61 kDa, Figure 6.21.

MSPSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMAIIRYIADKHNMLGG
CPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCL
DAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPGSTSGSGHHHHHHSAGLVPRGSTAIGMKETAAAKFERQH
MDSPDLGTGGGSGDDDDKMLEVLFQGPM**LEVLFQGP****AALKRLQDGEISRINAPSPEEFQPLVMQLNQLLDSLDKRLVRSRDALANL***
SHSVKTPIAAVRQILEDMDRPLPSDLRIQMAARLSDIDRQLEAEMRRSRFAGPQVGKSAYPVKQARDLLWMLGRLYPEKSFELSSS
LPEDTRWPIEEHDLNEVLGNLLDNAGKWSSRCVELSLKQDNNSRQIVVSDDGPGVNGDDLSSLGQRGLRLDEQTPGHGLGLAIVRE
IVARYEGNISFSTGPGSGLRVTIEF

Figure 6.20 Protein sequence of the CopS_CHis$_6$GST protein. Legend: In italic is the 33 kDa fragment which contains the tags inserted by the cloning methodology employed. Underlined is the HRV3C protease recognition sequence (0.9 kDa), followed by the C-terminal domain of CopS. The CopS_CHis$_6$GST, upon expression, should have a molecular mass of 61 kDa.

C.13.4 Elution volumes

Table 6.11 Elution volumes of the CopRHis$_6$ protein when the protein is injected into a Superdex 75 10/300 GL filtration column. The elution volumes are from the profiles in Figure 3.38. CopRHis$_6$-P corresponds to the phosphorylated protein.

Sample	Elution volume (mL)	Molecular mass (kDa)
Albumin	9.09	66.00
Ovalbumin	9.93	44.00
Chymotrypsinogen	11.96	25.00
Ribonuclease A	12.96	13.70
CopRHis$_6$	9.09	65.30
	10.49	38.18
CopRHis$_6$-P	9.09	65.30
	10.45	38.77

Table 6.12 Elution volumes of the CopR protein when the protein is injected into a Superdex 75 10/300 GL filtration column. The elution volumes are from the profiles in Figure 3.39. CopR-P corresponds to the phosphorylated protein.

Sample	Elution volume (mL)	Molecular mass (kDa)
CopR	11.25	28.45
CopR-P	11.25	28.45

Table 6.13 Elution volumes of the CopR protein in the presence and absence of DNA, when these are injected into a Superdex 75 10/300 GL filtration column. The elution volumes are from the profiles in Figure 3.44.

Sample	Elution volume (mL)	Molecular mass (kDa)
CopR-P	11.24	29.56
DNA	7.85	-
5CopR P:DNA $_{260nm}$	7.49	-
	11.24	29.56
5CopR P:DNA $_{280nm}$	7.53	-
	11.32	28.69

Table 6.14 Elution volumes of the CopR protein in the presence and absence of DNA and BSA, when these are injected into a Superdex 75 10/300 GL filtration column. The elution volumes are from the profiles in Figure 3.45.

Sample	Elution volume (mL)	Molecular mass (kDa)
CopR-P	11.40	27.84
BSA	9.13	66.0
DNA	7.85	-

Table 6.15 Elution volumes of the CopR_N and CopR_C protein when the protein is injected into a Superdex 75 10/300 GL filtration column. The elution volumes are from the profiles in Figure 3.67.

Sample	Elution volume (mL)	Molecular mass (kDa)
Albumin	9.05	66.00
Ovalbumin	9.93	44.00
Carbonic Anhydrase	11.25	29.00
Aprotinin	15.23	6.50
CopR_N	12.55	17.49
CopR_C	9.68	15.09
	12.95	50.69

Table 6.16 Elution volumes of the CopS_C protein when the protein is injected into a Superdex 75 10/300 GL filtration column. The elution volumes are from the profiles in Figure 3.89.

Sample	Elution volume (mL)	Molecular mass (kDa)
Albumin	9.05	66.00
Ovalbumin	9.93	44.00
Carbonic Anhydrase	11.25	29.00
Aprotinin	15.23	6.50
CopS_C	10.45	38.12
	11.68	24.21
	12.14	20.39
CopS_CGST	10.44	38.31
	11.63	24.59

C.13.4 Elution volumes – Influence of pH

Table 6.17 Elution volumes of the CopR protein when the protein is injected into a Superdex 75 10/300 GL filtration column in phosphorylation inducing conditions. The pH indicates the pH of the kinase buffer which was the only variable in this assay.

pH	Elution volume	Molecular mass
7.5	11.44	28
7.6	11.57	26
8	11.22	30
8.5	11.22	30
9	11.24	30

C.13.5 DNA Elution Profile

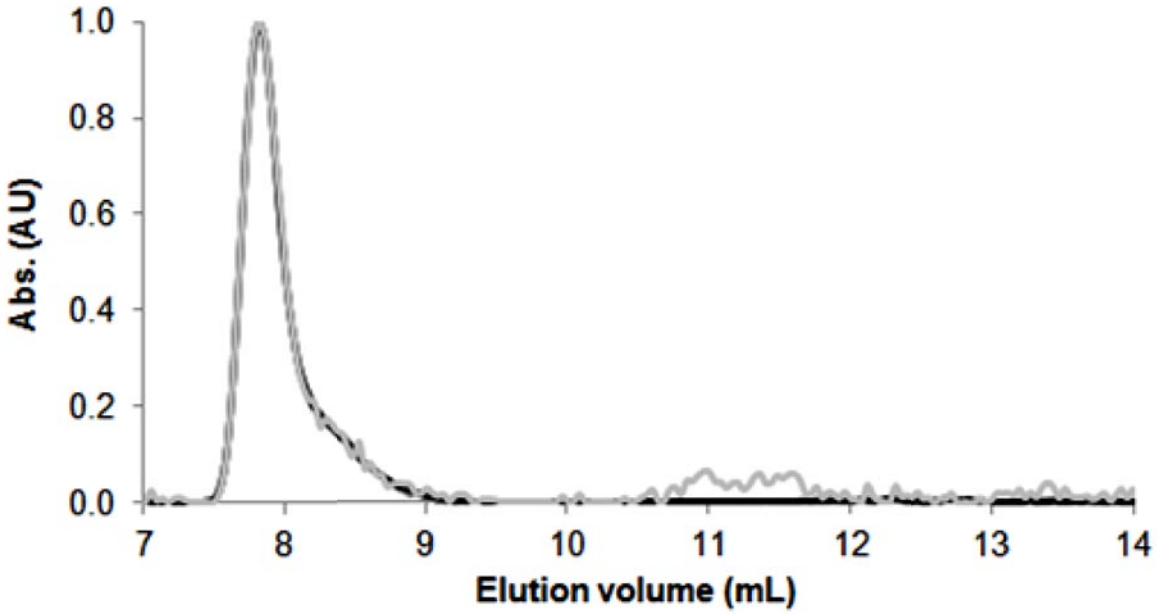


Figure 6.21 Elution profile of the *pro*DNA fragment in a Superdex 75 10/300 GL. The biomolecule was eluted with 50 mM Tris-HCl (pH 7.6), 150 mM NaCl. The profiles correspond to the same injection, measured at 215 nm (grey line) and 260 nm (black curve).

"A stumble may prevent a fall."

English Proverb