Biochemical basis of organohalide degradation:

Characterisation of a self-sufficient reductive dehalogenase from marine aerobic Proteobacteria

A thesis submitted to The University of Manchester
for the degree of Doctor in Philosophy (Ph.D.)
in the Faculty of Science and Engineering.

Department of Chemistry

School of Natural Sciences

M.Sc. Samantha A. Gaytán Mondragón

2020

Contents

Abstract

Organohalide respiring bacteria (OHRBs) can use halogenated compounds as terminal electron acceptors, leading to organohalide reduction and degradation. Organohalides are naturally produced as part of the halogen biogeochemical cycles, but anthropogenically introduced compounds are now considered major environmental pollutants. All OHRBs rely on reductive dehalogenases (RdhAs), a subfamily of B_{12}-dependent enzymes, that catalyse the final step of the organohalide respiration. However, RdhA enzymes are not exclusively found in anaerobic OHRBs, and a subset of RdhAs are part of catabolic pathways in marine Proteobacteria. These enzymes catalyse the reduction of organohalides to allow the complete degradation or reutilisation of the carbon backbone. A complete understanding of the structure, function, and regulation of the RdhAs is desirable for the application of reductive dehalogenation in the bioremediation of anaerobic polluted sites.

Here we present the initial biochemical characterisation of a catabolic RdhA that consists of a natural fusion between the dehalogenase B_{12}/Fe-S domains and an additional C-terminal reductase domain, similar to the iron-sulfur (Fe-S) flavoprotein phthalate dioxygenase reductase (PDR). The RdhA-PDR fusion renders the enzyme self-sufficient in terms of coupling NAD(P)H oxidation to organohalide reduction and are thus referred to as self-sufficient RdhAs (ssRdhAs). To achieve functional heterologous expression of ssRdhAs, we used the xylose-inducible *B. megaterium* and the standard *E. coli* system optimised with a B_{12}-uptake system (BtuB). Although the cofactor incorporation and protein yield remain a problem despite extensive screening of a range of conditions, we could obtain *in vivo* and *in vitro* activity with *ortho*-halogenated phenols. Furthermore, we confirmed that the PDR-like domain allows the intramolecular transfer of electrons from NAD(P)H to the active site of cobalamin without the need for an external reductase system to provide electrons. Given the initial challenges of producing soluble and active full-length enzymes, we also expressed the isolated PDR-like reductase domain and demonstrated it can support dehalogenation by channelling electrons to the dehalogenase active site when assayed in a one-pot reaction. Additionally, to test these results, we designed and produced an artificial fusion enzyme linking the PDR-like reductase domain with the previously characterised non-self-sufficient NpRdhA. Unfortunately, no consistent dehalogenation of the substrate 3,5-Bromo-4-hydroxybenzoic acid was obtained.

To address the regulation of the RdhAs, we completed the structural and functional characterisation of the MarR-type transcriptional regulator $RdhR_{Cbdb1625}$, previously studied in terms of ligand specificity towards dichlorinated phenols. In this work we present the crystal structure of $RdhR_{Cbdb1625}$, in complex with a 1,2,3-tricholorophenol, providing a rationale for the distinctive preference for specific halogenation substitution patterns. We also demonstrate the effect of the tight-binding ligands on the $RdhR_{Cbdb1625}$ DNA-binding affinity, confirming its putative role as a transcriptional repressor.

Declaration

A portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this University. The Chapter Three, Results Chapter I, "Crystal structures of $RdhR_{CbdbA1625}$ provide insight into sensing of chloroaromatic compounds by MarR-type regulators" presented in this thesis is a product of joint authorship with Carolina P. Quezada, who developed the groundwork and obtained relevant crystallographic data included in this work.

Copyright statement

Acknowledgments

I first want to thank Professor David Leys for allowing me to work in such a prestigious institution as the MIB and the University of Manchester. I want to express my deepest gratitude to my advisors Professors Jonathan Waltho and Sam Hay, and to my examiners Professors Eriko Takano and Gideon Grogan, for giving me an opportunity to make amends with my thesis.

Many thanks also to all the members of the Leys and the Enzymology Groups. Special thanks to Drs. Irina Gostismkaya, Godwin Aleku, Stephen Marshall, Deepankar Galoth, Hanan Latif Messiah, Karl Payne, Mark Dunstan, and the fabulous technicians Marina Golovanova and Michiyo Sakuma. And to those that are no longer in the MIB: Ziga Zebeç, Aitor Hernández, Bat-Erdene Judger, Roger Kutta, Nataliya Archipowa, Anna Peters, Lumi Gabi, and Leopoldo Machado, for giving me an example of research quality and camaraderie. I am also thankful to all the former/present students: Iaroslav Kosov, Arune Balaikaite, Annica Saaret, Gabriel Titchiner, Sam Bailey, Sam Bloor, Fraser Collins, and Carolina P. Quezada, for sharing with me the burden of the dehalogenase project! It has been a privilege to work with you all! Recognition as well to the MIB Media Kitchen for all their assistance.

This thesis is dedicated to my Mexican friends, who encouraged me to keep going and helped me with their knowledge, recommendations, and creativity, you are the best bunch of nerds that I would ever meet! I love you all!

Last but not least, I want to express all my gratitude to my handsome and intelligent super-husband David Apodaca, to my most excellent dogs Cajeta and Mermelada, and the thug-cat Toribio. Without you, nothing in my life would make sense.

Gracias a toda mi familia en México, particularmente a mi madre, padre y a mi querido hermano Javier.

"Ever tried. Ever failed. No matter. Try again. Fail again. Fail better."- S. Beckett.

❁ Note: This thesis was supported by CONACyT, Mexico (CVU 378154).

Preface to alternative format

This thesis is being presented in the alternative format in accordance with the rules and regulations of the University of Manchester. The results chapters presented herein are in manuscript form in the style suitable for their intended journal of submission. However, elements have been reformatted to ensure these form a cohesive body of work. Manuscripts included as results chapters are:

Chapter Three. Results Chapter I. Crystal structures of RdhR$_{CbdbA1625}$ provide insight into sensing of chloroaromatic compounds by MarR-type regulators.

Samantha Gaytán and Carolina P. Quezada[2], Thomas Jowitt, Mark Dunstan, Sam Hay, David Leys*.

Manuscript in preparation for submission to the Journal of Biological Chemistry (JBC).

Author contributions: Samantha Gaytán and Carolina P. Quezada contributed equally to this publication. Mark Dunstan and Carolina P. Quezada cloned, expressed, and purified the His-tag-RdhR$_{CbdbA1625}$ and obtained the first crystal structures and initial binding data. Carolina P. Quezada and Mark Dunstan developed the groundwork with the His-tag-RdhR$_{CbdbA1625}$ regarding the binding-assays (not included in this work). Samantha Gaytán cloned, express, and purified the Steptag-RdhR$_{CbdbA1625}$, performed the ligand and DNA binding assays by fluorescence and ITC, EMSAs and solved the Strep-RdhR$_{CbdbA1625}$ crystals structures. Thomas Jowitt performed AUC and ITC assays. Sam Hay assisted with the interpretation of the fluorescence quenching binding studies. Samantha Gaytán, Carolina P. Quezada, and David Leys wrote the manuscript. David Leys designed and directed the research leading to this paper.

Chapter Four. Results Chapter II. Heterologous expression of a self-sufficient catabolic reductive dehalogenase from marine Proteobacteria.

Samantha Gaytán, Godwin Aleku, Karl Payne, Irina Gostimskaya, Muralidharan Shanmugan, and David Leys*.

Manuscript written as a communication not currently submitted to any journal.

Author contributions:

Samantha Gaytán (S.G) and Irina Gostimskaya cloned the ssRdhA constructs. S.G and Karl Payne transformed *B. megaterium*. S.G expressed and purified the ssRdhAs, performed metal-cofactor determination, UV/Visible spectroscopy, and prepared the samples for Raman spectroscopy, and EPR. S.G and Godwin Aleku measured the activity of the enzyme. Muralidharan Shanmugan performed EPR studies and assisted with the interpretation. S.G. performed the *in silico* analysis presented in this work and the phylogenetics. S.G. wrote the manuscript. David Leys designed and directed the research.

Chapter Five. Results Chapter III. Designing an artificial self-sufficient catabolic reductive dehalogenase.

Samantha Gaytán, Karl Payne, Godwin Aleku, and David Leys*.

Manuscript not currently submitted to any journal.

Author contributions:

Karl Payne cloned the artificial ssRdhA construct and Samantha Gaytán (S.G) the individual domains. Karl Payne transformed *B. megaterium*. S.G expressed and purified the artificial RdhAs and individual domains, performed metal-cofactor determination, UV/Visible spectroscopy, and prepared the samples for EPR. S.G measured the activity of the enzyme. S.G. performed the *in silico* analysis presented in this work. S.G. wrote the manuscript. David Leys designed and directed the research.

List of figures and tables

Abbreviations

Å	Angstrom
aa	Amino acid residue
Ado-Cbl	5′-deoxy-5′-adenosylcobalamin
ADP	Adenosine-5'-diphosphate
Arg	Arginine (R)
Ar-X	Aromatic halogenated compound
Asn	Asparagine (N)
Asp	Aspartic acid (D)
ATP	Adenosine triphosphate
BhbA	Reductive dehalogenase of *Comamonas* sp. 7D-2
bp	Base pairs
BtuB	B_{12} uptake system
Cbl-C	Cobalamin decyanation/dealkylations enzymes
C-C	Carbon-carbon bond
CFC	Dichloro(difluoro)methane
C-H	Carbon-hydrogen bond
CN-Cbl	Cyanocobalamin (Vitamin B_{12})
Co-C	Cobalt-carbon bond
Cpr-Fnr	Cpr-Fnr transcriptional family factor
Cryo-EM	Cryo-Electron Microscopy
C-X	Carbon-halogen bond
Cys	Cystein (C)
Da	Dalton
DDT	Dichlorodiphenyltrichloroethane (isomers)
DFT	Density Functional Theory
D_o	Bond-dissociation energy
dsDNA	Double-stranded DNA
DT	Dithionite (sodium)
DTT	1,4-dithiothreitol
E^o	Redox potential (mV)
EPR	Electron Paramagnetic Resonance
E-S	Enzyme-Substrate complex
Fd	Ferredoxin
Fe-S	Iron-sulfur clusters
FMN	Flavin mononucleotide, Riboflavin-5′-phosphate
FNR	Ferredoxin $NADP^+$ reductase
Ga	Giga-annum (one billion years)
Gln	Glutamine (Q)
His	Histidine (H)
HOMO	Higher Occupied Molecular Orbital
HupL	Hydrogenases of the respiratory complex
LUMO	Last Unoccupied Molecular Orbital

Lys	Lysine (K)
MarR	Multiple-Antibiotic Resistance Regulator
Mbp	Mega base pairs
MD	Molecular Dynamics
Me-Cbl	Methylcobalamin
mM	Millimolar concentration
MMACHC	Methylmalonic aciduria and homocystinuria type C protein
NADH	Nicotinamide adenine dinucleotide
NADPH	Dihydronicotinamide-adenine dinucleotide phosphate
NpRdhA	Reductive dehalogenase from *Nitratireductor pacificus*
·OH	Hydroxyl group
OHRB	Organohalide Respiring Bacteria
OmeA	Fe-S-Mo dependent oxidoreductases
PCA	Perchloroethane (hexachloroethane)
PCB	Polychlorinated biphenyls
PCDF	Perchlorodibenzofuran (1,2,3,4,6,7,8,9-octachlorodibenzofuran)
PCE	Perchloroethylene (tetrachloroethylene)
PceA	Reductive dehalogenase from *Sulfospirillum multivorans*
PCP	Pentachlorophenol (2,3,4,5,6-pentachlorophenol)
PDB	Protein Data Bank
PDR	Flavoprotein Phthalate Dioxygenase Reductase
Phe	Phenylalanine (F)
Pi	Inorganic phosphate
POP	Persistent Organic Pollutants
QM/MM	Hybrid quantum mechanics / Molecular mechanics
QueG	Epoxyqueuosine reductases
r	Atomic radius
R-/R=	Organic moiety with a single or double bond
RBS	Ribosome Binding Site
RDases	Characterised reductive dehalogenases
RdhA	Reductive dehalogenases
rdhA	Reductive dehalogenase catalytic subunit gene
RdhB	Membrane anchor protein
rdhB	Reductive dehalogenase putative membrane-anchor gene
RNA	Ribonucleic acid
Ser	Serine (S)
SlyA	Transcription factor of *Enterococcus faecalis*
SN2	Nucleophilic substitution reaction
SQ	Flavin semiquinone
SSD	Spin-dependent delocalisation
ssRdhA/RdhA-PDR	Self-sufficient reductive dehalogenase, natural or artificial
TAT	Twin Arginine Translocation signal peptide
TCB	Trichlorobenzene
TCDD	Tetrachlorodibenzodioxin (2,3,7,8-tetrachlorodibenzo-p-dioxin)
TCE	Trichloroethane (1,1,1-trichloroethane)

TCM	Trichloromethane
Trp	Triptophan (W)
Tyr	Tyrosine (Y)
VcrA	Vinylchloride reductive dehalogenase
WingedHtH	Winged helix-turn-helix
X	Halogen atom
ΔG_f^o	Gibbs free energy of formation
ΔH	Change of Enthalpy

1.0 Introduction

1.1 Organohalide production and biogeochemical cycles

Aliphatic halogenated organic compounds are generated by the substitution of at least one C-H bond in order to generate one or more C-X[1] bonds in which the carbon atom has sp^3 hybridisation in case of the alkanes, sp^2 for the alkenes and sp for the alkynes. In aryl halides, the carbon atom has sp^2 hybridisation and is part of an aromatic system defined by the conjugation of the delocalised electronic density. Organohalide compounds have been used for decades in a wide and still growing range of industrial, agricultural, home, and medical applications as antiseptic, antihistaminic, analgesic, and antineoplastic drugs. In agriculture, organohalides are used as pesticides while their industrial applications include the use as synthetic precursors, solvents, additives, refrigerants, propellants, plastics, dry cleaning fluids, fire retardants and others **(Figure 1)** [1].

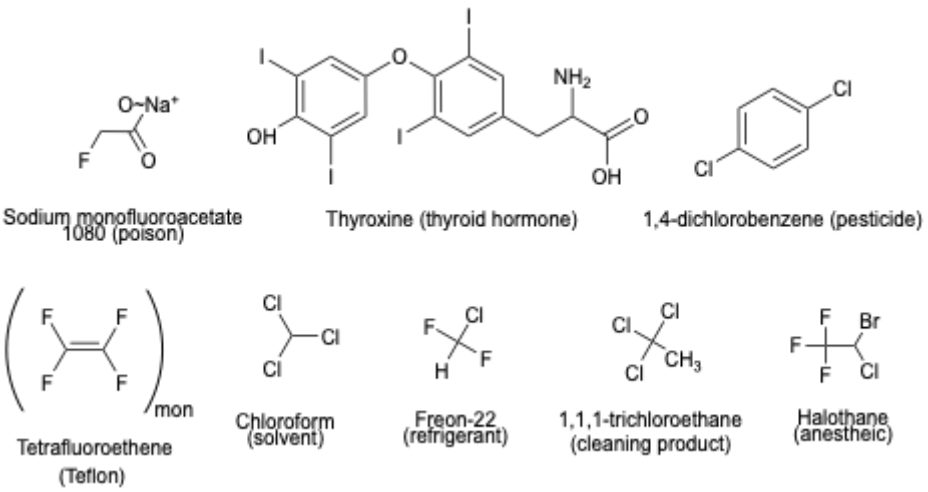


Figure 1. Common organohalides and their applications.

Organohalides are often described as xenobiotic and recalcitrant compounds given their persistence against chemical degradation, their low water solubility, and because they were thought to be un-natural compounds [2]. However, the synthesis and production of organohalides has not only been observed as a result of industrial activities but also as a

[1] Generally, X is used to refer to any halogen of group 17 or VIIA in the periodic table: fluoride (F), chloride (Cl), bromide (Br) and iodide (I). Astatine (At) is not considered due to its low abundance on Earth.

result of biogeochemical processes. The production of simple to structurally complex organohalides has been reported as a result of abiotic processes, like volcanic activity and forest fires **[3]**, and of biotic ones in marine, lentic and lotic environments as part of biosynthetic pathways **[1, 4-6]**.

Typical reactions for synthesising aromatic halogenated compounds (Ar-X) in a laboratory environment consist of the halogenation of aromatic rings with Lewis acid catalysts or by derivatisation of diazonium salts. Ar-X can undergo conventional nucleophilic substitutions in a similar manner to the aliphatic halides but in rather aggressive conditions of pressure and temperature. The C-X bond is strengthened and further stabilised because of the conjugation between the delocalised π electrons of the ring and the halogen lone electron pairs. Highly substituted aryl halides are susceptible to aromatic nucleophilic substitution and to the formation of Grignard reactants to carry out an *umpolung*[2] of the carbon atom bound to the halogen **[7]**. Elimination of the halogen substituents can occur when the X group is replaced by other nucleophiles or electrophiles **[8-10]**. The most common reactions for the industrial synthesis of organohalides are catalytic cracking of hydrocarbons in presence of the halogen or its corresponding binary acid (hydracid) or by nucleophilic substitution from alcohols **[11]**

Halogenated compounds are excellent electron acceptors, given the Gibbs free energy values for the reductive dehalogenation reaction by hydrogenolysis (-130 and -180 $kJmol^{-1}$) **[8]**. The possibility of some organohalides to undergo redox reactions depends on their chemical and energetic properties such as the Gibbs free energy of formation ΔG_f^o ($kJmol^{-1}$), redox potential E^o (mV) or ionisation potentials (eV) (defined as the energy gap between HOMO and LUMO molecular orbitals) and C-X bond strength and reactivity **[4]**. These in turn greatly depends on the nature of the halogen attached to the carbon atom, with iodide compounds being more reactive as the C-X bond weakens when the halogen atom becomes more polarisable, as shown in **Table 1** **[11, 12]**.

Bond	D_o	r
	$kJmol^{-1}$	*Å*
C-C	345.6	1.54
C-F	485	1.35
C-Cl	327.2	1.77
C-Br	285	1.94
C-I	213	2.14
C-O	357.7	1.43

[2] Polarity inversion in organic chemistry, defined by the chemical modification of a functional group.

C-H	411	1.20
C-S	272	1.82
C-P	264	1.84

Table 1. **Relevant C-X bond energies and lengths** [12]. Bond strength diminishes the more polarisable the halogen atom is.

Chlorine is one of the most abundant elements on Earth, whether as an anion in solution, a component of inorganic salts, or as part of a huge number of organic molecules produced by living organisms [13]. The combination of these abiotic and biotic processes involving the transport and chemical transformations of this element in the lithosphere, biosphere, and atmosphere has been named the chlorine cycle [13], illustrated in **Figure 2.** With the exception of astatine, the chlorine cycle can be extended to the rest of the halogens.

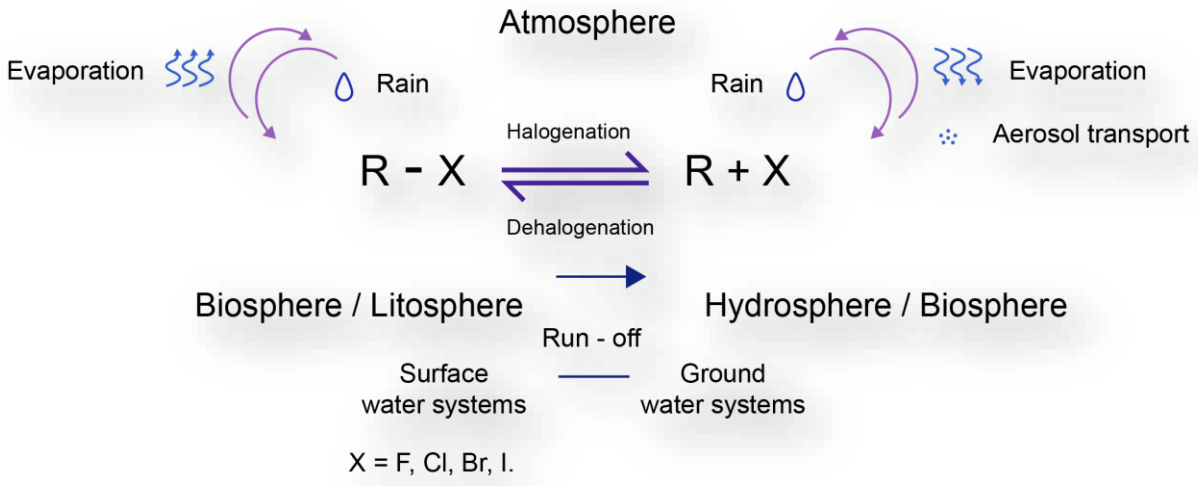


Figure 2. The chlorine or halogen cycle. Halogenation and dehalogenation are represented in equilibria due to the combination of the abiotic and biotic processes that allow global cycling through the Earth. Halogen ions (X^-) are moved between the atmosphere, lithosphere, hydrosphere, and biosphere (where organisms incorporate the most common halogens to natural products and where the anthropogenic production of organohalides takes place) [13].

Particularly in marine habitats, some organisms possess pathways that involve enzymes catalysing the halogenation of secondary metabolites from reactive halogen species derived from inorganic salts present in the ocean. Halogen-carbon metabolism is distributed throughout the three Domains of Life, and although it is much more common in prokaryotic organisms, it is also present in eukaryotes, particularly in protists, fungi, plants

but less frequently in animals [14, 15][3]. This, in turn, means that the metabolic pathways and enzymes that catalyse the incorporation or removal of X substituents from organic molecules are remarkably diverse in terms of the mechanistic strategies involved, that include reduction-oxidation reactions, nucleophilic and electrophilic substitutions, regioselective, and stereoselective functionalisation of C-H bonds, chain elongation, and cyclisations, where the organic scaffold and the halogen group determine the nature of the chemical transformation [16].

It is known that marine bacteria, algae, and fungi living in specific ecological niches as mangroves, seagrass, and coral reefs [17] produce most of the chlorinated and brominated metabolites known today [18]. These organohalides are produced for various purposes: as hormones, pheromones or to serve in the chemical warfare as defense mechanisms [3]. Interestingly, the bromination is far more common in aquatic ecosystems, even though the chlorine's concentration in the seawater is higher than that of bromine (approximately 500 mM to 0.9 mM) [15]. Halogenated natural products range from simply substituted methanes to some of the most synthetically complex structures, as the antibiotics vancomycin and balhymicin. Few iodinated natural products exist, possibly due to the low abundance of free iodide, but some macroalgae, sponges (Phylum Porifera), and terrestrial symbiotic associations like lichens are known to synthesise them [16, 19]. Despite the fact that enzymatic fluorination reactions are rare and thermodynamically expensive, given the electronegativity and energy barrier for the desolvation of F^-, organofluorinated natural products have been identified [20, 21] in marine and soil bacteria.

Evolving a strategy to further metabolise these compounds has enabled some organisms to transform or recycle halogen equivalents, thus completing the halogen cycle. Few examples of biological dehalogenation have been studied in the context of natural product chemistry, as research has been focused instead on their potential role as bioremediation tools for anthropogenic organohalide pollution. But it is known that in microbial communities, dehalogenation reactions are coupled to energy conservation and growth or as part of the catabolism of halogenated organic molecules. In the latter, the removal of the X substituents will allow its utilisation as carbon sources, as will be discussed in the following section of this work.

[3]In mammals, few halogenated metabolites are known, with the thyroid hormones triiodothyronine and prohormone thyroxine being the best examples [14]

Environmental levels of organohalides liberated as by-products of the chemical industry or as a result of biogeochemical processes (known or un-characterised) are not trivial. Even more, there is strong evidence to suggest that most anthropogenic organic halogenated molecules and their partial degradation derivatives are being deposited in low-temperature regions from terrestrial sources by a variety of atmospheric and oceanic transport mechanisms. These include non-destructive processes such as volatilisation, sorption phenomena and dispersion and destructive ones, like chemical transformation **[4]**. Some authors, *i.e.* Chatt *et al.* **[22]** have also suggested that fractionation of volatile compounds is occurring during transport through the atmosphere so that a gradient develops all around the world, making these compounds ubiquitous and an unquestionable issue for the environment **[2, 4, 23]**.

Box 1.0 Enzymatic halogenation

Biosynthetic halogenation involves enzymes that belong to diverse and evolutionary non-related families and can be subdivided into three classes **[15]** depending on the halogen oxidation state:

- The first class consists of the oxidation of the halide X^- to an halonium ion X^+ (or into XO^- in aqueous solution), that participates in electrophilic substitutions. Flavin-dependent halogenases are representative of this group.
- The second type of mechanism consists in the formation of a halogen radical $X^\bullet$ after a single electron oxidation process. The radical species can then react with inactivated carbon centres. Vanadium dependent haloperoxidases and non-heme-iron dependent halogenases exemplify this class.
- The third class of halogenation reaction is catalysed by S-adenosyl-methionine (SAM) dependent enzymes via S_N2 type mechanisms, like halide methyltransferases and 5'-halo-5'-deoxyadenosine synthases, like the FIA enzyme from *Streptomyces cattleya*, the only enzyme capable of catalysing the incorporation of fluorine to generate fluoroacetate and fluorthreonine **[24]**.

*

1.2 Organohalide pollution and impact in the environment

Although some efforts to control and restrict the use of organohalides started as early as the 1970s, it was not until 2001 when the industrial production of most anthropogenic organohalides was regulated by the Stockholm Convention (http://www.pops.int/) for persistent organic pollutants (POPs). Huge attention from researchers, environmental organisations and the public regarding the ozone layer depletion and its direct link with photooxidation process involving reactive halogen species, principally derived from fluorochlorocarbon compounds **[25]**[4], has brought the environmental effects of organohalide use into the limelight. However, the indiscriminate utilisation of organohalides during the last century and their chemical stability has led to its accumulation in soils and water sources in quantities that far exceed the resilience of ecosystems. Surface water sources, aquifers, and sewage are generally affected by the presence of halogenated aliphatic compounds of low molecular weight, such as trichloromethane (TCM), trichloroethane (TCE), and perchloroethane (PCE), but also by some highly functionalised aromatic halides and heterocycles **[27]**. Fluorinated, chlorinated, brominated, and iodinated benzenes, phenolic derivatives, cyclodienes, and polyhalogenated biphenyls are some of the most representative compounds due to their commercial significance and stability (**Figure 3**).

Halogenated organic compounds enter the trophic chain after the transfer of soil particles to plants by wind, by atmospheric deposition of contaminants after rain, and by direct consumption of contaminated water and organic matter **[28]** due to the bioaccumulation of these compounds in fat tissues, given their hydrophobic (lipophilic) nature. At the base of the marine ecosystems, organic pollutants are absorbed by plankton and krill and the effect is biomagnified to critical concentrations at higher trophic levels **[29]**. In the deep-sea ecosystems, like the Mariana trench, pollution with POPs is higher than in coastal ecosystems, as most populations analysed (including Amphipoda crustaceans) exhibit at least 50 times higher organohalide bioaccumulation than their shallow-water counterparts, perhaps because of run-off process from polluted land areas or because of the accumulation of organic matter, but also as a result of plastic waste transported by oceanic currents **[30]**.

[4] Halogen-induced catalytic ozone loss was first characterised in the Arctic troposphere because of bromine levels. Halogens act as strong oxidisers that reduce the amount of OH radical species that play an important role in atmospheric photooxidation process **[25]**,**[26]**.

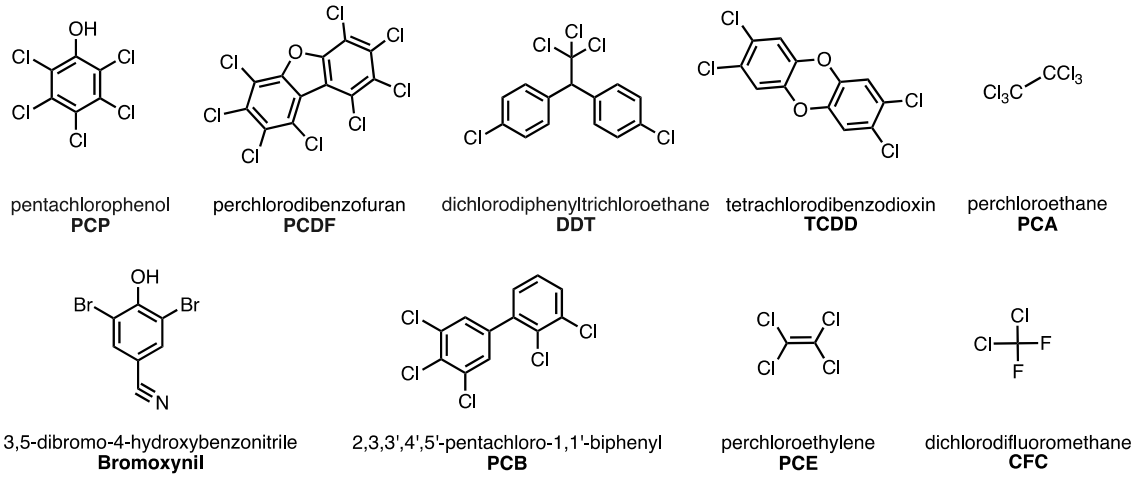


Figure 3. Chemical structures of aliphatic and aromatic organohalides are listed as POPs by the Stockholm Convention, given their industrial importance as solvents, synthetic precursors, and its applications in agriculture.

Pollution with some anthropogenic halogenated compounds like polychlorinated biphenyls (PCBs)[5] has impacted marine top-predators like sharks **[32]** and cetaceans, whose wild-life populations face decline as a direct consequence of the bioaccumulation and biomagnification of PCB by their trophic interactions, as was recently shown by Desforges *et al*. **[33]** on killer whales *(Orcinus orca)*. PCBs are known to be carcinogenic, disrupt endocrine and immune systems, and cause reproductive impairment **[34]** and this, in combination with other environmental pressures, might put this species and many others at risk of extinction. For that reason, finding appropriate decontamination strategies by physical, chemical, or biochemical methods is one of the major challenges in sustainable development **[35, 36]**.

However, traditional strategies of bioremediation of highly contaminated sites are expensive and inefficient. In contrast, new techniques that exploit the potential of microbiota, specifically the strategies evolved by some consortiums of both Gram-positive and Gram-negative microorganisms, offer promise. These consortia evolved adaptive mechanisms to survive under environmental stresses derived from the presence of naturally occurring organohalide compounds.

[5] Global production of PCBs in the 20th century, after its production became regulated under the Stockholm Convention, is estimated in 1.5 tonnes **[31]**.

1.3 Organic respiring bacteria

Since biological dehalogenation has been traditionally studied from the perspective of finding a solution to anthropogenic contamination, an increasing number of microorganisms that are able to dehalogenate both aromatic and aliphatic persistent pollutants have been isolated from different ecological niches **[37, 38]**. According to microbiologic and proteomic studies, it is known that organohalide microbial degradation occurs by a variety of strategies that are not limited to strict anaerobic habitats **[16]**. From a chemical perspective, organohalides are often considered as synthetic end-point products given their stability and the fact that these compounds are bad substrates for common oxidative degradation pathways. However, recent studies performed on marine aerobic organisms have shown that dehalogenation mechanisms are much more diverse than originally thought.

Box 2.0 Other enzymatic dehalogenating systems

Other types of molecular architectures exist that can link organohalide degradation to central catabolic metabolism in Bacteria, often by recruitment of promiscuous enzymes that participate in completely different routes and therefore might possess distinct catalytic mechanisms to those of the B_{12}-dependent reductive dehalogenases. Some examples are:

- The glutathione-S-transferase-like dehalogenase (GST superfamily) that participates in the catabolic pathway of tetrachlorohydroquinones in aerobic marine α-Proteobacteria is another type of reductive dehalogenase system that is not related to the B_{12}-dependent reductive dehalogenases **[21, 39]**.
- The flavin-dependent oxygenases, cytochrome P450 oxidases, and dehaloperoxidases are oxidative dehalogenases that catalyse the substitution of a halogen group for a hydroxyl group in halogenated aromatic compounds **[9]**.
- The hydrolytic dehalogenases (HDH), like the haloalkane and haloalcohol dehalogenases, are perhaps the most studied class of dehalogenases since its expression in heterologous systems is easy, in contrast to the B_{12}-dependent reductive dehalogenases. These enzymes catalyse the removal of halogens (including F-) and incorporation of hydroxyl groups via intramolecular o intermolecular redox-neutral reactions depending on the nature of the substrate **[39, 40]**.

Most of these organisms have been collected from highly specialised ecological niches in water ecosystems, where either oxidative or reductive conditions prevail. These selective pressures, in turn, have ensured the development and selection of non-evolutionary related enzymes that can catalyse the dehalogenation of organic compounds through oxidative, reductive, and hydrolytic reactions, not only as part of catabolic pathways but also during synthetic transformations of natural product intermediates [41].

In strict anaerobic bacteria, organohalides are degraded as part of metabolic pathways where they act either as carbon and/or energy sources, in what can be understood as a respiratory chain where the organohalide compounds act as final electron acceptors. This process has been named microbial organohalide respiration [27] and it is one of the many different strategies evolved by microorganisms to thrive in oxygen-depleted habitats, like industrial slurries or sludges.

Recently, reductive dehalogenation has also been observed in non-strictly anoxic conditions, where it can be linked to oxidative degradation processes [42]. Generally, reductive dehalogenation involves a chemiosmotic gradient that couples the catabolism of organohalides to the production of ATP equivalents [43], but it also occurs co-metabolically as an alternative respiratory chain in aerobic conditions or as part of catabolic pathways where the removal of the halogen atom allows the carbon backbone to be fully degraded [44].

The possibility of these compounds to be used as terminal electron acceptors by microorganisms largely depends on their specific reduction potentials. The more oxidised molecules are those that exhibit more positive potentials in a redox scale and have lower values of activation energy than the less oxidised compounds. Highly substituted aromatic compounds compete more effectively with terminal electron acceptors for growth in some cell cultures in anoxic conditions [4], as will be discussed later in this work.

From a thermodynamical perspective, organohalides are good electron acceptors during reductive dehalogenation in anaerobic conditions [45], as their redox potentials are highly positive (ranging between +240 and +560 mV) and are therefore comparable to those of the NO_3/NO_2 redox pairs (+433 mV), but nowhere near to those corresponding to the reduction of O_2 to H_2O (+818 mV), considering the H_2 electrode as reference at standard conditions **(Table 2)**.

Redox pair		E°'
		(mV)
hexachlorobenzene	pentachlorobenzene	478
pentachlorobenzene	1,2,3,4-tetrachlorobenze	421
1,2,4-trichlorobenzene	1,2-dichlorobenzene	349
monochlorobenzene	benzene	310
2,3,4-trichlorophenol	2,3-dichlorophenol	406
2,3,5-trichlorophenol	2,3-dichlorophenol	449
	3,5-dichlorophenol	393
3,5-dichlorobenzoate	3-chlorobenzoate	331
3-chlorobenzoate	benzoate	297
3-bromobenzoate	benzoate	379
3-fluorobenzoate	benzoate	299
pentachlorophenol	2,3,4,5-tetrachlorophenol	399
2,3,4-trichlorophenol	2,3,4-trichlorophenol	327
NO_3^-	NO_2^-	433
O_2	H_2O	818

Table 2. Oxidation and reduction potentials measured against the standard H_2 electrode for numerous organohalides [8]. The difference in the redox potentials between organohalide and O_2/H_2O pairs might explain why reductive dehalogenation preferentially occurs in anoxic environments.

According to the values of the free Gibbs energy for the enzymatically catalysed dehalogenation reaction, when H_2 is used as an electron donor, the process is highly exothermic and would prove suitable for energy production in anaerobic or moderately aerobic environments. However, as stated by Schubert *et al.* [43] there are thermodynamical and mechanistic restraints to the enzymatically catalysed reaction that prevent this energetic gain to be fully exploited for the production of ATP, given the low proton to electron ratio (H^+/e^-), considering the theoretical values calculated for the translocation of protons through the cell membrane by the ATP-synthase.

Microorganisms able to degrade organohalides by means of enzymatic catalysis are generally called "Organohalide Respiring Bacteria" (OHRB) [36], although Archaea are also capable of metabolising halogenated organic compounds and to synthesise cobalamin, particularly methanogens from the phylum Euryarchaeota; this feature will prove of high importance, as will be discussed later. Some Methanosarcina species can be found in anaerobic microbial communities with organohalide respiring Bacteria like *Dehalococcoides* spp. [46], as it has been reported that OHRBs are dependent on associations with other microorganisms for the supply of essential nutrients, specifically electron donors or cofactors [36]. These microbial ecological associations are particularly

relevant for the development of extremophile-based organohalide bioremediation strategies for industrial wastewaters where most prokaryotic cultures would not survive due to the high salinity and acidic or alkaline pH conditions.

Most OHRBs known to date belong to three Bacteria *phyla*: Firmicutes, Proteobacteria, and Chloroflexi **[36]** and can be further subdivided into obligate OHRBs and facultative OHRBs depending on the flexibility of their metabolism, as shown in **Figure 4** **[47]**. In the first category, strict anaerobes from the phylum Chloroflexi, that exclusively depend on H_2 as electron donor, use organohalides as terminal electron acceptors. In facultative organisms, most of them from the phylum Proteobacteria, the electron donors are diverse[6] and organohalides are one of many terminal electron acceptors, such as sulfates, nitrates, metal ions such as Fe (III) and even O_2, in micro-aerobic environments **[44, 48, 49]**.

Since the first organisms capable of reducing organohalides were discovered in different habitats, great interest in understanding the nature of the ecological interactions between microbiological communities in water and soils arose. Detailed microbiological studies with OHRBs cultures might provide us ultimately with valuable strategies for the implementation of novel long-term applications in bioremediation of heavily polluted sites. However, these bacteria grow at a slow rate and are highly sensitive to temperature, pH, redox conditions, light, and symbiotic interactions **[44, 50]**.

Biochemical studies on the pathways that involve the dehalogenation of organohalide compounds and their evolutionary background might help to solve some of the biggest unanswered questions in microbial ecology and provide an answer as to whether these pathways evolved as adaptive strategies for specific environmental conditions from the past or recently, in response to the anthropogenic pollution. Even when there is limited information about the biochemical context and catalytic mechanisms through which OHRBs perform the dehalogenation reactions and the substrate selectivity, it is now clear that such reactions are carried out by enzymes called reductive dehalogenases (RdhAs or RDases) [7] **[36, 51]**.

[6] Electron donors include pyruvate, formate, acetate, lactate and glucose, in facultative anaerobic bacteria as from the *Desulfitobacterium* genus.

[7] The abbreviation "RdhAs' refers to all reductive dehalogenases whose sequences are deposited in the databases. Whereas "RDases'' is the term used only for the enzymes that have been characterised. In this thesis, the RdhA notation is preferred.

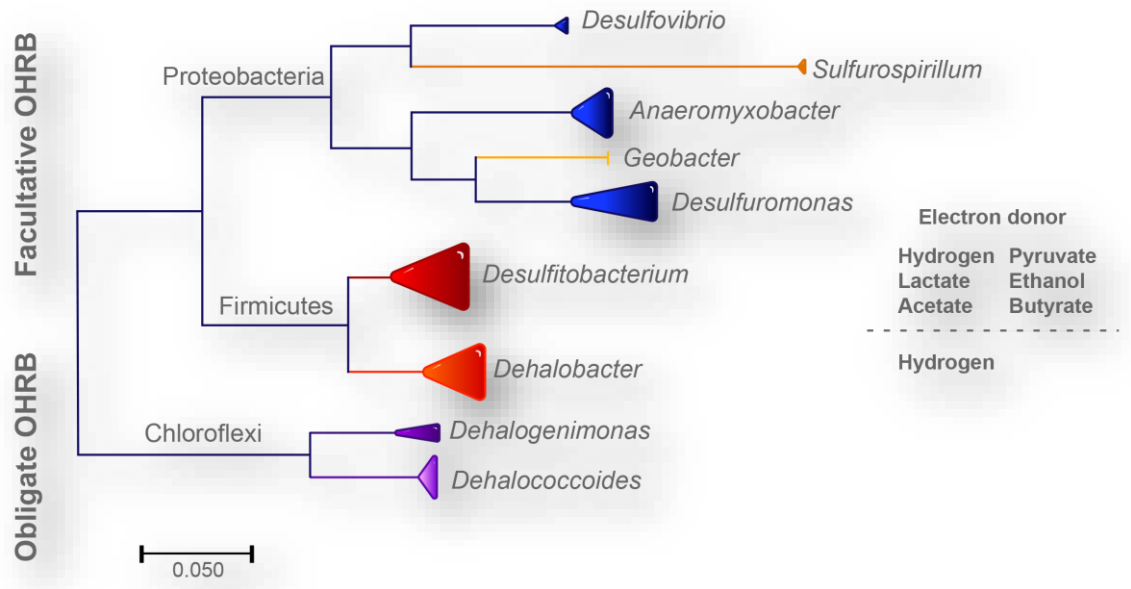


Figure 4. Maximum likelihood phylogenetic tree of organic respiring prokaryotes able to reduce aliphatic and aromatic organohalides, based in 16S ribosomal RNA sequences. Most ORBs known belong to the bacterial phyla Proteobacteria, Firmicutes, and Chloroflexi. Figure adapted from **[47]**.

1.4 Reductive dehalogenases RdhAs and other Cbl-dependent enzymes

RdhAs are a subclass of oxidoreductases (EC 1.97.1.8) that show dependence for corrinoid cofactors and iron-sulfur clusters, with exception of the *Desulfomonile tiedjei* DCB-1,3-chlorobenzoate RdhA, that instead of cobalamin possess a heme group in the active site, although this has not been verified independently by other research groups **[40, 52]**. Cobalamin and its derivatives participate in thermodynamically challenging reactions **[39]**, mainly because the central cobalt atom coordinated by the tetrapyrrolic macrocycle has three accessible oxidation states, all of them relevant for catalysis. A summary of the general features of the corrinoid cofactors and Fe-S clusters will be presented later in a different section, though a brief description of the B_{12}-dependent systems is to follow, in section **1.5**.

B_{12}-dependent enzymes can be subdivided into three classes, according to the type of cobamide derivative involved, the geometry of the complex, and the type of axial ligands: 1) 5'-deoxyadenosyl cobalamin (Ado-Cbl)-dependent proteins, 2) methyl-cobalamin (Me-Cbl)-dependent transferases and a third group that comprises the reductive dehalogenases, together with epoxyqueuosine reductases (QueG) and cobalamin decyanation/dealkylations enzymes (CblC) **[9, 10, 36]**. There is no indication that the central Co ion forms organometallic bonds during catalysis in the case of RdhAs and QueG;

as will be described later, the resting state of the central ion in the cobalamin is Co (II) [53]. Another homologous protein to the RdhAs is the B_{12}-trafficking chaperone methylmalonic aciduria Cbl-C protein (MMACHC) found in humans with homocystinuria, though the similarity is limited to the cobalamin binding domain. This enzyme catalyses the removal of the upper axial ligands from the CN-Cbl and alkylated-Cbl to form the typical organometallic cobalamin cofactors [54].

While it has been observed that the first two classes of Cbl-dependent enzymes exhibit a typical cobamide-binding motif (that usually comprise 5 stranded β-sheet flanked by 5 helixes similar to the typical Rossman fold) [9], this corrinoid binding motif common is not present in case of the RdhAs and QueG known to date [51].

B_{12}-dependent reductive dehalogenases can catalyse either a two-electron dependent α-elimination reaction that involves the cleavage of a C-X bond and the formation of a C-H bond by the formal addition of a hydride ion and dihaloelimination reactions that involve the removal of two halogens in adjacent positions that results in the formation of a double C-C bond, also called β-elimination, as exemplified in **Figure 5**.

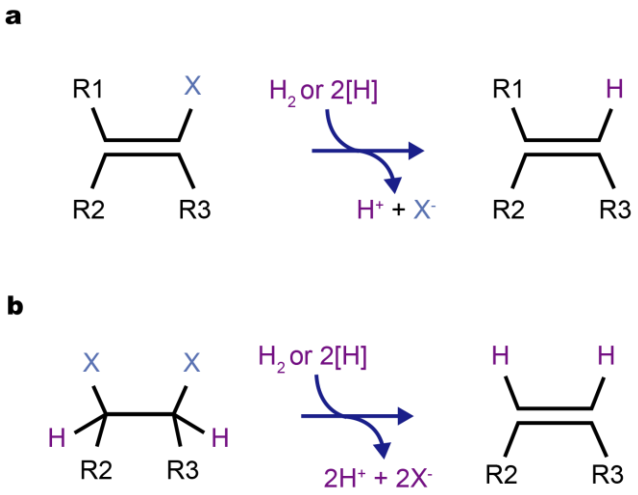


Figure 5. General reductive dehalogenation reaction schemes. The α elimination implies the substitution of a halogen group with an H^+ (a). The β elimination or dihaloelimination consists of the removal of two adjacent halogen substituents resulting in the formation of a double bond (b) [36, 47].

Experimental studies on the vinyl chloride reductive dehalogenase (VcrA) from *D. mccarty*i have proved that both reactions can be performed by the same RdhA and that the catalytic mechanism depends on the chemical structure of the substrate and not on the active site features [55]. This also hints at the fact that most RdhAs might be rather

promiscuous enzymes and possess a huge substrate scope, that might include both aromatic and aliphatic organohalides, as shown recently by Kunze *et al*. **[56]**.

1.5 Cofactors involved in reductive dehalogenation

1.5.1.A. Cobamide cofactors CN

Cobalt Co (with the electronic configuration [Ar]$3d^74s^2$) is an essential cofactor, particularly as the metal centre of cobamides[8], but there are few examples in Bacteria where non-corrin cobalt (II) is involved in enzymatic reduction and oxidation catalysis **[57]**. Cobalt is one of the rarest transitional metals in the Earth's crust, although it exists in numerous ores **[12, 58]**, it belongs to Group 9 (with rhodium Rh and iridium Ir) and when bound to cobamides has access to three different oxidation states in solution, as opposed to other transition metals, like iron Fe and copper Cu. Systems like d^6 (18 e-)/d^8 (16 e-) can readily react in oxidative addition and reductive eliminations.

Corrins are derived from the same porphyrin precursors as heme and chlorophyll. Despite its synthetic complexity, it has been speculated that corrinoids might have a prebiotic origin or that they might have originated biotically, early in evolutionary time **[59]**. These macrocycles can coordinate several transitional metals and form complexes whose chemistry or reactivity depends on the electronic configuration, oxidation state, and spin state of the metal centre, as well as the nature of the axial ligands. Corrinoids are smaller than porphyrins and its structure is uniquely asymmetric due to the loss of a carbon atom and the formation of a direct bond between two pyrroles **[59]**. Other modifications of the corrinoids include eight methylations and multiple substitutions with acetamide and propionamide groups **[10]** at the periphery of the macrocycle that prevent the oxidation of the double bonds **[59]**. Additionally, the conformational flexibility of corrin rings can confer stability to certain oxidation states in the central metal ion, in contrast to the porphyrins **[59, 60]**; the delocalised electronic density over the double bonds of the macrocycle causes a distortion of the planarity and contracts the coordination site of the metal centre, facilitating the coordination of the cobalt ion, instead of other ions with bigger atomic radius, like Fe. The metal centre is coordinated by the four equatorial pyrrolic nitrogens and different axial ligands. Cobalamin is the most complicated biological cofactor, but the most

[8] Cobamides are corrinoids with complete nucleotide loops, as opposed to cobinamides (were only the equatorial macrocyclic is present.

common of the cobalt-containing cofactors found in Nature **(Figure 6)** [61], its structural beauty and its remarkable catalytic features have led to decades of study since it was first isolated, crystallised and its total synthesis was reported by Woodward and Eschenmoser [10].

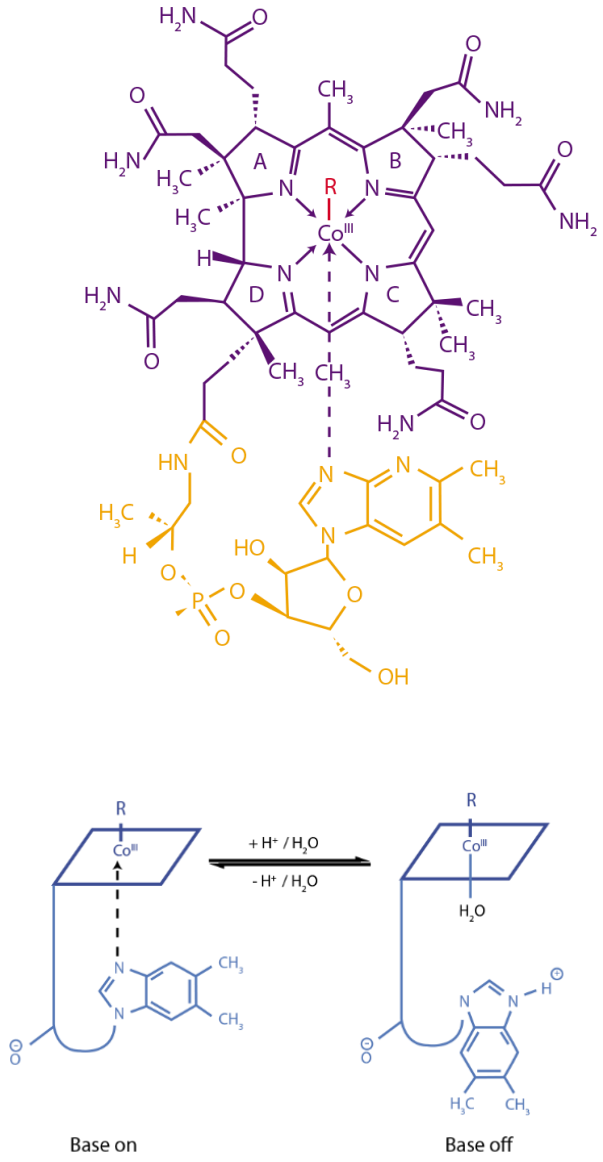


Figure 6. Structure of cobalamin. The axial ligand can be R= CN in cyanocobalamin (CN-Cbl), CH_3 in methylcobalamin (Me-Cbl), and 5'-deoxyadenosyl in adenosylcobalamin (Ado-Cbl). The corrin ligand is composed of four pyrrolic rings (A to D), here depicted in purple. The 5,6-dimethylbenzimidazole (DMB) nucleotide moiety is represented in yellow. Cobalamin is also presented in its base-on and base-off forms as a consequence of acid-base dependent equilibria. Figures are taken and adapted from [36].

The first cobalamin to be crystallised as a red cobalt complex was vitamin B_{12} (CN-Cbl), a derivative of the organometallic biologically active forms, that exhibits a nitrile group

in the upper axial position as a consequence of the isolation process [62]. The metal centre of cobalamin can exist in three different oxidation states in aqueous solutions and at physiological conditions: Co^{1+}, Co^{2+} and Co^{3+}, and for that reason, oxidation and reduction processes and electronic transfer reactions determine the reactivity of the B_{12} cofactors and its role in enzyme catalysis [62]. As with other metal centres, the number of ligands (and thus the coordination number) decreases with the oxidation state of the Co. The coordination geometry of most cobamides can be defined as pseudo-octahedral since it is tetragonally distorted due to the strain of the macrocycle [63]. In these complexes the most common oxidation state is Co^{3+}; the hexacoordinated d^6 low-spin Co (III) complexes are diamagnetic and normally observed in solution and in the active site of enzymes in their resting states, although that is not the case in RdhAs and QueG [9, 51].

It has been observed that cobalamin and its precursors, like cobyrinates and cobesters, are able to catalyse chemical regioselective and stereoselective transformations when free in solution [61]. This is particularly true in case of the dehalogenation of organohalides such as lindane, with (H_2O)-cobalamin, in presence of reductants like 1-thioglycerol, $NaBH_4$, DTT, and titanium (III)-citrate [39, 64], being the base-off forms of the cobalamin derivatives the most effective catalysts.

1.5.1.B. Organometallic B_{12}-dependent enzymes

In the biologically relevant complexes methylcobalamin (Me-Cbl), 5'-deoxyadenosyl cobalamin (Ado-Cbl), and cyanocobalamin (CN-Cbl), the metal centre is bound to the carbon-donating upper axial ligand. This feature means that these cob(III)alamin derivatives are one of the few naturally occurring organometallic complexes [58, 65]. Cobamide organometallic cofactors, though unstable and photosensitive, are involved in a broad range of remarkable enzyme catalysed reactions due to the reactivity and lability of the metal-carbon bond. The dissociation energy of the Co-C is about 125 $kJmol^{-1}$, relatively weak for covalent bonds, but it is this feature that confers unusual catalytic properties that are exploited by Nature and chemical synthesis for complicated transformations.

Macrocycles as corrins, chlorins, and porphyrins present electronic transitions within the UV/Visible region of the electromagnetic spectrum, a feature that enables them to participate in a number of photoactivation processes that involve electron relaxation dynamics that occur in picoseconds [60]. Light-induced reduction and oxidation processes

with physiological relevance have also been reported in case of cobalamin cofactors [66]. A new class of photoreceptors that depends on Ado-Cbl as chromophore was found in Bacteria [67]; it is involved in the metabolic response to light and dark periods by responding to electromagnetic radiation of different wavelengths. This prototypical light-sensitive receptor is the tetrameric protein CarH, that regulates the biosynthesis of carotenoids on a transcriptional level. It is a modular protein that possesses a DNA binding domain in addition to the Cbl-binding domain and effector domains, that are similar to the His-kinases. When exposed to light, the organometallic Co-C bond breaks causing large-scale conformational changes that destabilise the tetrameric complex, leading to the activation of the transcription.

The activity of both Ado-Cbl and the Me-Cbl containing enzymes depends largely on the chemical properties of the Co-C bond and the upper axial ligands of the cobalamin, thus supporting different mechanisms of catalysis. The Ado-Cbl-dependent enzymes are implicated in unusual isomerisation reactions and intramolecular rearrangements. Ado-Cbl-dependent isomerases constitute the biggest superfamily of B_{12}-dependent enzymes. The catalytic cycle of these enzymes consists of the homolytic cleavage of the organometallic bond in the Ado-Cbl that yields two carbon radicals in a reversible process (a cob(II)alamin radical and a 5'-deoxyadenosyl radical) [68]. The activation of the coenzyme that directs the cleavage of the Co-C bond is triggered by the binding of the substrate, as is represented in **Figure 7**.

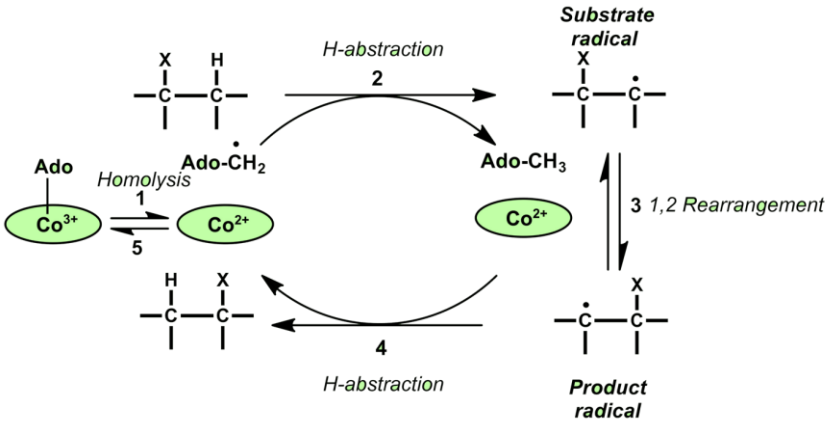


Figure 7. General Ado-Cbl-dependent enzymes catalytic mechanism of isomerisation. In most coenzyme B_{12}-dependent enzymes, it is the radical species the one that abstracts the hydrogen atom from the substrate, generating a 1,2-rearrangment in the isomeric radical intermediate that later abstracts a hydrogen atom from the Ado moiety to give a stable product and an Ado radical that finally reacts with the B_{12} to regenerate the organometallic catalyst. Adapted from [68].

The coenzyme B_{12} (or AdoCbl) enzymes have been subdivided into C-skeleton mutases, diol dehydratases and ethanolamine ammonia-lyase, amino mutases (where the

migrating groups are either amino or hydroxyl groups), and as B_{12}-dependent ribonucleotide reductases. In the B_{12} ribonucleotide reductases that catalyse the conversion of ribonucleotides to deoxyribonucleotides, the AdoCbl is bound weakly in the base-on conformation, as opposed to isomerases and mutases.

Me-Cbl-dependent enzymes catalyse methylation reactions that involve the heterolytic cleavage of the organometallic bond and the generation of a methyl carbocation and Co(I) intermediates; between the Me-Cbl-dependent enzymes group, the methionine synthases and methyltransferases are the most studied systems **[59, 69]**. In most Me-Cbl-transferases, methyl group transfer occurs via nucleophilic substitution reactions that depend directly on the heterolytic formation and cleavage of the organometallic bond **(Figure 8)**. In the B_{12}-dependent methyltransferase enzymes, such as methionine synthase, catalysis occurs in two steps in a sequential mechanism that is assisted by a catalytic triad of Ser, Asp, and His residues that coordinates the lower axial position of the cob(III)alamin.

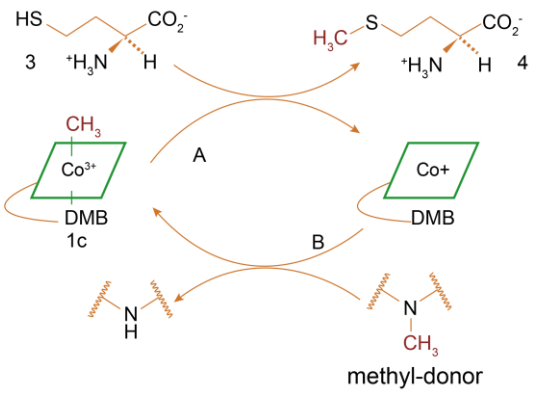


Figure 8. Methionine synthase (a Me-Cbl-dependent enzyme) mechanism of catalysis via two successive S_N2 reactions. During the first step of the catalysis, the methyl group bound in the upper position of the Cbl is abstracted by an activated homocysteine residue (due to the coordination of the thiol to a Zn(II) cation) producing the concomitant formation of methionine and a cob(I)alamin intermediate that later attacks the methyl group of N-methyltetrahydrofolate, thus regenerating the cofactor.

In B_{12}-dependent methyltransferases, the corrinoid is bound in the base-off conformation. The methyl transfer occurs via two S_N2 reactions, involving the formation of a *formal* carbocation species so that overall, the configuration of the substrate and the product is retained through the catalytic cycle. Analysis of crystal structures of the methionine synthase shows that there are major conformational changes in the His residue that control its coordination to the α-face of the cobalamin, ultimately determining the bond strength of the organometallic bond. The coordination of the His residue also has an effect

on the redox properties of the Cbl, so that the reduction of the cob(II)alamin only occurs to produce a cob(I)alamin intermediate when the axial coordination weakens [68].

There are multiple binding topologies for the complete cobalamin cofactors when interacting with the polypeptide chain, usually, the cobalamin plane is placed parallel to the protein interface between two domains or subunits so that the upper and lower axial positions are exposed to different electronic environments. Often, the lower axial position or α-face of the Cbl is bound via close contacts, whereas the upper axial position or β-face of Cbl faces an exposed or accessible solvent cavity of the activating domain. These different binding modes of the cobalamin cofactor aid in the positioning of the substrate or just provide conformational flexibility during catalysis.

1.5.1.C. Cobalamin cofactors in reductive dehalogenation

In OHRBs, structural variations in the loop composition have been observed, being the demethylations or changes in the lower base identity the most common [70]. It is thought that this might influence the catalytic behaviour of the cobalamin, although no involvement of the lower axial ligand during catalysis has been found [51, 71, 72]. Upon binding to the protein active-sites, the nucleotide loop coordinating the Co ion in the α-face can often be displaced by other ligands or even amino acid residues [59].

In RdhAs and QueG there are important differences in both the type of cobalamin and in the features of the binding site itself, with respect to the other B_{12}-dependent enzymes. Though the mechanistic details are still unknown, crystal structures and spectroscopic techniques indicate that in the resting state the DMB base is not coordinated to the Co. The lower axial ligand remains bound to the propionamide tail but steric hindrance of the aa side chains prevents the coordination to the metal centre, thus rendering a stable cob(II)alamin complex, where the upper axial position can be occupied by water molecules or by the substrate [51, 71, 73]. Experimental evidence suggests that these enzymes perform fascinating chemistry, that is markedly different from other cobalamin-dependent enzymes, that stabilise the cob(III)alamin complex instead and depend entirely on organometallic catalysis.

Cobamides can be synthesised by various anaerobic and aerobic microorganisms, in what is known as one of the longest biosynthetic pathways with approximately 30 biosynthetic steps [74], particularly in Bacteria (Proteobacteria) and Archaea (Thaumarchaeota and Cyanobacteria); for that reason, cobalamin derivatives are vital

nutrients that have to be acquired through diet in case of all eukaryotes [75]. Some strains of OHRBs synthesise cobalamin *de novo*, like is the case of *Sulfurospirillum multivorans, Dehalobacter hafniense,* or some *Dehalobacter* spp. [36, 76, 77]. However, in microbial dehalogenating communities, there are also auxotroph organisms that acquire their cobalamin cofactor by multiple symbiotic interactions, like is the case of *Dehalobacter restrictus* and *Dehalococcoides mccartyi* [78, 79]. This strongly suggests that in habitats where organohalide concentrations are considerable and where OHRBs thrive, some ecological interactions with cultures capable of biosynthesising corrinoids are present and necessary for the consortiums to survive [6].

The uptake of B_{12} requires a complex set of specialised transporters and chaperones. Those organisms unable to synthesise cobalamin (auxotrophs) can import it from the media via transmembrane proteins [80]. *Escherichia coli* is a cobalamin auxotroph, fact that constitutes a problem to overexpress B_{12}-containing enzymes. Most Gram-negative bacteria possess an external membrane that is impermeable to molecules that are bigger than 600 Da, thus requiring specialised transporters such as the B_{12} uptake (or Btu) system, that allows the active transport of the cobalamin from the media through the cell membrane. BtuB is a 594 aa β-barrel [81, 82] of the TonB type. The uptake initiates with the binding of the extracellular Cbl to BtuB, which translocates it into the periplasmic space where it binds to the BtuF carrier. This protein then transports it into the cytoplasm, aided by a complex conformed by BtuC, BtuD, and BtuF transmembrane proteins, that rely on the hydrolysis of ATP. Corrinoid transporters, like the TonB system, present different modes of binding of the Cbl, depending on both the identity of its upper ligand and loop composition [59, 68].

1.5.2 Iron-sulfur clusters (Fe-S)

Iron clusters and inorganic sulfide [Fe-S] clusters are the most ubiquitous prosthetic groups in nature [83]. In the primitive Earth, the environment was anaerobic, and free iron and sulfur were abundant; it has been proposed that their abundance facilitated the spontaneous assembly of Fe-S in the ancient protein scaffolds. The simplest biological iron-sulfur cluster consists of a single Fe atom coordinated by four Cys residues, but Fe-S can also exist in nature in the form of [2Fe-2S], [3Fe-4S], [4Fe-4S], or [8Fe-7S] core units bridged by inorganic sulfide, according to crystallographic studies **(Figure 9)**. Though the most common amino acid residue contributing to the formation of Fe-S is Cys, where the Fe is coordinated by the thiolate group, other residues like His, Asp, and Arg can also participate in the formation of clusters [84]. The usual oxidation states in the biological Fe-S are paramagnetic, but studies in magnetochemistry have indicated that the Fe-S are mixed

valent systems where the different metal centres present different oxidation states, causing spin-dependent delocalisation (SDD) phenomena [85]. Fe-S have received a lot of interest due to its different roles in enzymology and this has led to extensive studies leading to the total synthesis and characterisation of novel synthetic clusters with higher nuclearity. In these synthetic Fe-S, the Cys residues are substituted by inorganic sulfide [84]. Fe-S can undergo conversion and interconversion reactions, both in the protein-bound and free states, fact that suggest that the presence of the protein residues is not necessary for the assembly of the clusters. The comparison with the synthetic Fe-S has provided researchers with an explanation of the importance of the protein environment in determining the chemical properties and reactivity of the clusters in enzymes.

Most of the Fe-S clusters are major components of electron transport pathways, like in photosynthesis or oxidative phosphorylation [86]. They can act as single or double electron carriers, as well as redox reaction centres in several proteins, like nitrogenases, where the proton and electron transfer are coupled. Fe-S channel electrons between redox pairs that are physically separated by distances of approximately 10 – 14 Å and it is speculated that these distances might have been selected by evolution to allow rapid transfer of electrons between the clusters within the same, in different domains, and even between independent proteins, while tuned to diminish the number of clusters required for the process to occur [85, 87].

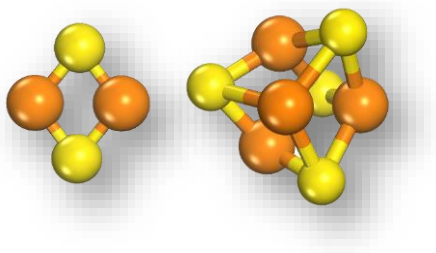


Figure 9. Iron-sulfur clusters: [2Fe-2S] (left) and [4Fe-4S] (right). Modeled using Materials Studio 8.0.

The electronic delocalisation found in the Fe-S and the electronic affinities of the metal centres, largely determine the rate of the electron transfer between the clusters and other redox centres or carriers and the substrates, so that the process is exergonic –(ΔH) (endergonic processes are minimised). Among other properties that define their chemical behaviour, the ability of Fe-S to support multiple stable oxidation states **(Table 3)** and the broad range of oxidation-reduction potentials of the different clusters, is what makes these systems the cofactors of choice for electron transfer reactions.

Oxidation states of common Fe-S		
[Fe-4S]	Stable/Active	2+, 3+
[2Fe-2S]	Stable/Active	1+, 2+
[3Fe-4S]	Stable/Often Inactive	0, 1+
[4Fe-4S]	Stable/Active	1+, 2+, 3+
	-600 to -450 mV	

Table 3. Oxidation states of the most common biological Fe-S. In comparison to other redox systems, Fe-S are known to work at the lowest range of reduction-oxidation potentials, in contrast to flavins, other metal cations, heme groups, pyridine nucleotides, and quinones [87].

It has been reported that Fe-S can also act as cationic substrate binding sites for oxoanionic and nitrogenated substrates or serve as sensors of free Fe, O_2, O_2^- [85]. This alternative role of the Fe-S has been extensively studied in dehydratases, where superficial Fe-S bind the substrate while providing a positive charge to stabilse the intermediate species during the catalysis. In general, the process occurs when the substrate substitutes an H_2O molecule placed instead of one of the four coordinating Cys of a central Fe atom. This behaviour depends on the capacity of Fe cations to behave as Lewis acids and change from tetrahedral to octahedral geometries in a non-redox process [87]. Interestingly, Fe-S clusters were conserved as the cofactors of choice in many respiratory systems, hinting to their important roles in microbial ecology even after the increase in atmospheric O_2 more than 2.5 Ga as a consequence of photosynthetic metabolism. Since O_2 has the potential to oxidise transition metals, when Fe^{2+} gets oxidised to Fe^{3+} it tends to precipitate as part of insoluble complexes, limiting the concentration of free iron to be incorporated into the cells [9] in aerobic environments. Fe-S can be destroyed by oxidation by univalent oxidants (superoxide, O_2, H_2O_2), for example, the cubane $[4Fe\text{-}4S]^{2+}$ clusters can be oxidised to inactive forms, like the $[3Fe\text{-}4S]^{1+}$ in the aconitase's family of enzymes. It has been observed that the oxygen-sensitivity in strict anaerobic respiratory chains is largely derived from the fact that most Fe-S are positioned near the protein surfaces and are therefore exposed to the solvent, whereas in aerobic pathways, the Fe-S tend to be occluded and protected by the polypeptide, preventing their oxidation and making these processes oxygen-tolerant. This perhaps shows the importance of the mechanisms that increase the bioavailability of iron

[9] Bacterial demands for intracellular Fe are high, reaching approximately mM concentrations and most of it is directed to the assembly of Fe-S.

and the expression of superoxide dismutases, reductases, and peroxidases, that prevent its oxidation [87]. Regardless of this, Fe-S systems were retained without dramatic changes as primary electron carriers, but it is thought that the selective pressure altered the clusters so that an increase in their redox potentials was acquired by the incorporation of subtle modifications in the protein environment, in contrast to their ancestral and/or strictly anaerobic counterparts [87].

In RdhAs, two typical cubane 4Fe-4S clusters are found in the dehalogenase domain and on occasion a 3Fe-4S cluster [51, 71]. In any case, it is plausible that all the Fe-S in reductive dehalogenation participate in the electron transfer process from the reducing species to the active site. A new type of RdhA, that will be introduced later in this thesis, possess an additional 2Fe-2S iron-sulfur cluster, that might participate in channelling electrons from NAD(P)H through FMN via a single-electron transport process to the cob(II)alamin.

1.6 Domain structure, genetic context, and regulation of RdhAs

Genes that encode RdhAs are found throughout all OHRBs, as a single unifying feature despite the different phylogenetic and ecologic origins of these microorganisms. Up until now, more than 300 genes coding for putative B_{12}-dependent reductive dehalogenases had been identified using metagenomic analysis [56], even in the archaeal phylum Asgard [88], this fact suggests that the evolutionary origin of the reductive dehalogenases might have preceded the diversification of the Bacteria and Archaea domains.

Despite the many advances in genome sequencing technology and the constant update of public databases, few *rdhA* gene products have been fully characterised due to their complex nature and the elevated experimental complexity [71]. Regardless of this, it has been established that many genes associated with halorespiration are present in multiple copies, as a result of gene duplication or horizontal gene transfer and further diversification processes. The presence of numerous paralogous genes throughout the genome corresponds to an expanse in the substrate specificity of the encoded RdhAs [6].

One of the major problems that have limited the progress of research focused on reductive dehalogenation is to achieve its heterologous expression in large quantities, particularly because of the cofactor requirements to obtain active proteins, but also because

most RdhAs are sensitive to oxidation, are associated to membranes or depend on the interaction with other proteins and therefore tend to be unstable and insoluble [36].

Some *Dehalococcoides* spp. strains of the Chloroflexi phylum contain up to 36 orthologous *rdhA* genes in its genome, while some *Dehalobacter* spp. of the Firmicutes have been found to have up to 25 genes coding for dehalogenases [89]. This diversity within the same microbial genome (that ranges from 2.5 to 3.1 megabases) may explain why OHRBs cultures can grow in presence of a pool of halogenated substrates, although mixtures of organohalides can result in inhibition of the organohalide respiration process [90]. The fact that various sets of functional enzymes have marked substrate specificities implies that there must be a subtle transcriptional regulation mechanism that allows the same OHRB to adapt to different niches, as will be discussed later in this work.

In accordance with its genomic context, most RdhAs genes are organised in bicistronic operons, composed mainly of two genes, usually known as *rdhA* and *rdhB*, encoding the catalytic subunit and the putative membrane-anchor protein respectively [91]. Some accessory sequences coding for proteins such as RdhK (possibly related to transcriptional regulation), RdhC (a membrane-bound regulatory protein), RdhD, and RdhT (chaperon proteins and the second one a fully characterised "trigger factor") [35] are also encountered in the gene clusters. By means of bioinformatic analysis, it has been established that the RdhAs proteins present low sequence similarity, as is the case of PceA of *Desulfitobacterium* sp. Y51, *D. multivorans*, TceA of *D. ethenogenes* 195 and CprA of *D. dehalogenans* where the identity values range between 20% and 30% amino acid identity, considering the total length of their sequence. Regardless of the poor similarity, these genes can be grouped as homologous [49, 91, 92]. The homology relationship between those proteins can be further analysed and clear orthologous groups that possess higher sequence similarity can be identified, but still, there is no obvious correlation about the substrate specificity of those proteins [52].

Although a classification system for all identified RdhAs does not exist, in general, all *rdhA* structural genes exhibit well-preserved domains and motifs **(Figure 10)**. Reductive dehalogenases can be divided as respiratory and catabolic RdhAs, as proposed by Payne *et al.* [51] depending on its specific domain features. The main differences between both groups are not reflected in their function as B_{12}-dependent reductive enzymes or their substrate specificity but in the type of metabolic pathways in which the enzymes are involved.

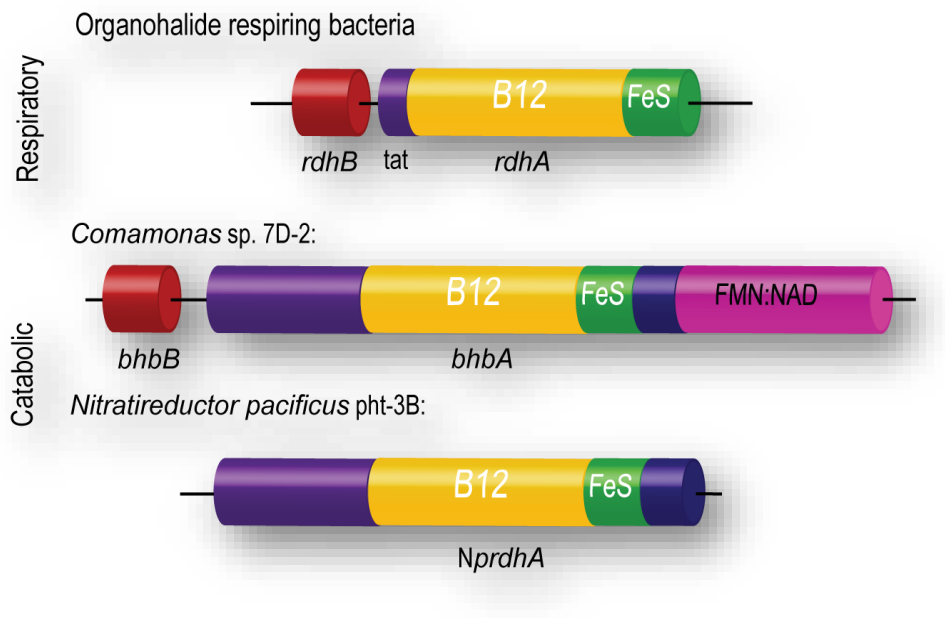


Figure 10. Representation of the domain structure of the respiratory and the catabolic reductive dehalogenases. The respiratory proteins have a molecular weight of around 35-65 kDa whereas the catabolic ones range from 75 to 150 kDa. In the scheme, the TAT (twin-arginine transit peptide), B_{12} (cobalamin-binding domain), Fe-S (iron-sulfur binding domain), FMN:NAD (reductase domain) are represented. The RdhB (equivalent to BhbB) corresponds to a membrane anchor. Adapted from Payne **[51]**.

As was stated before, organohalides are reduced during respiration by reductive dehalogenases in an energy conservation pathway that involves a chemiosmotic process coupled to energy production. However, under anaerobic conditions, reductive dehalogenation does not occur exclusively in primary metabolism like respiration, but also in organohalide degradation pathways where cytosolic reductive dehalogenases remove the halogen substituents enabling the carbon backbone to be catabolised as a carbon source (hence the term catabolic). It is worth mentioning that RdhAs may also be associated with fermentative pathways rather than a respiratory chain. Fermentative metabolism implies substrate phosphorylation level processes, where the organohalides serve as electron sinks and not terminal electron acceptors. This occurs under anaerobic conditions, when the pool of electron donors includes complex metabolites, like pyruvate and fumarate **[93]**.

1.7 Transcriptional regulation of the expression of RdhAs

In line with the metabolic costs of producing reductive dehalogenases and their associated molecular components, the expression of the structural genes *rdhA* (and *rdhB,* if

present) is under tight transcriptional control. Given the availability of new metagenomic tools, there has been an increase in the number of OHRBs genomes being deposited in the public databases and a wide range of distinct transcriptional regulators have been identified as part of the signal transduction processes that control the expression of the *rdhA* structural genes and it has been reported that some of these regulators have a restricted ligand specificity, as opposed to the low substrate selectivity of their corresponding enzymes **[35]**.

To understand organohalide respiration, studies have focused in analysing the highly variable operons where the *rdhA* genes cluster together with multiple structural genes that encode proteins that participate in the maturation of RdhAs **[36],** plus numerous additional regulatory components. Some transcription factors repress transcription and others activate it, but some can function either way according to the position of their transcription binding sites relative to the σ-factor binding sites (proteins that associate with the RNA polymerase and allow the recognition of the -35 and -10 sequences in the prokaryotic promoter region), thus controlling the expression of genes of whole regulons **[94]**.

Analysis of these transcriptional elements, that are associated with the *rdhAB* operons, have allowed the identification of at least three major regulatory systems at the transcriptional level **[89]**: the Crp/Fnr family (represented in OHRBs by CprK regulators), homologous proteins to the membrane-bound NirI/NosR regulators present in all OHRBs (represented by the *cprC* or *pceC* genes) and often located downstream of the structural *rdhA* genes **[95]**, the two-component regulatory systems or TCS (containing the stimulus-sensing self-phosphorylating His-kinases and the response regulators that bind the DNA) **[96]** and MarR type regulators.

Though most of the dehalogenase genes in *Dehalococcoides* spp. are associated with genes that encode MarR transcriptional regulators or the two-component signal transduction systems, the *Desulfitobacterium* genus, in contrast, depends on Cpr-Fnr-type regulators, a family known to regulate the transcription of proteins involved in oxidative stress responses, stationary phase survival, quorum sensing, nitrogen fixation, denitrification and pathogenesis **[97, 98]**. The transcriptional regulator CprK from the Gram-positive organism *Desulfitobacterium hafiense* DCB-2 induces the expression of the RdhAs upon binding of *ortho*-chlorophenols at μM affinities, thus enabling the binding of DNA and leading to transcriptional activation. The allosteric rearrangements in CprK were described by Levy *et al.* **[97]** by crystallographic studies **(Figure 11)**.

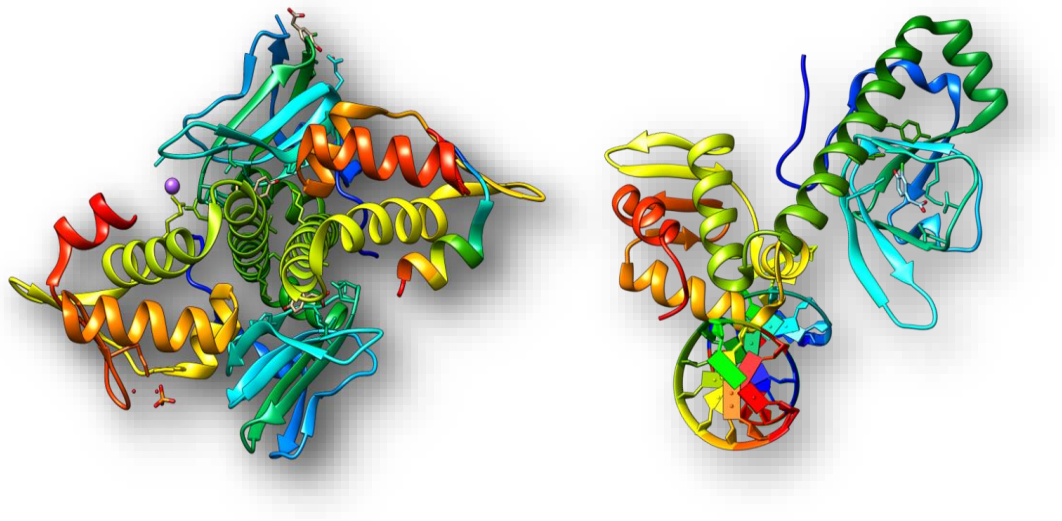


Figure 11. CprK crystal structure *Desulfitobacterium hafiense* DCB-2. Left: A cartoon representation of the homodimer structure is presented. The N-terminal β-barrel corresponds to the sensor domain linked via an α-helix Leu-zipper to the HtH domain for DNA binding (PDB ID 2H6C). Right: View of the ligand and dsDNA binding sites per monomer [97]. Figure generated in Chimera.

Members of this superfamily of regulators are homodimeric and contain an N-terminal β-barrel sensor domain linked via an α-helix Leu-zipper to a typical HtH domain for DNA binding. The linker region between domains is involved in dimerisation. Ligand binding induces a rotation of the sensory domain β-barrel of one of the monomers, decreasing the flexibility of the whole dimer and positioning the DNA binding domain in a conformation that is compatible with the 3D structure of the palindromic DNA molecule called also "dehalobox" (TTAAT-N4-ATTAA) [77].

Ligand binding occurs with positive cooperativity, so the weak binding of the first ligand molecule leads to the tight binding of the second. The first binding event induces conformational changes that are concerted in both monomers, suggesting the CprK physiological role as a highly sensitive and specialised transcriptional regulator that can activate the transcription of the structural *rdhA* gene in response to different ligand concentrations. Additionally, the study showed that CprK can be reversibly inactivated in the presence of O_2, though the precise physiological relevance is still unknown.

Transcription factors belonging to the Multiple Antibiotic Resistance Regulators (MarR) superfamily control the expression of proteins that are involved in the biochemical and physiological responses to several biotic and abiotic stresses. The proteins thus expressed confer resistance to some of the most widely used antibiotics, like penicillins, lactams, tetracycline, and chloramphenicol, but also to have important roles in regulating

the metabolic response to oxidative stress and in the catabolism of organic solvents, disinfectants and controlling the virulence factor production in response to pathogens [99-101]. Many MarR homologous proteins have been found in the phylogenetic domains of Bacteria and Archaea and its biochemical and structural characterisation has been completed, with about 273 crystal structures with the MarR fold solved.

MarR-type regulators are small proteins between 17-22 kDa, that present an overall α/β fold. In general, these transcriptional factors form triangular homodimers in solution, that bind specific palindromic double-stranded DNA (dsDNA) near the -35 and -10 promoter regions [102, 103]. Their 3D structure exhibits 6 α-helixes and 2 β-strands, interestingly, the residues in both N-terminal and C-terminal domains contribute to the interface of dimerisation. The MarR homologues contain a conserved winged helix-turn-helix (winged HtH) motif that permits the recognition of the target sequence, a dimerisation domain, and often present more than one binding site per dimer for the putative physiological effectors **(Figure 12)**.

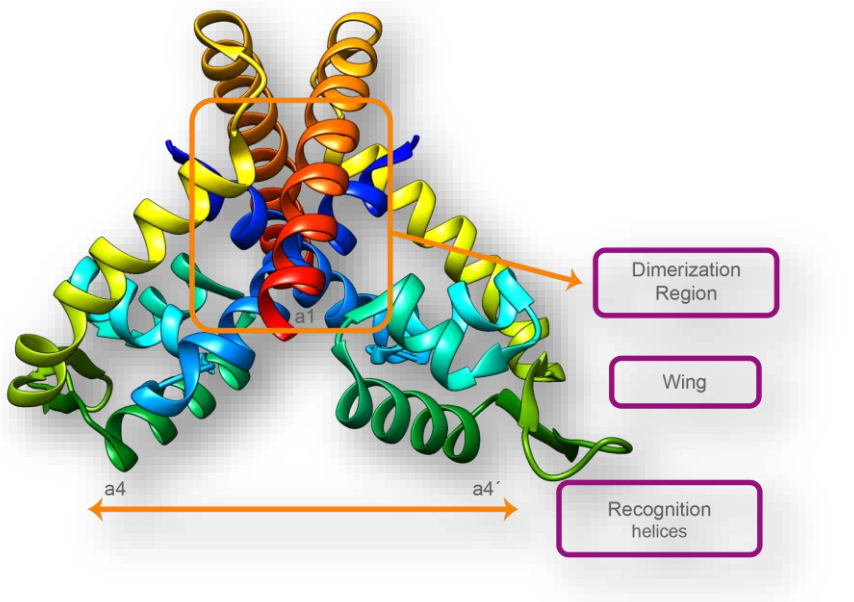


Figure 12. Transcription factor of the MarR superfamily. Cartoon representation of the dimeric SlyA from *Enterococcus faecalis*. The crystal structure was solved at 1.6 Å. The dimerisation interface and the dsDNA recognition helixes are showcased (PDB ID 1LJ9) [104]. Figure generated in Chimera.

The interaction between the dsDNA and the transcription factors *in vivo* is controlled or regulated by the presence of anionic hydrophobic molecules, often aromatic compounds like salicylates, benzoates, and m-chlorophenylhydrazone derivatives [99]. The

MarR transcription factor from *Dehalococcoides* spp. (or DmRdhR) is seemingly a negative response regulator homologous to that encoded by the *marRAB* locus in *Escherichia coli* **[35, 105, 106]**, but its response signal *in vivo* remains unknown. In this thesis, we present our results regarding the binding specificity of the $DmRdhR_{Cbdb1625}$ for certain chlorinated aromatic compounds and its dsDNA palindrome in solution (based on previous work by Krasper *et al.,* and Wagner *et al.*) **[105, 106]**.

1.8 Organohalide respiration and maturation of RdhAs

Until very recently, all RdhAs were thought to participate in respiration or fermentative pathways, attached to the cytoplasmic or periplasmic membranes, coupling electron transport with ATP synthesis via a proton motive force. However, no physiological evidence for this was available until subcellular localisation studies demonstrated that organohalide respiration indeed occurs in the periplasm or the membrane interface, as both the primary electron donor protein and a dimer of the terminal oxidoreductase (respiratory RdhA) are localised in the outer face of the membrane, associated to a protein complex of approximately 250 kDa **[71, 107]**. Although little is known about the exact composition of the respiratory complex in OHRBs and variations are observed in different phyla, its position facing the outer membrane ensures that a chemiosmotic transport can happen. The efficiency of the process can be measured as the ratio of H^+/e^- transferred through the system and it is known that for every halogen equivalent that is substituted just about $\frac{2}{3}$ ATP are formed **[108]**. This, in turn, explains the low growth yields of multiple OHRBs cultures, when respiration of organohalides in presence of H_2 takes place.

Mass spectrometry studies have shown that the respiratory complex is constituted by hydrogenases and reductases that possess multiple Fe-S and other metal cofactors that participate in the electron transfer reactions, but also by membrane anchors that stabilise the complex to the lipid bilayer. RdhAs do not possess transmembrane domains within its primary sequence so they must interact directly with exoplasmic regions of other proteins. It is thought that the small RdhB proteins are responsible for anchoring the RdhA to the membranes. Despite little sequence similarity between RdhBs, in general, these hydrophobic proteins exhibit 2 or 3 transmembrane helixes of variable length (66 – 116 amino acid residues), plus an exoplasmic region where two conserved Glu residues are present **[109]**. Though the *rdhB* gene accompanies most *rdhA* genes, it is not present in all OHRBs operons, as is the case of *Dehalogenimonas* spp. This might indicate that the

membrane anchor is not required by all reductive dehalogenases contained in their genomes or that the stabilisation of the respiratory complex in the lipid bilayer does not depend solely on the RdhB. A new type of reductive dehalogenase was reported by Atashgahi *et al.* [42]. This reductive dehalogenase has an N-terminal transmembrane domain and appears to be a hybrid protein between the RdhB membrane anchor and the RdhA. According to genomic analysis, multiple sequences from Bacterioidetes and δ-Proteobacteria that lack the *rdhB* gene, instead possess this type of hybrid-dehalogenases. Little is known yet about their metabolic function, but its domain structure suggests that it might also have a role in respiration.

Interestingly, the presence of a twin-arginine signal sequence (TAT) that is located in the N-terminal domain of most respiratory RdhAs [110], is implicated in the transport of folded proteins through the cell membrane by specialised molecular machinery in a process called maturation, that was first studied in *Dehalobacter restrictus* for the PceA dehalogenase **(Figure 13)** [36, 48].

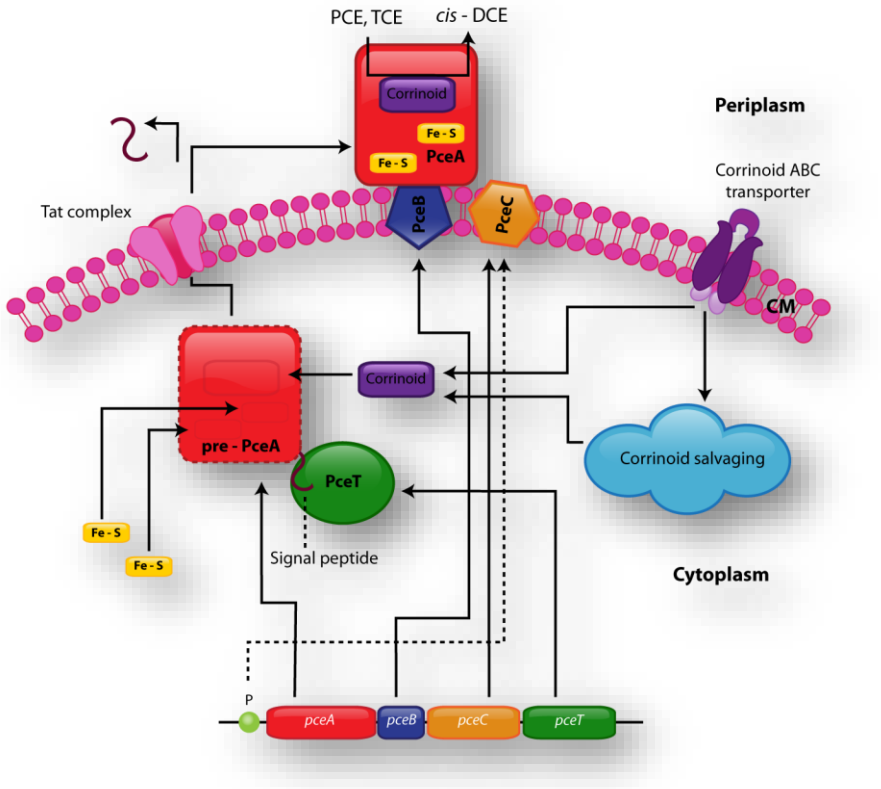


Figure 13. Schematic representation of the respiratory RdhAs maturation process in *D. restrictus*. The process initiates with the translation of PceA, the membrane anchor PceB, a chaperone PceT and the transcriptional regulator PceC, The loading of the cobamide and Fe-S cofactors occurs in the cytosol, followed by the translocation of the pre-PceA by specialised machinery through the cell membrane to the periplasm [36].

RdhAs are synthesised, folded, and loaded with cofactors in the cytoplasm by chaperones (like the RdhT), and then the precursor polypeptides are translocated to the periplasmic space by a set of integral membrane translocases (TatA, TatC, and TatB). The signal peptide (RRXFXK) is then removed by protease activity **[6, 111]**. TAT motifs are hydrophobic peptides of about 50 aa long and common in bacterial proteins that are to be transported to the periplasm or in those that require the incorporation of highly specific and complex metal cofactors **[112]**, like is the case of the Fe-S, cobamides, and porphyrins.

1.9 Electron transfer in organohalide respiration

During respiration, electron transport flows from the electron donor to the terminal acceptor, building a chemiosmotic gradient through the cell membrane that is coupled to the phosphorylation of ADP. Although there are slight variations between different OHRBs and the precise mechanism remains elusive, the involvement of quinone or menaquinone-dependent electron transfer chains or proton pumps during organohalide respiration has been ascertained **[43]**. In the first case, electron transfer depends on the quinone reduction and quinol oxidation[10] reactions. Energy conservation during respiration in organisms like *S. multivorans* does not occur in the RdhA enzyme and instead it is linked to the electron transfer process from the quinol pool in the periplasmic space to the dehalogenase system via a putative quinol-dehydrogenase **[47]**; although the precise identity of the components in this pathway is not known yet, there are some candidates like the NapGH and NrfAH proteins, plus a cytochrome C nitrite reductase and flavoproteins like the Desde_3368, sometimes present in the same or near the operon where the structural gene *rdhA* is encoded. This process might be involved in fermentative metabolic routes in strictly anaerobic bacteria **[93]**. From a thermodynamical perspective, however, there are some inconsistencies given the difference in redox potentials of the quinol/quinone (-74 mV) and the Co (II)/Co (I) (-360 mV) redox pairs. Reverse electron transfer or bifurcation process might explain the generation of low potential electrons to make the process quantitative, even when considering the energetic restraints **[43]**. The presence of extracellular flavin-dependent proteins and the protein FixABCX support the hypothesis **[113]**.

[10] Both menaquinones and ubiquinones biosynthetic genes have been identified in some OHRBs, like *Desulfitobacterium, Sulfurospirillum* and *Dehalobacter* spp. **[47]**.

Some OHRBs lack the set of genes related to the synthesis of quinones and experimental evidence has suggested the lack of quinoid derivatives in cell extracts. So, an alternative to quinone-dependent pathways that instead involves a hydrogenase mediated electron transfer has been suggested. By mass-spectrometry and fingerprinting studies it was determined that a set of proteins that include four hydrogenases (Ni-Fe dependent HupL, Fe-dependent Hym, Ech, and Hyc), plus the Fe-S-Mo-dependent oxidoreductases (OmeA) and a NAD(P)H dependent dehydrogenase, are indeed part of a respiratory complex with the RdhAs in *D. mccartyi*) **(Figure 14)**. It is the HupL enzyme the one that oxidises molecular H_2 and channels the electrons into the system. Although the precise nature of the intermediary steps remains unknown, multiple Fe-S clusters are also implicated. While the HupL does not interact directly with the RdhA dimer, both enzymes are interacting with the Mo-dependent oxidoreductase **[107]**. The complex is stabilised by membrane anchor proteins with multiple transmembrane helixes, like is the case of OmeB. Interestingly, all the proteins that constitute the complex also exhibit TAT sequences before being processed and transported to the outer face of the membrane **[43]**.

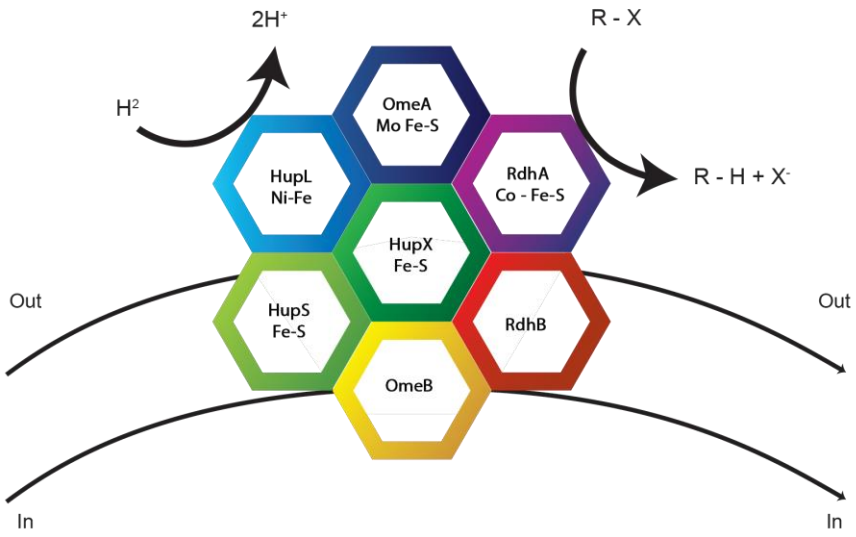


Figure 14. Representation of the hypothetical respiratory complex anchored to the cell membrane in *D. mccartyi*, illustrating the electron transfer process from H_2 to the RdhA protein. The illustration includes the hydrogenases Ni-Fe-dependent HupL plus the Fe-S-Mo-dependent oxidoreductases and a NAD(P)H dependent dehydrogenase depicted as hexagons.

So far, the preference for one-electron transfer mechanism during reductive dehalogenation seems to stem from the different ecophysiology of particular OHRBs and their phylogenetic backgrounds, as it seems to be that only those organisms that belong to the phylum Chloroflexi (and are obligate anaerobes) lack the sets of genes that might imply

a quinone/quinol dependent electron transfer process. This may suggest that perhaps in the non-obligate OHRBs, the electron transfer to the reductive dehalogenase is not directly implicated in a strict respiratory chain and can instead be involved in catabolic pathways to degrade organohalides or fermentative process [47].

1.10 Structural features of RdhAs

Both the respiratory and catabolic reductive dehalogenases display a huge structural diversity, as can be inferred from their low sequence similarity and ecophysiology. Although these proteins share common domains and motifs, particularly regarding their cofactor dependence for both cobamide derivatives and iron-sulfur clusters, not all RdhAs can be described as orthologous. Paralogous RdhAs might be more common than originally expected, given the fact that some of the more recently found sequences exhibit, apart from the duplication and diversification of the central B_{12}-binding domain, additional Fe-S reductase or transmembrane domains similar to the RdhB protein, as was mentioned before, making these new architectures extremely interesting. In addition to this, gene transfer events between organisms that share the same ecological niches but belong to different phyla may also be considered, as some of the RdhAs may be xenologous. Although there is sequence similarity between the cobamide binding central domain and the typical nitroreductase fold [114], there is no evidence to ascertain common ancestry between these protein families.

1.10.1 Crystal structure of the respiratory RdhA

The crystal structure of the respiratory reductive dehalogenase PceA of the microaerophilic ε-proteobacterium *Sulfospirillum multivorans*, a 464 aa (89 kDa) protein, was solved at a resolution of 1.6 Å in presence and without its substrate TCE by Bommer *et al*. [71], showing that the protein is structured in α/β domains. A stable dimer was found in the asymmetric unit with two-fold non-crystallographic symmetry. Although most of the respiratory reductive dehalogenases constitute dimers, this is not the case for the catabolic RdhAs, even when only the crystal structure of the NpRdhA has been solved to date [51]. Solution data supports the claim that indeed the catabolic enzymes are monomeric.

PceA is attached to the periplasmic side of the cytoplasmic membrane, as most respiratory RdhAs and it has been speculated that for both monomers to function

independently, the 2-fold non-crystallographic symmetry axis of the dimer has to be perpendicular to the membrane plane in the complex that PceA forms with PceB [71].

According to Bommer *et al.*, each monomer or protomer of PceA can be divided into an N-terminal unit (1-138), plus the central nor-pseudo-B_{12} binding core (conformed by the residues 139-163 and 216-323), an insertion domain (324 to 394), and the C-terminal unit (395-464) for the binding of Fe-S clusters. The Fe-S are located within 6 Å of the enzyme surface, accessible to the solvent, but the nor-pseudo B_{12} is deeply buried in the structure, stabilised by hydrogen bonds. Two α-helixes from the B_{12}-binding domain and one α-helix from the N-terminal domain contribute to the dimer interface, as well as the loop regions of both terminal domains. Interestingly, in PceA, the distance between the Co atoms in each monomer active site is approximately 42 Å. Electron transfer suitable paths, plus a hydrophobic substrate channel of approximately 12 Å long that leads to the active site were also found in the structure **(Figure 15)**.

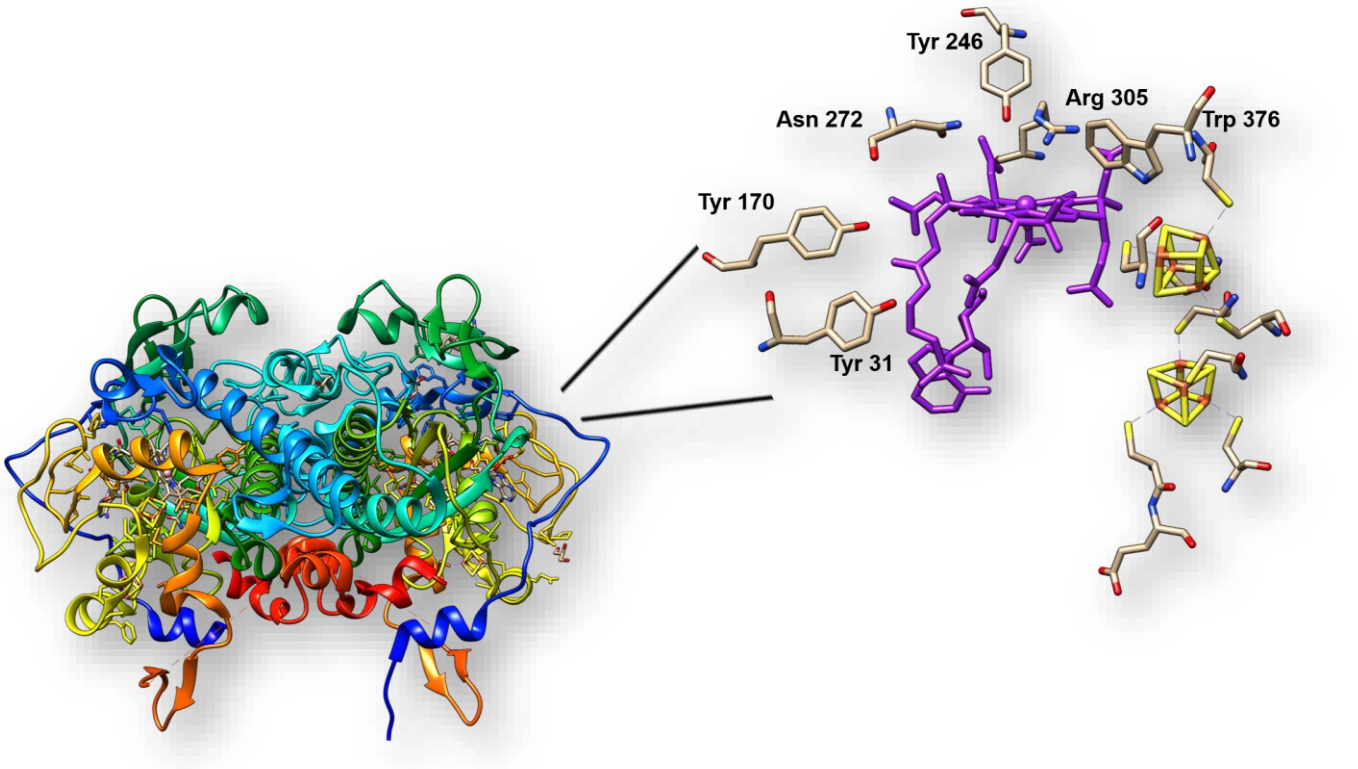


Figure 15. Cartoon representation of the crystal structure of the dimeric PceA (PDB ID 4UQU) of *Sulfospirillum multivorans*, solved at a resolution of 1.6 Å in presence and absence of TCE [71]. The cobalamin binding sites are located in the central core of the structure and a detailed view of the active site residues that are relevant for the catalysis or that stabilise the Fe-S are shown, the Cbl moiety is highlighted in purple. All residues are coloured according to the IUPAC system. Figure generated in Chimera.

The active site cavity is mainly composed of Trp and Tyr residues, making the pocket amphiphilic. The residues Tyr 246, plus Arg 305 and Asn 272 are highly conserved in all RdhAs. The Tyr 246 remains unchanged in all the sequences known to date, but the Arg and Asn residues are often substituted by residues with similar physicochemical properties, like Lys. The side chains of these additional residues face the Cbl-binding site and contribute to the positioning of the Tyr 246 hydroxyl group within hydrogen-bonding distance of the Arg residue, that probably plays a role during catalysis, by facilitating the proton transfer from the Tyr residue and stabilising the intermittent negative charges, according to the mechanistic proposal that will be discussed later. The substrate access to the active site is restricted by a letterbox section of 3.0 per 5.5 Å. The TCE binds the upper axial face of the cobalamin, but it has been found in at least two orientations, with the lone chloride in *cis* or *trans* configurations. The PCE/TCE binding pocket is restricted by Van der Waals contacts with the Tyr 246 that is placed within hydrogen-bonding distance to the Cl substituent **[71]**.

1.10.2 Crystal structure of the catabolic NpRdhA

Previously published results on the structural and kinetical characterisation of the catabolic RdhA from *Nitratireductor pacificus* pht-3B or NpRdhA, revealed that it is closely related to the catabolic reductive dehalogenases BhbA, though it lacks the C-terminal PDR-like domain. NpRdhA catalyses the reduction of *ortho*-halogenated phenols with a clear selectivity for 2,6-dihalogenated phenolic compounds using methyl viologen as electron donor, even in the presence of O_2, as opposed to many respiratory RdhAs. The highest activity of the enzyme was displayed towards the 3,5-dibromo-4-hydroxybenzoic acid, which is reported to be a degradation product of the herbicide.

NpRdhA was overexpressed in *Bacillus megaterium* without losing its activity or its cofactor content in aerobic conditions. It lacks the characteristic twin-arginine signal (TAT) and the associated transmembrane RdhB protein, hence its solubility, in contrast again to the respiratory RdhAs. Crystal structures of the NpRdhA were obtained at 2.3 Å (see **Figure 16**), crystallising in the C2 space group (monoclinic). The crystal structure reveals a globular α/β fold with a central core where the cofactor bindings sites are located. This region, comprising the amino acidic residues between 244–606, is comparable in size to the respiratory RdhA proteins (**Figure 15**, in blue). The core (244-505) is composed of a similar domain to the vitamin B_{12} processing enzyme in humans and a C-terminal domain where two [4Fe–4S] clusters bind to the enzyme. It has been determined that the cobalamin moiety is bound in its base-off form, at the interfacial region between the cobalamin-domain and

the iron-sulfur binding C-terminal segment via the formation of hydrogen bonds with at least 11 aa residues.

The 5,6-DMB is kept in the base-off conformation by water-mediated polar interactions and by four hydrogen bonds that stabilise its binding, as well as an ionic bond or salt bridge between the Lys 542 and the phosphate group and the stacking or π interactions between the benzimidazole ring, Tyr 538 and Pro 461. The upper side of the corrinoid ligand plane is located in a solvent-exposed cavity, whereas the other axial position of the Co complex is not available for ligand binding due to the steric hindrance of the Asp 476 Cβ **[51]**. This protein displays the N-terminal domain (1-243) that was originated during a gene duplication event and further divergence of the B_{12}-binding domain previously described. The C-terminal domain (506 - 694) consists of a region of α helixes and loops very similar to the bacterial ferredoxin domain (**Figure 16**, in orange), where the [4Fe–4S] clusters are bound near the surface of the enzyme at a distance of 9.8 Å between the nearest iron ions, according to structural data. This domain wraps around the functional B_{12}- domain.

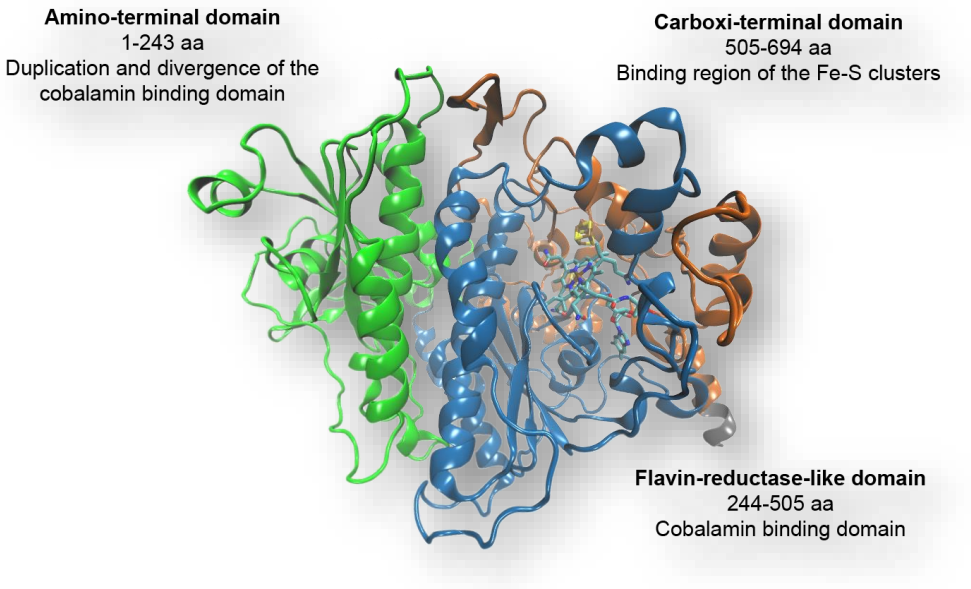


Figure 16. Crystal structure of the NpRdhA (PDB ID 4RAS), shown in different colours according to its domain structure. The overall structure corresponds to a globular α/β fold. Figure generated in VMD 1.9.2.

Both [4Fe–4S] clusters are bound to the enzyme in different ways at approximately 9.8 Å; the first [4Fe–4S] cluster is located at approximately 12.6 Å of the cobalamin, interacting with three Cys residues of the common bacterial ferredoxin motif CXXCXXCXXXCP, while the second [4Fe–4S] cluster (often substituted by a [3Fe–4S] in some

RdhAs due to a variable insertion between two of the first Cys in the motif) is in direct Van der Waals contact with one of the pyrrolic rings of the cobalamin via the Cys 548 of the iron-sulfur binding motif. This is a remarkable feature of the RdhA **(Figure 17)** and QueG enzymes **(Figure 18)**, as in most enzymes the [4Fe–4S] rarely are near enough to the corrin derivatives or other cofactors as porphyrins [51].

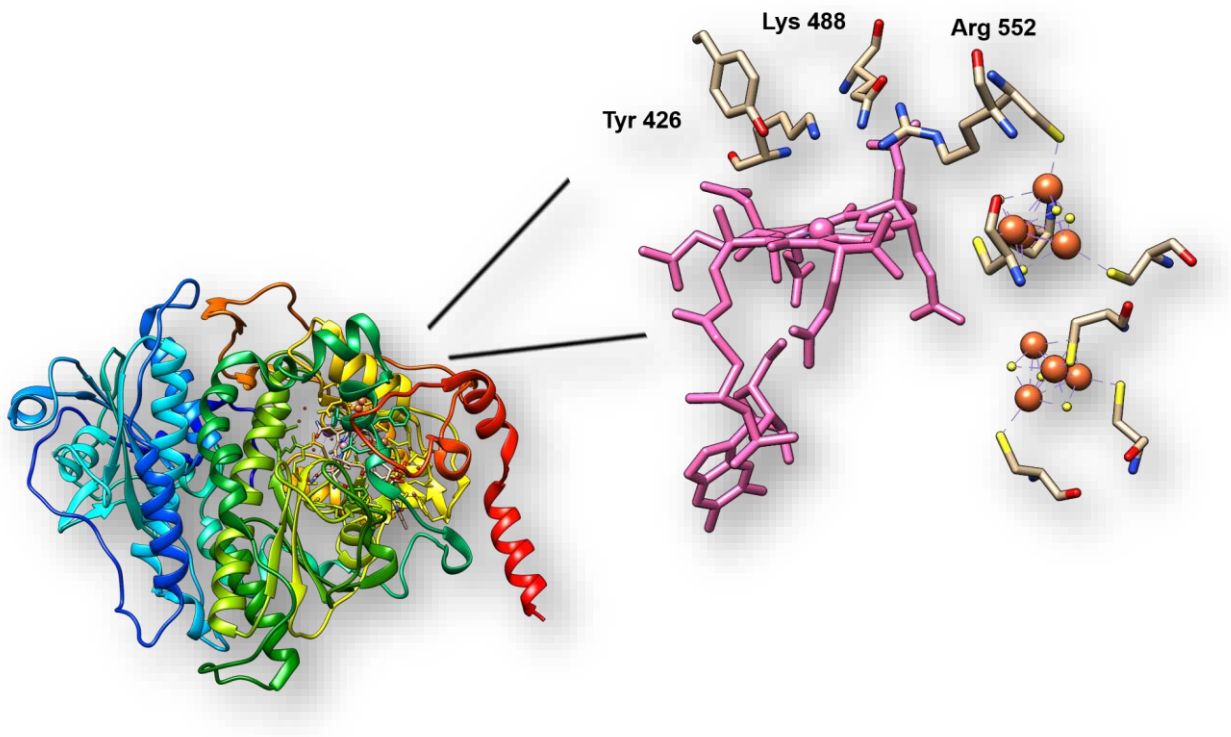


Figure 17. Cartoon representation of the crystal structure of the NpRdhA solved at 2.3 Å (PDB ID 4RAS). The cobalamin binding site, located in the central core of the structure, is shown in detail and the Cbl is highlighted in pink, while some of the active site residues, that participate in the catalysis or substrate stabilisation are shown coloured according to the IUPAC system for heteroatoms. Figure generated in Chimera.

As no crystal structures were obtained of the NpRdhA with the substrate 3,5-dibromo-4-hydroxybenzoic acid, computational docking of the ligand helped to clarify that the active site is located above the cobalamin plane, exposed to the solvent molecules and that the aromatic plane of the aryl halide is perpendicular to the corrin ring moiety. The hydroxyl group of the substrate is stabilised by hydrogen bonds with the Ser 422, Arg 488 and Arg 552 residues, suggesting that it will preferably bind in a deprotonated state. More interestingly, it also shows that one of the halogen substituents can be placed near the Co ion and the side chains of the Tyr 426, Lys 488, and Arg 552 residues, whereas the second bromide can be accommodated in a hydrophobic cavity. This feature suggests that a Co-X bond (and not a Co-C as in other cobalamin-dependent enzymes) can be formed during catalysis. The interaction between Co-X determined *in silico* was confirmed experimentally

by EPR spectroscopy. Further analysis of the metal-halogen bond between de Co (II) of the cobalamin and bromide was made using DFT calculations. It was established that the possible reduction of the Co (II) to Co (I) can result in the elongation (or rupture) of the C-Br bond in the substrate and the contraction (or formation) of the Co-Br bond; this suggests mechanistic information on the catalytic behaviour of the NpRdhA.

Box 3.0 Epoxyqueuosine reductase (QueG) structure

Queuosine (Q) is a modified nucleoside, one of many post-transcriptional modifications often found in the wobble position of the anticodon of Asp, Asn, His, and Tyr tRNAs in Bacteria and Eukarya domains, however, only Bacteria have been reported to be able to synthesise it. The last enzyme participating in the biosynthetic pathway of Q is the epoxyqueuosine reductase (QueG) that catalyses the reduction and dehydration of an epoxyqueuosine (oQ) moiety when bound to the tRNA. QueG is a Cbl-dependent enzyme, homologous to the RdhAs and the Cbl-C enzyme. The structure possesses three domains: an N-terminal Cbl-binding domain, a ferredoxin-like Fe-S binding domain, and the C-terminal tRNA binding domain, which is similar to the HEAT domain specialised in protein-nucleic acid interactions. In QueG, both the Fd and the Cbl-binding domains present insertions that create a scaffold where the tRNA binding site is formed. The position of both 4Fe-4S is comparable to that of the clusters found in the RdhAs, suggesting a similar role during the Cbl reduction. A structure of QueG from *Bacillus subtilis* with Q-tRNA was solved at 2.1 Å by Dowling *et al.* **[73]** shows that at least 8 nucleotides from the Tyr tRNA contribute to the crystal packing; the Q moiety has been bound placed directly in the upper axial position of the Cbl plane at approximately 4.0 Å, a distance that is expected to decrease in case of the substrate oQ. The structure shows that there are no large-scale conformational changes upon the substrate binding, as opposed to Ado and Me-Cbl-dependent enzymes.

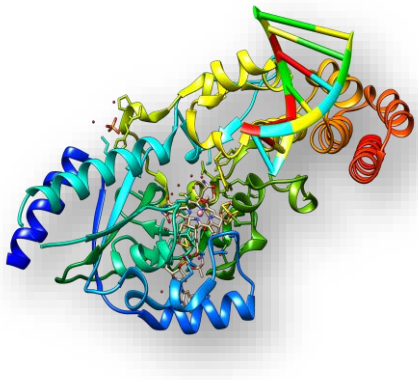


Figure 18. Structure of QueG from *Bacillus subtilis,* (PDB ID 5D08) crystallised with a fragment of 8 bp of Tyr-tRNA and solved at 2.1 Å by Dowling *et al.* **[73]**. Figure generated in Chimera.

1.11 Catalytic mechanism of the reductive dehalogenases RdhAs

The initial mechanistic proposals for the dehalogenation of organohalides were developed to explain the reduction of alkyl halides. Originally, it was hypothesised that the dehalogenation of perchloroethylene PCE could either proceed via the formation of an organocobalt adduct or by a single electron transfer from the Co (I), but there is no experimental support of either the formation of organic radicals or organometallic intermediates between the Co ion and the C atom **[36, 48, 91, 115]**.

Therefore, the current mechanistic proposal for the reductive dehalogenation reaction catalysed by RdhAs is quite different from other cobalamin-dependent enzymes, as it does imply the possibility of a direct attack of the metal centre of the cobalamin to the X atom of the substrate. The oxidative addition of the halogen atom to the Co may be a unique feature of the RdhAs and QueG enzymes, and it now constitutes an entirely new catalytic behaviour, both in terms of the cobalamin cofactor reactivity and because of its relevance in organohalide and supramolecular chemistry **[71, 116]**. This third mechanism is supported by the experimental results on the NpRdhA and the PceA from *Sulfurospirillum multivorans*.

The metal-halogen single and secondary bonds were also recently characterised in *ortho*-chloro and *ortho*-bromophenols coordinated to divalent transition metals **[117]**. The formation of this type of interactions may be responsible for the substrate specificity in a number of enzymes like the PceA **[118]**. The chemistry and electronic features of metal-halogen bonding are now being widely studied, particularly for its possible applications in supramolecular chemistry **[116]**. Metal-halogen bonding can be defined as the interaction of a metallic species acting as a Lewis base and halogen X acting as a Lewis acid. In cob(I)alamin, the interaction occurs between the Co (I) occupied dz^2 orbitals and the substrate's apical σ-hole **[116]**. Sigma holes are non-covalent interactions between the lone pair of a Lewis base and atoms from the groups' IV, V, and VI **[119]**. Halogen atoms are polarisable and can exhibit an anisotropic charge distribution, in secondary metal-halogen bonds a transient positively charged region or charged crown, interacts with a Lewis base (like the organic moiety), while the negatively charged region interacts with the positive charge of a transition metal ion.

The overall proposal for the mechanism of dehalogenation catalysed by NpRdhA can be summarised as follows (see **Figure 19**):

- The substrate binds to the active site and the halogen atom replaces the axial ligand of the Co (II) centre. In the crystal structure of NpRdhA the Cl^- anion is the best-modelled ligand according to electron density assignments, but a water molecule is also plausible **[51]**.

- The C-X bond can be broken via homolytic or heterolytic mechanisms: in the homolytic proposal, the formation of an aryl radical by an electron transfer from the Co (I) or the nearest [4Fe–4S] cluster is required (see **Figure 19**).

- The heterolytic mechanism involves the attack of the cob(I)alamin to the X atom of the organohalide and the formation of an X-cob(III)alamin transient complex. The attack over the X atom leads to the protonation of the leaving group (in this case the aryl moiety) and the concomitant rupture of the C-X bond. The C-X cleavage occurs while the second electron transfer from the [4Fe–4S] to the metal centre regenerates the cob(II)alamin complex.

Additional considerations for the heterolytic mechanism, which are not yet supported by experimental evidence and therefore are intuitive or speculative, are:

  - The C-X cleavage can occur in a concerted manner with the protonation of the substrate during the additive oxidation of the Co and before the second electron transfer from the Fe-S to the metal centre. Proton donor species are present in the active site, being the Tyr 426 residue the most probable proton donor.

  - The anionic species formed in the substrate can be stabilised either by resonance or shielding effects coming from positively charged residues in the binding site, like Lys 488 and Arg 552.

  - The substitution of the Br^- in the axial position of cob(III)alamin complex by another ligand and therefore the regeneration of cob(II)alamin by the second single electron transfer might occur in the presence of a stronger ligand than water and the halogen itself.

The residues Tyr 426 and Lys 448 (that stabilise the nearest [4Fe–4S] cluster, as described previously) apparently play an important role during catalysis, as mutation abolishes enzyme activity **[51, 119]**.

It is speculated that the reaction mechanism may change according to the structural and electronic features of the substrates **[116]**. The heterolytic mechanistic proposal is perhaps the most plausible mechanism of reaction given the type of intermediates that can be formed, in accordance to energy calculations made by Liao, *et al.* **[120]** using DFT and

QM/MM hybrid approaches for analysing the dehalogenation reaction of 3,5-dibromo-4-hydroxybenzoic acid carried out by NpRdhA by assigning three different oxidation states to the metal centre.

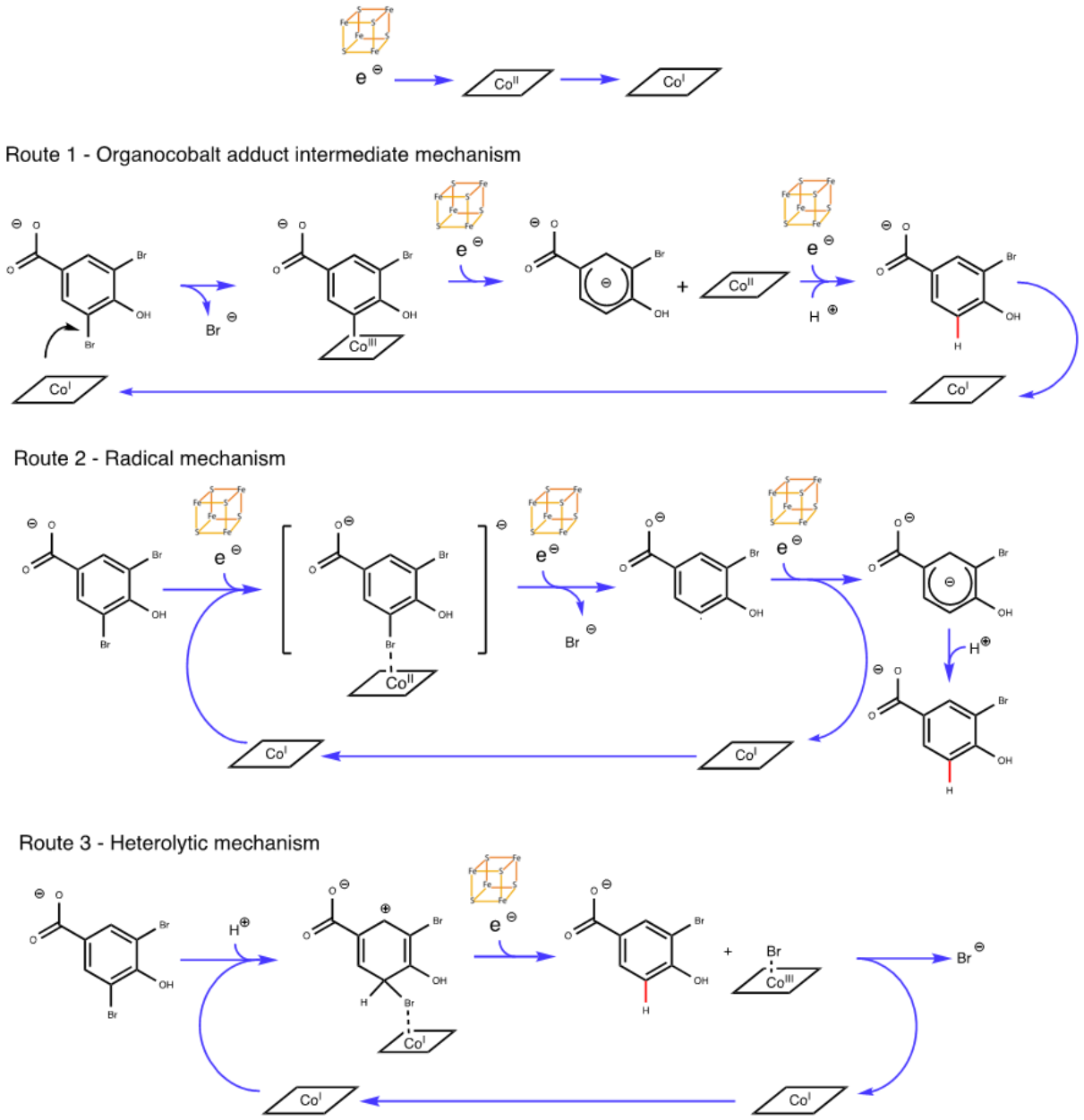


Figure 19. Proposals for the reaction mechanism of reductive dehalogenation. These reaction mechanisms illustrate the formation of an organocobalt adduct intermediate (Route 1), the long-range electron transfer from the Co(I) or the Fe-S that leads to a substrate radical, followed by the formation of a carbanion after the elimination of the halogen substituent (Route 2), and finally, the formation of a metal-halogen bond by nucleophilic attack of the Co(I), tailed by the heterolytic cleavage of the Co-X bond (Route 3).

In silico results also show that the dehalogenation reaction cannot proceed via the formation of a Co (0) intermediate and that the formation of the cob(I)alamin nucleophile

and the H^+ transfer from Tyr 426 lower the energetic barrier with the formation of a stable tetrahedral transition state. Finally, they conclude that enzymatic dehalogenation reactions follow the trend of C-X bond cleavage: I>Br>Cl, with the energy barrier for the defluorination reaction too high for the reaction to occur.

1.12 Catabolic reductive dehalogenases

The search for more catabolic RdhAs sequences depends on metagenomic analysis of microbial cultures, particularly from marine ecosystems (including coastal or tidal and deep-sea habitats) where aerobic (or facultative) bacteria from the superphylum Proteobacteria and Bacteroidetes thrive **[121, 122]**. Aquatic Proteobacteria contain an array of Gram-negative bacteria that can be further subdivided in α, γ, δ, and ε classes); these microorganisms have evolved metabolic routes that involve halogenation and dehalogenation reactions to produce secondary metabolites or degrade them, as well as biosynthetic routes of unusual heterocycles and peptides **[123]**. Although little is known about the catabolic reductive dehalogenases physiological and metabolic role, a chemiosmotic coupling of the reductive dehalogenation process to the production of ATP has been ruled out, and instead of a degradative pathway in which the removal of the halogen substituent renders the reduced product as a carbon source is suspected. The products of reductive dehalogenation during respiration are generally not fully degraded in anaerobes, in contrast to aerobes **[124]**.

There are few examples of catabolic reductive dehalogenases known to date; these enzymes are different to their respiratory counterparts in many respects, being the absence of the TAT signal peptide and sometimes the lack of the membrane anchor RdhB the most obvious ones; its basic domain structure is depicted also in the schematic representation of **Figure 10**. The lack of the *rdhB* gene within the same operon as the *rdhA* hints to the fact that these proteins might be not attached to the membrane as part of a complex and localised in the cytosol instead. Perhaps the most important feature of the catabolic RdhAs is the presence of an extra but non-functional B_{12}-binding domain located at the N-terminal. This domain of about 250 aa is probably the result of a gene duplication event, which indicates that these proteins have diverted from their probably ancestral respiratory counterparts. Bioinformatic analysis has helped to determine that some catabolic RdhAs also contain a C-terminal reductase domain, that is similar to the iron-sulfur (Fe-S) flavoprotein phthalate dioxygenase reductase (PDR). This additional domain is suspected

to allow them to transfer electrons intramolecularly from pyridine nucleotides like NAD(P)H to FMN, a simple 2Fe-2S cluster and then to the pair of 4Fe-4S clusters and finally to the cobamide in the active site of the RdhA domain, in a self-sufficient way, without the need of external reductase systems to provide electrons. This remarkable fused architecture has been called a self-sufficient reductive dehalogenase (ssRdhA).

The only example of a self-sufficient system was characterised by Chen *et al.* [50]; the RdhA from *Comamonas* sp. 7D-2 **(Figure 10)** also called BhbA, that catalyses the debromination of the herbicide bromoxynil. The degradation pathway of bromoxynil to 4-carboxy-2-hydroxymuconate-6-semialdehyde was investigated *in vivo* and in addition to the RdhA, nitrilases, monooxygenases, and dioxygenases were identified **(Figure 20)**. Homologous proteins to the BhbA are distributed in the superphylum Proteobacteria, both as full-length and natural truncated versions of the protein from *Nitratireductur pacificus* pht-3B, structurally characterised recently [51]. Structural information obtained by Payne *et al.* [51] shows that NpRdhA is a monomer in solution and therefore it is thought that most catabolic reductive dehalogenases are. Catabolic RdhAs are much larger enzymes in comparison to the respiratory RdhAs, with molecular weights between 75-150 kDa. Preliminary data hints at the possibility of these enzymes being O_2 tolerant, in contrast to the respiratory proteins.

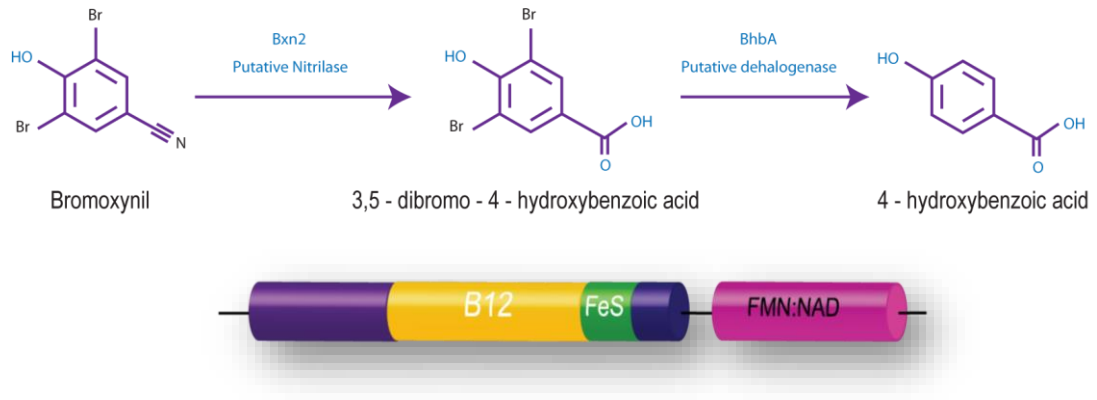


Figure 20. Catabolism of bromoxynil in *Comamonas 7D-2.* The activity of the putative bromoxynil-specific nitrilase Bxn2 is shown as the first step of the degradation pathway, followed by the reductive dehalogenation of the 3,5-Br-4-hydroxybenzoic acid by the self-sufficient BhbA. The domain structure of the catabolic full-length or self-sufficient RdhAs is presented here for convenience.

It is probable that this catabolic type of self-sufficient architectures will offer a more direct possibility to study the mechanistic details of reductive dehalogenation, since there is no requirement for exogenous redox partners, as is the case of the naturally truncated

protein NpRdhA, that lacks the PDR-like reductase domain and therefore depends on ferredoxin and other reductases to shuttle electrons to the active site cobalamin **[92, 125]**.

Until more genomic and metabolomic data is gathered and better strategies for heterologous expression of these RdhAs are explored, the biochemical information in the topic is limited, but still, the catabolic RdhAs constitute an excellent research target, that poses new challenges and interesting questions, like:

- Are the catabolic reductive dehalogenases interacting with other proteins to form complexes in a similar manner to their respiratory counterparts?
- Do they interact with membrane anchors to be further stabilised?
- What is the ecophysiological and biochemical role of the self-sufficient catabolic RdhAs?
- Can we heterologously express and purify these proteins in large amounts to be characterised by biochemical methods?
- Are these proteins a suitable target for X-ray crystallography and Cryo-EM?
- Will the presence of the PDR-like reductase domain help us to obtain mechanistic information in a more direct way?
- Which is the rate-limiting step for the catalysis: the electron transfer process or the formation of the E-S complex?
- What is the substrate specificity of these proteins, are they promiscuous?
- How, when, and why did these fusion proteins evolve?

The main objective of this thesis is to achieve the initial biochemical characterisation of at least one homologue of the self-sufficient RdhA from aerobic marine α-Proteobacteria. The particular objectives of this project will be presented in the next section.

1.2 Overview and objectives of the project

Research on reductive dehalogenases has been historically hindered by the many challenges of expressing soluble and active enzymes in quantities that allow its characterisation by biophysical methods, not only because of the intrinsic difficulties of dealing with membrane proteins or with those that depend on the interaction with a complex to be stabilised, but also because of the many cofactor requirements of the RdhAs **[36, 43]**. Previous efforts at expressing the respiratory reductive dehalogenases in the native OHRB organisms faced different problems derived from the slow growth rates of strict anaerobic bacteria when grown in isolation from other microorganisms, as is the case of *Dehalococcoides* spp., that has doubling times of nearly four days when cultured in the

laboratory **[79]**, and that the media must be supplemented with numerous nutrients, electron donors, and vitamins to ensure growth.

Since *Escherichia coli* commercial strains are cobalamin auxotrophic (unless the BtuB cobalamin transporter is included by transformation), overexpression of these B_{12}-dependent enzymes with other systems as *Bacillus megaterium*, had to be considered. After the success story with NpRdhA in our group **[51]**, a natural truncated catabolic RdhA using *B. megaterium*, we decided to express several orthologous proteins to the first self-sufficient system characterised *in vivo* by Chen *et al.* **[50]** from *Comamonas* sp. 7D-2. For that purpose we selected various homologues with low sequence similarity (30 to 50 %) that belong to marine aerobic α and β Proteobacteria phylum: *Ottowia thiooxydans* DSM14619, *Salinarimonas rosea* DSM21201, *Ruegeria pomeroyi* DSS3, *Tropicibacter phthalicicus, Pseudovibrio* sp. FO-BEG, and *Comamonas* sp. 7D-2), all encoding RdhAs that exhibit a fused PDR-like flavoprotein domain that is similar to the already characterised phthalate dioxygenase reductase, that catalyses two-electron transfer reactions utilising NAD(P)H **[126]**.

The overexpression of soluble and fully complemented active protein will enable us to accomplish the biochemical characterisation of the first catabolic full-length RdhA in a self-sufficient process without requiring the presence of redox partners to transfer electrons to the cob(II)alamin during the catalysis, as was shown by Collins *et al*. **[125]**, with the ferredoxin from *Spinacea olereacea* and the flavodoxin reductase from *Escherichia coli*. For this reason, there was a huge interest in designing a robust methodology for the heterologous production of a self-sufficient enzyme; we approached the subject by producing different constructs, with affinity and solubility tags that are suitable for expression in various hosts, plus an iterative approach at purification strategies to tackle the many issues encountered when first dealing with the ssRhdAs until we managed to produce a stable and active protein with good yields. We summarise these efforts in Results Chapter II.

In line with the original idea of producing a self-sufficient RdhA reductive dehalogenase and given the experimental challenges of expressing the natural self-sufficient proteins, a chimeric fusion of the in-house dehalogenase NpRdhA and the PDR-like reductase from *Pseudovibrio* sp. FO-BEG was produced. Incidentally, as part of this work, we also proved that the natural truncated reductase domains are able to transfer electrons to the cubane Fe-S clusters and the cob(II)alamin in the dehalogenase active site,

suggesting that *in vivo*, the PDR-like reductase might act as a physiological redox partner for the truncated RdhAs enzymes. This work is presented in Results Chapter III.

As a complementary project, we continued previous studies, started by Carolina P. Quezada, on the $RdhR_{Cbdb1625}$ a transcriptional regulator of the Multiple Antibiotic Resistance (MarR) superfamily, from *Dehalococcoides mccartyi* strain CBDB1, that dehalogenates some of the most recalcitrant aromatic organohalides. For that purpose, we determined the binding parameters for various constitutional isomers of 1,2,3-trichlorophenols. We also tested the effect of these putative physiological effectors in the recognition of its 42 bp palindromic dsDNA sequence, determining that indeed the $RdhR_{Cbdb1625}$ acts as a repressor of transcription. We obtained crystals structures of both the free and ligand-bound $RdhR_{Cbdb1625}$. Our results are presented in Results Chapter I.

In summary, I hope that this work might be a stepping stone for the group and that in the next years the relationship between the structural features and biochemical function of the self-sufficient RdhAs might be understood, shedding some light on mechanistic details of the catalysis, complementing our knowledge of yet another B_{12}-dependent system **[74]**, and finally providing us with the answer as to which cobalamin intermediate participates in the reduction of organohalides. A full characterisation might also facilitate subsequent studies focused on directed evolution, that perhaps will enable us to produce novel enzymes and the possibility to export the active site features to new molecular architectures. Also, by achieving the overexpression in heterologous systems, some efficient and competitive bioremediation strategies could be developed to treat highly contaminated sites **[36]**, thus helping us to diminish the impact of anthropogenic pollution in soils, lotic, lentic, and marine ecosystems; this will also be translated in the improvement of the quality of life in human populations living near heavily contaminated places throughout the world.

References

1. J. C. Theiss, M. B. Shimkin and L. A. Poirier, *Cancer Res.*, 1979, **39**, pp. 391-395.
2. K. H. Engesser and P. Fischer, in *Biodegradation: Natural and Synthetic Materials*, ed. W. B. Betts, Springer, London, 1991, ch. 2, pp. 15-54.
3. G. W. Gribble, *Am. Sci.*, 2004, **92**, pp. 342-349.
4. P. Adriaens and A. L. Barkovskii, in *Encyclopedia of Environmental Microbiology*, ed. G. Bitton, Wiley-Interscience, New York, 2003, p. 509.
5. S. Fetzner and F. Lingens, *Microbiol. Rev.*, 1994, **58**, pp. 641-685.
6. H. Smidt and W. M. d. Vos, *Annu. Rev. Microbiol.*, 2004, **58**, pp. 43-73.
7. A. Srivastava, in *Alkyl Halides and Aryl Halides*, NISCAIR, New Delhi, 2014.
8. J. Dolfing and I. Novak, *Biodegradation*, 2015, **26**, pp. 15-27.
9. L. C. Blasiak and C. L. Drennan, *Acc. Chem. Res.*, 2009, **42**, pp. 147-155.
10. R. Banerjee and S. W. Ragsdale, *Annu. Rev. Biochem.*, 2003, **72**, pp. 209-247.
11. L. G. Wade, *Organic Chemistry*, Pearson, London, 8th edn., 2013.
12. J. E. Huheey, E. A. Keiter and R. L. Keiter, *Inorganic Chemistry: principles of structure and reactivity.*, Oxford University Press, México, 4th edn., 2005.
13. G. Öberg, *Applied Microbiology and Biotechnolgy*, 2002, **58**, pp. 565-581.
14. W. Fenical, in *Marine Organic Chemistry: Evolution, Composition, Interactions and Chemistry of Organic Matter in Seawater*, eds. E. K. Duursma and R. Dawson, Elsevier, London, 1981, vol. 31, ch. 12, pp. 375-393.
15. G. W. Gribble, *Chemosphere*, 2003, **52**, pp. 289-297.
16. V. Agarwal, Z. D. Miles, J. M. Winter, A. S. Eustáquio, A. E. Gamal and B. S. Moore, *Chem. Rev.*, 2017, **117**, p. 5619−5674.
17. A. E. Gamal, V. Agarwal, S. Diethelma, I. Rahman, M. A. Schorna, J. M. Sneedb, G. V. Louiec, K. E. Whalene, T. J. Mincere, J. P. Noelc, V. J. Paulb and B. S. Moore, *PNAS*, 2016, **113**, pp. 3797–3802.
18. L. Fowden, R. Robinson and N. W. Pirie, *Proc. Royal Soc. B*, 1968, **171**, pp. 5-18.
19. A. K. Croft, W. Groenewald and M. S. Tierney, in *Marine Bioactive Compounds: Sources, Characterization and Applications*, ed. M. Hayes, Springer, New York, 1 edn., 2012, ch. 7, p. 232.
20. S. Emerson and J. Hedges, in *Chemical Oceanography and the Marine Carbon Cycle*, Cambridge University Press, Cambridge, 12 edn., 2008, ch. 12, pp. 404-444.
21. D. C. Lohman, D. R. Edwards and R. Wolfenden, *J. Am. Chem. Soc.*, 2013, **135**, pp. 14473-14475.
22. A. Chatt, presented in part at the Research co-ordination meeting on use of nuclear and related analytical techniques in studying human health impacts of toxic elements consumed through foodstuffs contaminated by industrial activities, Vienna, Austria, 18-22 March, 2002.
23. E. D. Goldberg, *Proc. Royal Soc. B*, 1975, **189**, pp. 277-289.
24. C. Dong, F. Huang, H. Deng, C. Schaffrath, J. B. Spencer, D. O'Hagan and J. H. Naismith, *Nature*, 2004, **427**, pp. 561–565.
25. T. E. Graedel and W. C. Keene, *Global Biogeochem. Cycles*, 1995, **9**, pp. 47-77.
26. W. R. Simpson, S. S. Brown, A. Saiz-Lopez, J. A. Thornton and R. V. Glasow, *Chem. Rev.*, 2015, **115**, pp. 4035-4062.
27. E. J. Bouwer, B. E. Rittmann and P. L. McCarty, *Environ. Sci. Technol.*, 1981, **15**, pp. 596-599.
28. M. S. McLachlan, *Chemosphere*, 1997, **34**, pp. 1263-1276.
29. V. Tornero and G. Hanke, *Mar. Pollut. Bull.*, 2016, **112**, pp. 17-38.
30. A. J. Jamieson, T. Malkocs, S. B. Piertney, T. Fujii and Z. Zhang, *Nat. Ecol. Evol.*, 2017, **1**. 0051.
31. B. Hens and L. Hens, *Toxics*, 2018, **6**. 1.

32. A. Strid, I. Athanassiadis, M. Athanasiadou, J. Svavarsson, O. Papke and A. Bergman, *Environ. Toxicol. Chem.*, 2010, **29**, pp. 2653–2659.
33. J.-P. Desforges, A. Hall, B. McConnell, A. Rosing-Asvid, J. L. Barber, A. Brownlow, S. D. Guise, I. Eulaers, P. D. Jepson, R. J. Letcher, M. Levin, P. S. Ross, F. Samarra, G. Víkingson, C. Sonne and R. Dietz, *Science*, 2018, **361**, pp. 1373–1376.
34. R. J.Letcher, J. O. Bustnesb, R. Dietz, B. M. Jenssend, E. H. Jørgenseneh, C. Sonne, J. Verreault, M. M.Vijayan and G. W.Gabrielsen, *Sci. Total Environ.*, 2010, **408**, pp. 2995-3043.
35. D. Leys, L. Adrian and H. Smidt, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120316.
36. B.-E. Jugder, H. Ertan, M. Lee, M. Manefield and C. P. Marquis, *Trends Biotechnol.*, 2015, **33**, pp. 595-610.
37. A. M. P. Escobar, *Apuntes de Investigación*, 2015, **11**. 2.
38. J. Liu and M. M. Häggbloma, *Mbio*, 2018, **9**. e02471-18.
39. M. Giedyk, K. Goliszewska and D. Gryko, *Chem. Soc. Rev.*, 2015, **44**, pp. 3391-3404.
40. R. M. deJong, P. Bazzacco, G. J. Poelarends, W. H. J. Jr., Y. J. Kim, E. A. Burks, H. Serrano, A. M. W. H. Thunnissen, C. P. Whitman and B. W. Dijkstra, *J. Biol. Chem.*, 2007, **282**, pp. 2440-2449.
41. A. E. Gamal, V. Agarwal, I. Rahman and B. S. Moore, *J. Am. Chem. Soc.*, 2016, **138**, pp. 13167–13170.
42. S. Atashgahi, Y. Lu, Y. Zheng, E. Saccenti, M. Suarez-Diez, J. Ramiro-Garcia, H. Eisenmann, M. Elsner, A. J. M. Stams, D. Springael, W. Dejonghe and H. Smidt, *Environ. Microbiol.*, 2017, **19**, pp. 968–981.
43. T. Schubert, L. Adrian, R. G. Sawers and G. Diekert, *FEMS Microbiol. Ecol.*, 2018, **94**. fiy035.
44. F. Maphosa, S. H. Lieten, I. Dinkla, A. J. Stams, H. Smidt and D. E. Fennell, *Front. Microbiol.*, 2012, **3**. 351.
45. J. Dolfing and D. B. Janssen, *Biodegradation*, 1994, **5**, pp. 21-28.
46. M. J. Krzmarzick, D. K. Taylor, X. Fu and A. L. McCutchan, *Hindawi*, 2018, **2018**. 3194108.
47. M. Fincker and A. M. Spormann, *Annu. Rev. Biochem.*, 2017, **86**, pp. 357–386.
48. J. Maillard, W. Schumacher, F. Vazquez, C. Regeard, W. R. Hagen and C. Holliger, *Appl Environ Microbiol.*, 2003, **69**, pp. 4628–4638.
49. A. Suyama, M. Yamashita, S. Yoshino and K. Furukawa, *J. Bacteriol.*, 2002, **184**, pp. 3419-3425.
50. K. Chen, L. Huang, C. Xu, X. Liu, J. He, S. H. Zinder, S. Li and J. Jiang, *Mol. Microbiol.*, 2013, **89**, pp. 1121–1139.
51. K. A. P. Payne, C. P. Quezada, K. Fisher, M. S. Dunstan, F. A. Collins, H. Sjuts, C. Levy, S. Hay, S. E. J. Rigby and D. Leys, *Nature*, 2014, **517**, pp. 513-516.
52. L. A. Hug, F. Maphosa, D. Leys, F. E. Loffler, H. Smidt, E. A. Edwards and L. Adrian, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120322.
53. K. A. P. Payne, K. Fisher, H. Sjuts, M. S. Dunstan, B. Bellina, L. Johannissen, P. Barran, S. Hay, S. E. J. Rigby and D. Leys, *J. Biol. Chem.*, 2015, **290**, pp. 27572-27581.
54. J. Kim, C. Gherasim and R. Banerjee, *PNAS*, 2008, **105**, pp. 14551-14554.
55. A. Parthasarathy, T. A. Stich, S. T. Lohner, A. Lesnefsky, R. D. Britt and A. M. Spormann, *J. Am. Chem. Soc.*, 2015, **137**, p. 3525–3532.
56. C. Kunze, M. Bommer, W. R. Hagen, M. Uksa, H. Dobbek, T. Schubert and G. Diekert, *Nat. Commun.*, 2017, **8**. 15858.
57. W. Maret, *Metallomics: A primer of integrated Biometal Sciences*, Imperial College Press, London, 2016.
58. D. Lindsay and W. Kerr, *Nat. Chem.*, 2011, **3**, p. 494.
59. E. Neil and G. M. Marsh, *Essays Biochem.*, 1999, **34**, pp. 139-154.
60. A. S. Rury, T. E. Wiley and R. J. Sension, *Acc. Chem. Res.*, 2015, **48**, p. 860–867.
61. I. A. Dereven'kov, D. S. Salnikov, R. Silaghi-Dumitrescu, S. V. Makarov and O. I. Koifman, *Coord. Chem. Rev.*, 2016, **309**, pp. 68-83.

62. B. Krautler, *Biochem. Soc. Trans.*, 2005, **33**, pp. 806-810.
63. K. L. Brown, *Chem. Rev.*, 2005, **105**, pp. 2075-2149.
64. N. Assaf-Anid, K. F. Hayes and T. M. Vogel, *Environ. Sci. Technol.*, 1994, **26**, pp. 246-252.
65. N. Metzler-Nolte, in *Comprehensive Organometallic Chemistry III. From Fundamentals to Applications*, eds. D. M. P. Mingos and R. H. Crabtree, Elsevier Science, New York, 2007, vol. 1, ch. 1.31, pp. 883-920.
66. R. J. Kutta, S. J. O. Hardman, L. O. Johannissen, B. Bellina, H. L. Messiha, J. M. Ortiz-Guerrero, M. Elías-Arnanz, S. Padmanabhan, P. Barran, N. S. Scrutton and A. R. Jones, *Nat. Commun.*, 2015, **6**. 7907.
67. M. Jost, J. Fernández-Zapata, M. C. Polanco, J. M. Ortiz-Guerrero, P. Y.-T. Chen, G. Kang, S. Padmanabhan, M. Elías-Arnarnz and C. L. Drennan, *Nature*, 2015, **526**, pp. 536-541.
68. K. Gruber, B. Puffer and B. Krautler, *Chem. Soc. Rev.*, 2011, **40**, pp. 4346-4363.
69. V. Bandarian and R. G. Matthews, in *Methods in Enzymology*, Academic Press, 2004, vol. 380, pp. 152-169.
70. B. Kräutler, in *Metal-Carbon Bonds in Enzymes and Cofactors*, eds. A. Sigel, H. Sigel and R. K. O. Sigel, Royal Society of Chemistry, London, 2009, ch. 1, pp. 1-51.
71. M. Bommer, C. Kunze, J. Fesseler, T. Schubert, G. Diekert and H. Dobbek, *Science*, 2014, **346**, pp. 455-458.
72. S. Keller, C. Kunze, M. Bommer, C. Paetz, R. C. Menezes, A. Svatoš, H. Dobbek and T. Schubert, *J. Bacteriol.*, 2018, **200**. e00584-17.
73. D. P. Dowling, Z. D. Miles, C. Köhrer, S. J. Maiocco, S. J. Elliott, V. Bandarian and C. L. Drennan, *Nucleic Acids Res.*, 2016, **44**, pp. 9965–9976.
74. J. Bridwell-Rabb and C. L. Drennan, *Curr. Opin. Chem. Biol.*, 2017, **37**, pp. 63-70.
75. K. R. Heal, W. Qinb, F. Ribaleta, A. D. Bertagnollib, W. Coyote-Maestasa, L. R. Hmeloa, J. W. Moffettc, A. H. Devola, E. V. Armbrusta, D. A. Stahlb and A. E. Ingallsa, *PNAS*, 2017, **114**, pp. 364–369.
76. T. Goris, T. Schubert, J. Gadkari, T. Wubet, M. Tarkka, F. Buscot, L. Adrian and G. Diekert, *Environ. Microbiol.*, 2014, **16**, pp. 3562-3580.
77. H. Nonaka, G. Keresztes, Y. Shinoda, Y. Ikenaga, M. Abe, K. Naito, K. Inatomi, K. Furukawa, M. Inui and H. Yukawa, *J. Bacteriol.*, 2006, **188**, pp. 2262-2274.
78. A. Rupakula, Y. Lu, T. Kruse, S. Boeren, C. Holliger, H. Smidt and J. Maillard, *Front. Microbiol.*, 2015, **5**. 751.
79. F. E. Löffler, J. Yan, K. M. Ritalahti, L. Adrian, E. A. Edwards, K. T. Konstantinidis, J. A. Müller, H. Fullerton, S. H. Zinder and A. M. Spormann, *Int. J. Syst. Evol. Microbiol.*, 2013, **63**, pp. 625-635.
80. S. J. Moore, M. J. Mayer, R. Biedendieck, E. Deery and M. J. Warren, *N. Biotechno.l*, 2014, **31**, pp. 553-561.
81. D. R. DiMasi, J. C. White, C. A. Schnaitman and C. Bradbeer, *J. Bacteriol.*, 1973, **115**, pp. 506-513.
82. D. P. Chimento, A. K. Mohanty, R. J. Kadner and M. C. Wiener, *Nature*, 2003, **10**, pp. 394-401.
83. D. C. Johnson, D. R. Dean, A. D. Smith and M. K. Johnson, *Annu. Rev. Biochem.*, 2005, **74**, pp. 247-281.
84. R. H. Holm, *ChemInform*, 2004, **35**. doi:10.1002/chin.200449262.
85. H. Beinert, R. H. Holm and E. Munck, *Science*, 1997, **277**, pp. 653-658.
86. L. M. Hunsicker-Wang, A. Heine, Y. Chen, E. P. Luna, T. Todaro, Y. M. Zhang, P. A. Williams, D. E. McRee, J. Hirst, C. D. Stout and J. A. Fee, *Biochemistry*, 2003, **42**, pp. 7303-7317.
87. J. A. Inlay, *Mol. Microbiol.*, 2006, **59**, pp. 1073-1082.
88. K. Zaremba-Niedzwiedzka, E. F. Caceres, J. H. Saw, D. Bäckström, L. Juzokaite, E. Vancaester, K. W. Seitz, K. Anantharaman, P. Starnawski, K. U. Kjeldsen, M. B. Stott,

T. Nunoura, J. F. Banfield, A. Schramm, B. J. Baker, A. Spang and T. J. G. Ettema, *Nature*, 2017, **541**, p. 353.
89. R. E. Richardson, *Curr. Opin. Microbiol.*, 2013, **24**, pp. 498-505.
90. L.-L. Wen, J.-X. Chen, J.-Y. Fang, A. Li and H.-P. Zhao, *Front. Microbiol.*, 2017, **8**. 1439.
91. T. Futagami, M. Goto and K. Furukawa, *Chem. Rec.*, 2008, **8**, pp. 1-12.
92. F. A. Collins, PhD thesis, The University of Manchester, 2017.
93. B. A. VandePas, H. Smidt, W. R. Hageni, J. v. d. Oost, G. Schraa, A. J. M. Stams and W. M. d. Vos, *J. Biol. Chem.*, 1999, **274**, pp. 20287–20292.
94. S. A. F. T. v. Hijum, M. H.Medema and O. P. Cuipers, *Microbiol. Mol. Biol. R.*, 2009, **73**, pp. 481-509.
95. H. Smidt, M. v. Leest, J. v. d. Oost and W. M. d. Vos, *J. Bacteriol.*, 2000, **182**, pp. 5683-5691.
96. G. D. Sankhe, N. M. Dixit and D. K. Saini, *mSphere*, 2018, **3**. e00111-18.
97. C. Levy, K. Pike, D. J. Heyes, M. G. Joyce, K. Gabor, H. Smidt, J. v. d. Oost and D. Leys, *Mol. Microbiol.*, 2008, **70**, pp. 151-167.
98. M. Kube, A. Beck, S. H. Zinder, H. Kuhl, R. Reinhardt and L. Adrian, *Nat. Biotechnol.*, 2005, **23**, pp. 1269–1273.
99. T. Kumarevel, in *Antibiotic Resistant Bacteria - A Continuous Challenge in the New Millennium*, ed. M. Pana, Intech, London, 2012, ch. 16, pp. 403-418.
100. S. P. Wilkinson and A. Grove, *Curr. Issues Mol. Biol.*, 2006, **8**, pp. 51-62.
101. M. N. Alekshun, S. B. Levy, T. R. Mealy, B. A. Seaton and J. F. Head, *Nature*, 2001, **8**, pp. 710-714.
102. D. K. Deochand and A. Grove, *Crit. Rev. Biochem. Mol. Biol.*, 2017, **52**, pp. 595-613.
103. A. Grove, *Comput. Struct. Biotechnol. J.*, 2017, **15**, pp. 366-371.
104. R.-y. Wu, R.-g. Zhang, O. Zagnitko, I. Dementieva, N. Maltzev, J. D. Watson, R. Laskowski, P. Gornicki and A. Joachimiak, *J. Biol. Chem.*, 2003, **278**, pp. 20240-20244.
105. L. Krasper, H. Lilie, A. Kublik, L. Adrian, R. Golbik and U. Lechner, *J. Bacteriol.*, **198**, pp. 3130-3141.
106. A. Wagner, L. Segler, S. Kleinsteuber, G. Sawers, H. Smidt and U. Lechner, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120317.
107. A. Kublik, D. Deobald, S. Hartwig, C. L. Schiffmann, A. Andrades, M. v. Bergen, R. G. Sawers and L. Adrian, *Environ. Microbiol.*, 2016, **18**, pp. 3044–3056.
108. H. Scholz-Muramatsu, *FEMS Microbiol. Lett.*, 1990, **66**, pp. 81-85.
109. A. Neumann, G. Wohlfarth and G. Diekert, *J. Biol. Chem.*, 1996, **271**, pp. 16515-16519.
110. W. Schumacher, C. Holliger, A. J. B. Zehnderb and W. R. Hagen, *FEBS Lett.*, 1997, **409**, pp. 421-425.
111. Y. Morita, T. Futagami, M. Goto and K. Furukawa, *Appl. Microbiol. Biotechnol.*, 2009, **83**, pp. 775–781.
112. T. Palmer, B. C. Berks and F. Sargent, *Methods Mol. Biol.*, 2010, **619**, pp. 191-216.
113. T. Kruse, B. A. VanDePas, A. Atteia, K. Krab, W. R. Hagen, L. Goodwin, P. Chain, S. Boeren, F. Maphosa, G. Schraa, W. M. d. Vos, J. v. d. Oost, H. Smidt and A. J. M. Stams, *J. Bacteriol.*, 2015, **1975**, pp. 893-904.
114. E. Akiva, J. N. Copp, N. Tokuriki and P. C. Babbitt, *PNAS*, 2017, **114**, pp. E9549-E9558.
115. J. Krasotkina, T. Walters, K. A. Maruya and S. W. Ragsdale, *J. Biol. Chem.*, 2001, **276**, pp. 40991–40997.
116. M. Cooper, A. Wagner, D. Wondrousch, F. Sonntag, A. Sonnabend, M. Brehm, G. Schüürmann and L. Adrian, *Environ. Sci. Technol.*, 2015, **49**, pp. 6018-6028.
117. J. A. Schofield, W. W. Brennessel, E. Urnezius, D. Rokhsana, M. D. Boshart, D. H. Juers, P. L. Holland and T. E. Machonkin, *Eur. J. Inorg. Chem.*, 2015, **2105**, pp. 4643-4647.
118. T. E. Machonkin, M. D. Boshart, J. A. Schofield, M. M. Rodriguez, K. Grubel, D. Rokhsana, W. W. Brennessel and P. L. Holland, *Inorg. Chem.*, 2014, **53**, pp. 9837-9848.
119. P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.*, 2013, **15**, pp. 11178-11189.

120. R.-Z. Liao, S.-L. Chen and P. E. M. Siegbahn, *ACS Catal.*, 2015, **5**, pp. 7350-7358.
121. M. Kawai, T. Futagami, A. Toyoda, Y. Takaki, S. Nishi, S. Hori, W. Arai, T. Tsubouchi, Y. Morono, I. Uchiyama, T. Ito, A. Fujiyama, F. Inagaki and H. Takami, *Front. Microbiol.*, 2014, **5**, pp. 1-15.
122. H. Stevens, M. Stubner, M. Simon and T. Brinkhoff, *FEMS Microbiol. Ecol.*, 2005, **54**, pp. 351-365.
123. M. L. Timmermans, Y. P. Paudel and A. C. Ross, *Mar. Drugs*, 2017, **15**. 235.
124. A. M. Cupples, R. A. Sanford and G. K. Sims, *Appl Environ Microbiol.*, 2005, **71**, pp. 3741–3746.
125. F. A. Collins, K. Fisher, K. A. P. Payne, S. Gaytan, S. E. J. Rigby and D. Leys, *Biochemistry*, 2018, **57**, p. 3493–3502.
126. C. C. Corell, C. J. Batie, D. P. Ballou and M. L. Ludwig, *Science*, 1992, **258**, pp. 1604-1610.

2.0 Chapter Two. Materials and Methods

The Second Chapter describes all the materials and methods used in this work. The methodologies have been organised in different categories, between experimental and *in silico*. In the first category we have included: molecular biology, protein expression, and purification, electrophoretic mobility shift assays (EMSAs), enzymatic assays, biotransformation, and biophysical techniques used for protein characterisation or binding assays (fluorescence-quenching experiments, isothermal titration calorimetry (ITC), analytical ultracentrifugation (AUC), circular dichroism (CD), electron paramagnetic resonance, RAMAN spectroscopy, and finally X-ray crystallography. In the second category, I describe the bioinformatics methods, homology modelling, molecular dynamics (MD), and rigid-body docking.

2.1 Materials

Chemical reagents were purchased from Sigma-Aldrich, Fisher Scientific, Fluka, and Fluorochem, with analytic grade purity (>99%), although in some cases a standard laboratory-grade had to be used (>95%).

2.2 Bacterial strains and plasmids

2.2.1 *Escherichia coli* strains

DH5α and all DE3 or HMS174 strains of competent cells were supplied by Novagen (Merck, Germany). Stellar Clonetech competent cells were supplied by Takara (Clonetech, USA). Genes were synthesised by GeneArt (Thermo Fisher Scientific, UK) and were codon optimised for *E. coli* expression.

Strain	Genotype
***Escherichia coli* DH5α**	*[F- ϕ80lacZΔM15 Δ(ΔlacZYA-argF) U169 recA1 endA1 hsdR17(rk-, mk+) phoA supE44 thi-1 gyrA96 relA1 λ-]*
***E. coli* SoluBL21™**	*F- ompT hsdSB (rB - mB -) gal dcm (DE3)*
***E. coli* Arctic Express (DE3)**	*[F– ompT hsdS(rB-mB-) dcm+ Tetr gal λ(DE3) endA Hte [cpn10 cpn60 Gentr]*

***E. coli* BL21(DE3)**	*[F– ompT hsdSB(rB-mB-) gal dcm rne131 (DE3)]*
***E. coli* HMS174 (DE3)**	*[F– ompT hsdSB(rB-mB-) recA1, rpoB331*

Table 1. *E. coli* strains utilised in this work.

2.2.2 *Bacillus megaterium* strains

Protoplasts from *Bacillus megaterium* strain DSM319 were obtained from Dr. Karl Payne, as a generous donation from Professor Martin Warren's laboratory (University of Kent, United Kingdom). Genes were synthesised by GeneArt (Thermo Fisher Scientific, UK) and were codon optimised for *B. megaterium* expression. *Bacillus megaterium* (a Gram-positive soil bacteria) does not produce alkaline proteases, a fact that allows the heterologous expression of proteins without the risks of proteolytic degradation. The *B. megaterium* strain MS941 was generated from the wild-type strain DSM319 by deletion of an extracellular protease (nprM), allowing high levels of extracellular protein production. All cloning vectors of the *B. megaterium* are derived from the pWH1520, a xylose inducible expression system. Two parallel-replicating plasmids are extensively used for transformation: pT7-RNAP and pPT7. pT7-RNAP contains the genes for ampicillin and chloramphenicol resistance for easy selection when cloning in *E. coli* (AmpR) and pPT7 is responsible for the T7 RNA polymerase dependent expression of the gene of interest. Downstream of the T7 promoter, the pPT7 contains the multiple cloning site and two selection markers for ampicillin (in *E. coli*) and tetracycline (in *B. megaterium*) **[1]**.

2.2.3 $RdhR_{cbdb1625}$ construction

The $RdhR_{cbdb1625}$ construct was a gift from Professor Anke Wagner.

2.2.4 Plasmids

Most of the plasmids used in this study belong to the pEt System series of vectors, originally derived from the pBR322 plasmid, one of the most widely used for cloning and expression of recombinant proteins in *E. coli.* The genes cloned into the pEt systems are under the control of the T7 bacteriophage gene *10* to promote high levels of transcription and translation. The viral T7 polymerase is highly specific for its promoter sequences guaranteeing that only the target gene will be transcribed, and in turn, when the inductor is not present in the media, the target gene is transcriptionally silent (since the promoter is not recognised by the host polymerases) **[2]**.

2.2.5 Agar Plates

Plates for bacterial growth were prepared using LB-agar media (Formedium, UK). After sterilisation, the appropriate antibiotics were added to the molten agar before solidification and the mixture was poured into sterile plastic Petri dishes (100 x 20 mm). The antibiotic stock concentrations might vary when preparing plates. *B. megaterium* MS941 is asporogenic and cannot be kept in plates at 4 °C for more than a week, so glycerol stocks must be prepared immediately after transformation and working colonies have to be streaked continually on fresh plates.

2.3 Molecular Biology methods

2.3.1 PCR and ligation independent cloning

For cloning of the reductive dehalogenase RdhA genes, the lyophilised DNA was resuspended in 50 µL of nuclease-free water dH_2O and stored at -20 °C, when not used immediately. Coding sequences were amplified by PCR using Q5 HF polymerase (New England Biolabs, UK) and specific oligonucleotides designed to have a T_m of 58 °C to allow annealing using the SnapGene© to incorporate the genes into different T7 promoter-based vectors with ligation independent cloning or by normal ligation using T4 DNA-ligase (New England Biolabs, UK). PCR reactions were prepared using nuclease-free water, containing approximately 50 ng of the template DNA. Reactions were cycled using a BioRad C1000 TM thermocycler, with the following conditions:

Thermocycling conditions:

Backbone (5782bp)			Insert (3245 bp):		
98 C	30 s	1x	98 C	30 s	1x
98 C	10 s	31x	98 C	10 s	31x
50 C	15s		50 C	15 s	
72 C	3 min		72 C	1min 45s	
72 C	7 min	1x	72 C	5 min	1x
4 C	End		4 C	End	

Table 2. PCR thermocycling conditions.

Destination vectors (pET32b(+), pET28a(+), pNIC28 and pET30a(+)) were prepared by PCR linearisation, according to the protocol outlined before or by restriction digestion reactions set according to the enzymes manufacturer's protocol (New England Biolabs, UK): restriction enzyme digest reactions contained 1.5 µg of plasmid DNA plus 10 U of the restriction enzymes (New England Biolabs, UK), made up to a final volume of 20 µL with dH_2O in Cut-Smart® Buffer and incubated at 37 °C for 1 h. Restriction enzyme digests using endonucleases pairs were performed in 2 h, with the addition of the enzymes one-by-one. When digesting PCR inserts with *DpnI*, the enzyme was added directly to the PCR reaction mixture, without the addition of Cut-Smart®.

Successful amplification and digestion were confirmed by agarose gel electrophoresis. Agarose gels for the electrophoretic separation of nucleic acids were prepared at 0.4 w/v of ultrapure agarose in 50 mL of TAE Buffer 1x [Tris-HCl 40 mM, CH_3CO_2H 20 mM, EDTA 1 mM, pH 8.0], adding Safeview 1.0 μgmL^{-1} (NBS Biologicals, UK). DNA samples were prepared adding 2 µL of DNA loading buffer 10 x [bromophenol blue 0.25% w/v, xylene cyanole 0.25% w/v and glycerol 30%] per 10 µL. The molecular weight marker was the 1 kb DNA ladder (New England Biolabs, UK).

The In-fusion cloning reaction mixture is composed as follows: 2 µL In-Fusion Enzyme (Clonetech, USA), 10-200 ng of the digested PCR insert (previously treated with DpnI (New England Biolabs, UK)) and 50-200 ng of the digested vector, made up to 10 µL with nuclease-free dH2O. The reactions were incubated at 50 °C for 15 mins. 2.5 µL of the In-Fusion products were used to transform DH5α maintenance cells (Stellar or NEB5α), employing the usual transformation protocols.

2.3.2 DNA electrophoresis

For electrophoretic separation of nucleic acids, agarose gels were prepared at 1%. Electrophoresis was performed at 100 mV for 30 min. DNA visualisation was done using a UV transilluminator (Cole-Parmer, USA). The remaining reaction mixture was purified using the QIAquick PCR Purification Kit (Qiagen, UK) according to the protocol.

2.3.3 DNA concentration determination

The concentration of DNA was measured using a Nanodrop (Thermo Fisher Scientific, UK) following the manufacturer's instructions.

2.3.4 DNA sequencing

Before protein expression trials and scale-ups were carried out, the correct orientation and reading frame of all gene constructs were verified by DNA sequencing (MWG Eurofins, Germany).

2.4 *E. coli* and *B. megaterium* methods

2.4.1 *E. coli* and *B. megaterium* growth conditions

All strains of *E. coli* were cultured in previously sterilised LB broth liquid media (Formedium, UK) or M9 media, following the instructions of the manufacturer. Cultures were incubated at 37 °C with shaking (200 rpm). *Bacillus megaterium* was cultured in TB broth liquid media (Formedium, UK) and incubated at 30 °C with shaking (200 rpm). The media contains varying concentrations of bactotryptone, yeast extract, NaCl, and micronutrients, such as inorganic salts and amino acids. The final concentration for standard antibiotics used for the selection and maintenance of both organism strains are summarised below **(Table 3)**.

All stocks were sterilised using a 0.25 µm filter and stored at -30 °C. Tetracycline stocks were covered in foil due to its light sensitivity.

Antibiotic	Stock concentration and solvent ($mgmL^{-1}$)	
Ampicillin	50	ddH_2O
Kanamycin	50	ddH_2O
Chloramphenicol	34/4.5	EtOH
Tetracycline	5	EtOH
Rifampicine	50	DMSO/EtOH

Table 3. Antibiotic concentrations for selection and maintenance of organisms.

2.4.2 Transformation into *Escherichia coli* strains

E. coli DH5α maintenance cells (Stellar or NEB5α) and multiple expression strains like BL21 (DE3), soluBL21, Artic Express, and HMS174 (NEB, Amsbio, Agilent, and Novagen) were transformed by T7 promoter-based plasmids according to the manufacturer protocols and plated onto LB agar containing the appropriate antibiotics. The general methodology is described as follows: approximately 50 µL of competent cells were thawed on ice for 10 min, then 2.5 µL containing 50-100 $ngmL^{-1}$ of the DNA constructs were added and mixed

carefully under sterile conditions. The cells were incubated on ice for 30 min. The cells were then heat-shocked for 10 to 45 s (depending on the strain) at 42 °C and immediately transferred to ice for an additional incubation time of 2 to 5 min. About 450 µL of preheated SOC medium (Novagen, UK) were added and the cells were incubated at 37 °C for 1 hour in a shaker. 100 µL cells were spread onto fresh LB-agar plates (Formedium, UK) containing the appropriate antibiotic to a final concentration of 50 µg mL^{-1}. Plates inoculated with *E. coli* strains were incubated overnight at 37 °C and then stored at 4 °C. When the cells were transformed with multiple plasmids, the incubation in SOC media was extended by 30 minutes or done overnight, before being centrifuged at 13 200 g for 5 minutes. The supernatant was discarded and the cell pellet was gently resuspended in 50 µL fresh SOC media to be plated on LB agar. The protocol was varied due to a lower transformation efficiency when transforming cells with multiple plasmids.

2.4.3 Transformation into *Bacillus megaterium* strains

2.4.3.1 Media preparation

The following solutions/media were prepared for the transformation protocol:

Prot medium

Component	Quantity	Preparation
NH_4Cl	10 g	Dissolve all components in approximately 500 mL of distilled water, adjust pH to 7.5 with concentrated HCl, and add distilled water to a final volume of 864 mL.
Tris/HCl	120 g	
KCl	350 mg	
NaCl	580 mg	
$MgSO_4$ x 7 H_2O	2.67 g	
$MgCl_2$ x 6 H_2O	46.7 g	

Table 4. Components for preparation of Prot-medium.

Hyp-medium

Component	Quantity	Preparation
Prot-medium	432 mL	Autoclave all components separately before mixing on a 1L bottle
Sucrose [50 % (v/w)]	68 mL	
Glucose [20 % (v/w)]	5 mL	
Yeast extract 10%	3 mL	
KH_2PO_4 x H_2O [0.5 M]	1 mL	
$MnSO_4$ [2 mg/mL]	50 µL	

Table 5. Components for preparation of Hyp-medium.

Hyp-agar plates

Component	Quantity	Preparation
Hyp medium	216 mL	Add the agar into the previously prepared Hyp medium.
Agar	3.75 g	
Hyp agar – top		
Hyp medium	216 mL	Add the agar into the previously prepared Hyp medium.
Agar	2 g	

Table 6. Preparation of Hyp agar for base plates and top plates.

40% PEG-6000-Lösung (in Prot-medium)

Component	Quantity	Preparation
NH4Cl	0.125 g	Components for 125 mL 40% PEG-6000-Prot-medium. All components were dissolved in a 500mL bottle with 20 mL of distilled water, and pH was adjusted to 7.5 using stock HCl solution. 50 g PEG-6000 were added and volume adjusted to 108mL with distilled water. After shaking until solubilsed in an incubator at 37 °C the solution was autoclaved. Once autoclaved, 17 mL of 50 % sucrose solution (previously autoclaved) were added and the solution was stored in refrigeration under sterile conditions for future use
Tris-HCl	1.5 g	
KCl-KCl	4.375 g	
NaCl	7.25 g	
$MgSO_4$ x 7 H_2O	33.37 mg	
$MgCl_2$ x 6 H_2O	0.58 g	

Table 7. Components for preparation of 40% PEG-6000 in Prot-medium.

2.4.3.2 Protoplast preparation

In preparation for the transformation protocol, falcon tubes were chilled at 4 °C, pre-cultures were checked for contamination, agar was pre-warmed in a water bath at 45 °C, and a sterile lysozyme solution 5 $mgmL^{-1}$ was freshly prepared in Hyp-medium and kept on ice. A 250 mL flask with 50 mL of Hyp-medium was inoculated using 1.5 mL of the pre-culture and incubated at 37 °C and 200 rpm until OD_{578nm}=0.9. Cells were then incubated for 30 min on ice in the pre-chilled Falcons. In parallel, two additional 50 mL Falcon tubes with Hyp-medium were prepared and stored on ice. After 10 minutes, cells were harvested by centrifugation (7 min, 5000 rpm at 4 °C). Pellets were then resuspended in the pre-chilled Hyp-medium for a second centrifugation.

Protoplasts were generated by adding 250 µL of the freshly prepared lysozyme solution to the resuspended pellets and incubating at 30 °C and slow shaking (<100 rpm). Progress was monitored by observation under the microscope. Once approximately 60% of the cells were protoplasts, these were placed on ice for 5 min, and 3 cycles of centrifugation (4min, 2400 g at 4 °C) and washing with 5 mL of Hyp-medium were performed to harvest them. The resulting pellet was resuspended on 2 mL of Hyp-medium and stored on ice (maximum of 2 h).

2.4.3.3 Protoplast transformation

For each plasmid to transform, 3x 1.5 Eppendorf tubes with 100 µL of Hyp-medium and approximately 5 µg of DNA were prepared and labelled as viability control (V), negative control (Neg), and the corresponding plasmid (P). 3x 15 mL Falcon tubes were equally labelled and 1.5mL of 40% PEG-6000 solution added. About 500 µL of the protoplast suspension were pipetted into each P, V, and Neg Eppendorfs and carefully mixed. The solution was then added into the corresponding Falcon tube (V, Neg, P) and mixed by rolling the tube on the table. Transformed protoplasts were left to rest for 4 minutes before proceeding. After the incubation time, 5 mL of pre-chilled Hyp-medium were added into each Falcon tube, contents were carefully mixed, and the transformed protoplasts were harvested by centrifugation (4 min, 1400 g, 4 °C), discarding the supernatant. The resulting pellets were re-suspended in 1 mL of Hyp-medium and incubated for approximately 1.5 h at 30 °C, and agitation of 100 rpm.

2.4.3.4 Preparation of plates agar and top agar

Two thin base plates were made without tetracycline before mixing the remaining media (approximately 200 mL) with 400 µL of tetracycline 5 $mgmL^{-1}$; 10 additional plates were made using the Hyp-agar + tetracycline.

Hyp-agar for base plates

Component	Quantity	Preparation
Hyp-agar	216 mL	All components were mixed.
Sucrose [50 % (v/w)]	34 mL	
Glucose [20 % v/w]	2.5 mL	
Yeast extract 10%	1.5 mL	
KH_2PO_4 x H_2O [0.5 M]	512 µL	
$MnSO_4$ [2 mg/mL]	25 µL	

Table 8. Components for preparation of Hyp-agar for base plates.

Hyp-agar for top plates

Component	Quantity	Preparation
Hyp-top-agar	216 mL	All components were mixed.
Sucrose [50 % (v/w)]	34 mL	
Glucose [20 % v/w]	2.5 mL	
Yeast extract 10%	1.5 mL	
KH_2PO_4 x H_2O [0.5 M]	512 µL	
$MnSO_4$ [2 mg/mL]	25 µL	

Table 9. Components for preparation of Hyp-agar for top plates.

Top agar was prepared according to the above formulation, and 7 mL aliquots were added to 15 mL Falcon tubes pre-warmed at 45 °C. Tubes were kept in a water bath at 45 °C. About 900 µL and 100 µL fractions of the resuspended transformed protoplasts were added to the Falcon tubes with the Hyp-top-agar according to the table below, and contents were mixed by gently rolling the tube in the table before pouring the contents over the previously prepared Hyp-agar plates containing tetracycline, except the viability controls which will be made in the plates without antibiotic.

Aliquots

Plasmid	Viability control	Negative control
1 x 100 µL	1 x 100 µL	2 x 100 µL
1 x 900 µL	1 x 900 µL	2 x 900 µL
Hyp-tet agar	Hyp-agar	Hyp-tet agar

Table 10. Aliquots of transformed protoplasts and indication of media to use.

Plates were left for 30 min at ambient conditions before incubating for 16h at 30 °C. Plates were not inverted to prevent top-agar from collapsing.

2.4.3.5 Evaluation of transformation

Plates were monitored after the 16 h period for presence/absence of colonies. If no colonies were present, temperature was raised to 37 °C and plates were incubated for a maximum of 10 additional hours.

*

2.4.4 Colony PCR

Successfully transformed DH5α or the *B. megaterium* transformed colonies were checked for the incorporation of the insert by colony PCR amplification. The reaction mix containing 50% v/v CloneAmp HiFi 2X mastermix (Takara, USA) and 1 μM of each T7 forward and reverse standard primers, plus a sample of an individual colony by washing the tip into 10 μL of PCR reaction mix. Positive colonies were streaked on a fresh LB agar plate containing the appropriate antibiotic and cultured overnight. Positive colonies were cultured overnight before minipreps of plasmids were purified using the Qiagen standard kit and protocols. Purified plasmids were sequenced using Eurofins Genomics sequencing services.

2.4.5 Small-scale overnight growths

Multiple Erlenmeyer flasks of 250 mL of LB or TB (Formedium, UK) with 0.4% (v/v) glycerol added were used for small-scale growths. Cultures were inoculated by addition of 1 % (v/v) a single bacterial colony. All cultures were grown with μgL^{-1} concentrations of the appropriate antibiotic(s) overnight, typically for 12 hours at the standard temperature for each host (37 °C for *E. coli* or 30 °C for *B. megaterium*). These cultures were used for inoculation.

2.4.6 Plasmid purification and glycerol stocks

After inoculating 5 mL of LB broth media supplemented with the appropriate antibiotics (grown overnight) with single colonies after the transformation, plasmid purification was carried out using a Spin Miniprep Kit (QIAGEN, UK) following the manufacturer's instructions. DNA concentration was determined in a Thermo Scientific Nanodrop. Screening for colonies was made after purifying the plasmid DNA and analysing the DNA via PCR and digestion reactions. Glycerol stocks of the positive colonies were made by adding 0.250 mL of cells to sterile cryogenic vials with 0.250 mL of autoclaved glycerol 75% in liquid LB media. The vial was mixed by inversion and stored at - 80 °C.

2.5. Protein expression and purification

2.5.1 Protein expression in *E. coli*

E. coli DE3 strains (Novagen, Germany) containing the pEt constructs of interest were inoculated in 50 mL of LB media supplemented with the appropriate antibiotics and were grown in a 37 °C shaker overnight (200 rpm). The next morning flasks containing 0.5 or 1 L of sterile media supplemented with the same antibiotics were inoculated with 2-5 mL of the overnight cultures. Cells were grown at 37 °C (200 rpm) until an OD_{600nm} nm ≤ 0.6. Then, IPTG was added for induction of the expression and the temperature was lowered to less 20 °C. Cultures were grown overnight. Cells were harvested by centrifugation at 6 000 rpm for 10 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). Pellets were stored at -80 °C.

2.5.2 Protein expression in *B. megaterium*

Single colonies of *B. megaterium* DSM319 containing the RdhA-pPT7 constructs were inoculated in two flasks of 250 mL of LB supplemented with tetracycline 10 μgmL^{-1} and chloramphenicol 4.5 μgmL^{-1} and grown overnight in a shaking incubator at 30 °C and 180 rpm. The next morning, 24 flasks containing 1L of sterile LB supplemented with tetracycline 10 μgmL^{-1} and chloramphenicol 4.5 μgmL^{-1} were inoculated with 5 mL of the overnight culture. Cultures were grown at 30 °C until an OD_{600nm} ≤ 0.5 was reached. The cultures were induced with varying concentrations of xylose (0.25 or 0.50 % w/v).

After the addition of filter-sterilised solutions of hydroxocobalamin 2 µM and $(NH_4)_2FeSO_4 \cdot 6H_2O$ 50 µM, the temperature was lowered to 17 °C. Cultures were induced overnight. Cells were harvested by centrifugation at 6 000 rpm for 15 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). Pellets were stored at -80 °C.

2.5.3 Immobilised metal ion affinity chromatography

Immobilised metal ion affinity chromatography (IMAC) is a purification technique based on the interaction between a specific region of the recombinant protein and the stationary phase. It consists of 4 basic steps: cell lysis, binding, washing, and elution. In this work, IMAC purification was made with a Protino® Ni-TED or a Ni-NTA column (Qiagen, UK), both of which bind polyhistidine tags. The chelating group of Protino® Ni-TED is based on TED (Tris-carboxymethyl ethylene diamine), a pentadentate metal chelator. TED occupies five of the six binding sites in the coordination sphere of the Ni^{2+} ion, the remaining coordination site of Ni^{2+} is available for protein binding. Ni-NTA (nitrilotriacetic acid) has four binding sites available for the Ni^{2+} ion, the remaining sites of Ni^{2+} are available for protein binding. Purification was carried according to the manufacturer recommendations and is described in more detail in the corresponding Results Chapter's sections.

2.5.4 Streptactin XT

The Strep-tag®II peptide (IBA Life Sciences, Germany) is an 8 amino acid peptidic tagging sequence used to improve purification of recombinant proteins. It binds with high selectivity to Strep-Tactin® and Strep-Tactin®XT – both constituted by engineered streptavidin with improved binding properties. When using Strep-Tactin, specific elution has to be performed with 50 mM biotin or 2.5 mM d-desthiobiotin. Specific experimental details are further described in the relevant section (Results Chapter I).

2.5.5 Ion-exchange chromatography

Ion exchange chromatography (IEX) separates molecules on the basis of differences in their net surface charge. Molecules will exhibit different degrees of interaction with charged chromatography media according to differences in their overall charge, charge density, and surface charge distribution. IEX takes advantage of the fact that the relationship between the net surface charge and pH is unique for a specific protein, mainly because the charged groups within a molecule that contribute to the net surface charge possess different pK_a values depending on their structure and chemical microenvironment.

In IEX purification, reversible interactions between charged molecules and oppositely charged IEX media are controlled to favour binding or elution of specific molecules and achieve separation. A protein that has no net charge at a pH equivalent to its isoelectric point (pI) will not interact with a charged medium **[3]**. In this thesis, ion-exchange chromatography was made using a variety of Q resins for the ÄKTA Purifier (GE). Specific methods are described in Results Chapters II and III.

2.5.6 Hydroxyapatite

Hydroxyapatite ($Ca_5(PO_4)_3OH)_2$ is a form of calcium phosphate used in the chromatographic separation of biomolecules. Sets of five calcium doublets (C-sites) and pairs of –OH containing phosphate triplets (P-sites) are arranged in a geometric pattern. Hydroxyapatite contains two types of binding sites, positively charged calcium and negatively charged phosphate groups. These sites are distributed regularly throughout the crystal structure of the matrix. Cation exchange occurs when protein amino groups interact ionically with the negatively charged phosphates. The amino groups are similarly repelled by the calcium sites. Binding depends upon the combined effects of these interactions. The addition of salt or phosphate, or an increase in pH, can be used to weaken the interaction

and elute the solute. Calcium affinity occurs via interactions with carboxyl clusters and/or phosphoryl groups on proteins or other molecules (e.g., nucleic acids); these groups are simultaneously repelled by the negative charge of the hydroxyapatite phosphate groups. The affinity interaction is between 15 and 60 times stronger than ionic interactions alone and, like classical metal-affinity interactions, is not affected by increasing ionic strength using typical elution ions (e.g., chloride anions). Metal affinity interactions can be dissociated by phosphate in the mobile phase. In this work, CHT ceramic hydroxyapatite (Bio-Rad Laboratories, UK) columns were used. Specific conditions of use and experimental parameters are described in Results Chapter II.

2.5.7 Size-exclusion chromatography

SEC separates molecules according to their differences in size as they pass through a chromatography medium packed in a column. SEC resins consist of a porous matrix of chemically and physically stable spherical particles with properties that minimise the adsorption of biomolecules. Sample components are eluted isocratically, meaning that the elution buffer does not need to be changed during the process. In this work, SEC resin Superdex-200 (GE Healthcare, UK) was used. Specific conditions of use and experimental parameters are described in the relevant section (Results Chapters II and III).

2.5.8 SDS-PAGE electrophoresis

Samples were made up of 20 µL with 10 µL SDS 2 x Sample Buffer and mixed before boiling at 100 °C for 5 mins. 10 µL of samples were loaded onto a Precast gel 10 % precast SDS-PAGE gel cassette (Expedeon, UK) in addition to 5 µL Page ruler pre-stained protein ladder (Fermentas, UK). SDS 1X running buffer and a voltage of 150 V was applied to the gel for a good separation of the protein samples. The gel was stained with Instant Blue (Gentaur, UK).

2.5.9 Determination of protein concentration by UV-Visible spectroscopy

Protein concentration was determined by UV-Visible spectrophotometry at $\lambda^{280\ nm}$ in a Cary 200 UV-Vis Spectrophotometer (Varian, USA), using the value for the molar extinction coefficient generated from the primary sequences by the server ProtParam, included in ExPASy [4] . Samples were baseline corrected using the working buffer solution

and by measuring transmittance 0 %. Scans were recorded at 0.5 ns between 200 – 800 nm. Alternatively, the concentration can be determined using the BioRad.

2.6 Protein characterisation

2.6.1 Protein fluorescence spectroscopy

Luminescence is defined as the light emission caused by the relaxation of excited electronic states in a substance; it is formally defined as either fluorescence or phosphorescence depending on the nature of the excited state. In excited singlet states, the electron in the excited orbital is paired to the second electron in the ground-state orbital. Consequently, return to the ground state is allowed and occurs rapidly by the emission of a photon. The emission rates of fluorescence are typically 108 s^{-1} so that a typical fluorescence lifetime is near 10 ns.

Fluorescence data are generally presented as emission spectra. A fluorescence emission spectrum is a plot of the fluorescence intensity versus wavelength (nm) or wavenumber (cm^{-1}). Emission spectra vary widely and are dependent upon the chemical structure of the fluorophore and its interactions with the solvent in which it is dissolved. Fluorescence typically occurs from aromatic molecules; in proteins, fluorescence is an intrinsic property of aromatic residues, which is particularly useful for monitoring the electronic microenvironments in a peptide, even providing information from conformational changes.

Intrinsic fluorescence experiments require the aromatic amino acid side chains to be in the regions of interest, but not too close together to avoid charge transfer phenomena, also known as quenching, and diminish the fluorescence intensity. In proteins, only aromatic amino acids will fluoresce. Phenylalanine (F), tyrosine (Y), and tryptophan (W) fluoresce when excited by UV radiation. In presence of tyrosine and tryptophan, phenylalanine fluorescence cannot be detected as it is too weak due to the low quantic yield in solution (0.04 for F in comparison to 0.2 and 0.21 for W and Y, respectively). Tyrosine residues also present a partial or total quenching when ionised or when present near the amine or carboxyl groups or tryptophan residues. Tryptophan fluorescence is usually monitored to determine protein intrinsic fluorescence variations, considering it will also diminish its intensity by quenching or by the proximity of protonated acid groups. **[5]**.

2.6.2 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) allows the biophysical characterisation of binding between unlabelled (no fluorescent tags), and non-immobilised biomolecules in solution. ITC is the only technique that can simultaneously determine all binding parameters in a single experiment, providing not only the binding constraints at equilibrium but also the stoichiometry and the thermodynamic parameters.

Measuring heat transfer during binding enables accurate determination of binding constants (KD), reaction stoichiometry (n), enthalpy (ΔH), and entropy (ΔS). This provides a complete thermodynamic profile of the molecular interaction. ITC goes beyond binding affinities and can elucidate the mechanisms underlying molecular interactions.

ITC monitors the difference in enthalpy or heat (released and/or absorbed in exothermic or endothermic processes) caused by the interactions or binding between molecules in multiple buffer solutions, independently of the molecular weight of the species of interest. ITC monitors ΔH by determining the differential heat provided by heaters placed in the instrument on the reference and sample cells, needed for counteracting any temperature difference between the two cells during the binding events. When binding occurs, heat is either absorbed or released and measured by the sensitive calorimeter during gradual titration of the ligand into the sample cell containing the biomolecule of interest **[6, 7]**.

2.6.3 Analytical ultracentrifugation

Analytical Ultracentrifugation (AUC) is a powerful technique that allows the characterisation of solutions of macromolecules and has been a very relevant tool for the quantitative analysis of macromolecular interactions since the 1920s when it was first developed. AUC has broad applicability and can be used to analyse the solution behaviour of a variety of molecules in a wide range of solvents and over a wide range of solute concentrations, and has the key advantage that samples are characterised in their native state under biologically-relevant solution conditions **[8]**.

Its basic principle is the property of mass and the fundamental laws of gravitation **[8]**, where the application of a centrifugal force is coupled to the real-time observation of the redistribution of the macromolecule in solution. The acquired data report on the spatial gradients that result from the application of a centrifugal field, and their evolution with time **[9]**. The analytical ultracentrifuge basic arrangement contains a built-in optical system (absorbance, interference, or fluorescence); the most common being a combination of

UV/Visible absorbance and a LASER interferometer to record refractive index gradients, allowing the observation of the movement of a sample as it is spun down in a centrifuge rotor **[8-10]**. Although a large number of specialised centrifugation techniques have been developed for a variety of studies, the vast majority of ultracentrifugation experiments for the characterisation of proteins are conducted by sedimentation velocity (SV) or sedimentation equilibrium (SE) **[9]**.

In SV high rotor speeds (<40,000 rpm) are used to cause the vast majority of proteins to sediment to the bottom of the cell. As the sedimentation depends both on the size and shape of the protein, valuable information can be extracted from this experiment: concentration-related self-association, solution heterogeneity amongst others. The major parameter derived from SV experiments is the sedimentation coefficient (s), through the application of the Lamm and Svedberg equations **[8]**:

$$\left(\frac{\delta c}{\delta t}\right)_r = D\left[\frac{\delta_c^2}{\delta r^2} + \frac{1}{r}\frac{\delta c}{\delta r}\right] - s\omega^2\left[r\frac{\delta c}{\delta r} + 2c\right] \quad \textbf{(Equation 1)}$$

$$\frac{s}{D} = \frac{M_b}{RT} \quad \textbf{(Equation 2)}$$

Equations 1 and 2. Lamm and Svedberg equations upon which the theory behind analytical ultracentrifugation is based.

In Lamm's equation **(Equation 1)**, c is the weight concentration of the molecule, t is time, D is the translation diffusion coefficient, r is the distance from the centre of the rotor, s is the sedimentation coefficient and ω is the rotor speed in radians per second. In Svedberg's equation **(Equation 2)**, M_b is the buoyant molar mass, R is the gas constant and T is the absolute temperature.

The second experiment that can be run by analytical Ultracentrifugation is sedimentation equilibrium (SE). At lower rotor speeds, the transport of samples down the centrifuge cell is balanced by its diffusion back up the cell due to the creation of a concentration gradient. SE is established when no change in the concentration distribution of any component is detectable. SE is a very powerful way to assess self-associating systems, where different oligomeric states are populated at different concentration ranges.

Several software packages are available for the analysis of SE and SV, with varying degree of flexibility and capabilities for the analysis of self-association and hetero-association, such as BPCFIT, SEDANAL, SEDFIT, SEDPHAT, amongst others.

2.6.4 Circular Dichroism

Circular dichroism (CD) is a technique that measures the asymmetry of a chromophore or its environment, by using polarised light. When polarised light interacts with an asymmetric medium it is reoriented, a phenomenon called optical rotation, and there is also a certain loss of polarisation, called ellipticity, which is measured as the circular dichroism. When peptide groups are organised in an asymmetric environment, such as in the protein native state forming α-helices, their near-UV CD will be much stronger than in a or β-sheets or than in a denaturated state, giving rise to distinctive CD signals sensitive to changes to the secondary structure **[11]**.

The CD spectrum of a protein is the sum of contributions from its secondary structures. By measuring CD spectra of a set of proteins with known secondary structure contents it is possible to solve for the spectra of α-helices and β-sheets, and the CD spectrum of a protein with unknown secondary structure can then be fitted to a sum of contributions from helix, sheet and a residual signal from other parts of the structure **[11]**.

2.6.5 Raman scattering

When a beam of monochromatic radiation is passed through a transparent substance, a small amount of its energy will be scattered. This scattered energy will consist almost entirely of radiation of the incident frequency (Rayleigh scattering) but a fraction of it will be scattered in a discrete range above and below the original beam, this is referred to as Raman scattering.

Raman scattering occurs when energy is exchanged in an inelastic collision between a photon with energy $h\nu$ and a molecule. The energy variation ΔE in the molecule must be the difference in energy between two of its allowed states following the quantum laws, so ΔE must represent a change in the vibrational and/or rotational energy of the molecule **[5]**. If these vibrational and/or rotational changes modify the polarisability of the molecule, it will be detectable by Raman spectroscopy through a characteristic shift in frequency. The multiple chemical bonds in a molecule will therefore generate a Raman fingerprint that can be used to detect and quantify it.

The similarity in the composition of most biologically relevant molecules makes discriminating between them using Raman spectroscopy challenging. Vitamin B_{12}, however, is unique in that it contains a cobalt ion bound to an organic corrin ring – a structure not found across other biological molecules. This makes Raman spectroscopy a particularly

attractive technique for measuring vitamin B_{12} because its unique Raman signature can provide direct molecule identification [12]. Cobalamin presents a characteristic Raman peak at 1504 cm^{-1}, corresponding to the ground state vibration of the corrin ring when excited at 532 nm [13].

2.6.6 X-ray crystallography

X-ray crystallography relies on the interaction between solid-state (crystalline) matter and the electromagnetic radiation in the range of 0.01–10 nm, where the electronic clouds will cause a beam of X-rays to diffract into many specific directions. By measuring the angles and intensities of these diffracted beams, a diffraction pattern is obtained, and an electron density map is produced from which, once analysed and refined, the mean positions of the atoms in the crystal can be determined, as well as their connectivity, and other information including the three-dimensional structure of biological models [14].

X-ray crystallography vs other techniques. Few techniques are available to determine structures at atomic resolution: structures can be visualised only if radiation with a wavelength comparable to the dimensions of the feature observed is used, and at the Angstrom scale of atomic radii and bonds, visualisation is only possible with electrons, neutrons or X-rays. X-rays and neutrons are scattered by proteins in solution, but they cannot be focused to create a reconstructed image of a single molecule. Therefore, at present only three techniques allow visualisation of three-dimensional structures at high resolution: X-ray or neutron diffraction analysis of single crystals, NMR of small proteins in solution, and low-temperature electron microscopy (Cryo-EM) of large protein aggregates [11, 15, 16].

Crystallisation of proteins. Adequate diffraction patterns can only be achieved from a suitable crystal. Protein crystallisation, however, is mainly a trial-and-error procedure involving 4 main steps:

1) Protein purification. Although purity requirements are not perfectly understood, the general understanding is that higher sample purities will result in better crystals. Protein purification techniques are described elsewhere in this work, so it will not be described in this section.

2) Solubilisation in a suitable solvent. Solubilisation is normally made in a water-based buffer solution, to which inorganic or organic salts and other agents that compete with the protein for interaction with water molecules can be added to reduce its solubility but without causing it to precipitate.

3) Supersaturation of the solution and nucleation. Supersaturation is the critical step, as it will force the protein out of the solution and small aggregates will start to form; these will be the nuclei or seeds for the crystal growth phase. Supersaturation is normally achieved by making the interactions between the protein and solvent less favourable, usually by adding a salt, polyethylene glycol (PEG), or an organic solvent. Other methods are also available, including changes in temperature and pH.

4) Crystal growth. Maintaining a high saturation during the growth phase would lead to the formation of multiple nuclei and therefore smaller crystals. Because of this, the crystal growth phase is better at a lower concentration. This also allows for a slower growth that, in theory, results in a maximum degree of order in its structure.

Crystallisation techniques. As the exact conditions to form suitable crystals are seldom known, screening of very many different conditions is required followed by optimisation of the selected conditions. Screening of a wide spectrum of physical and chemical parameters to identify the optimum sequence for achieving protein crystals is nowadays automated (Mosquito protein crystallisation robot, SPT Labtech), but it relies upon one or a combination of the following basic techniques:

1. Batch crystallisation. Precipitating reagent is added to the protein solution, instantaneously generating a supersaturated solution from which the protein will precipitate and in some cases form crystals.

2. Liquid-liquid diffusion. The protein solution and the solution containing the precipitant are layered on top of each other in a small-bore capillary, and they will gradually diffuse into each other.

*

3. Vapour diffusion – hanging drop/sitting drop. Drops are prepared on a siliconised glass coverslip by mixing equal volumes of protein solution and precipitant solution. The slip is placed upside-down over a depression in a tray partly filled with the precipitant solution. The chamber is sealed by grease, and equilibrium is reached by diffusion of vapour from the drop to the precipitating solution or vice versa. If the protein solution has low surface tension, a sitting drop arrangement can be used as an alternative to the hanging drop. In this case, the protein/precipitant mix is placed in a well sitting in a sealed chamber, with wells containing a precipitant solution. Vapour diffusion applies in the same way as the hanging drop technique.

4. Dialysis. A semi-permeable membrane is loaded with the protein solution and is suspended in a container with the precipitant solution. Equilibration kinetics depend upon the molecular weight cut-off of the membrane, the precipitant, the volume ratio, the concentration of the components inside and outside of the dialysis cell, and the geometry of the cell. Several variations of this technique exist, such as macro/microdialysis, Zeppenzauer capillary technique, and dialysis button.

Crystals and diffraction. X-ray crystallography is based on the analysis of the diffraction pattern caused by the sum of the interactions of the approximately 10^{15} molecules that make up a suitable crystal. The molecules in a crystal are arranged and oriented in specific 3-dimensional positions and orientations that cause the diffraction patterns of the individual molecules to add up, so the crystalline arrangement effectively amplifies the diffraction pattern that a single molecule would show.

Unit cells are the basic unit of a crystal; these are defined as the smallest parallelepiped described by vectors a, b and c and angles α, β, and γ, that when repeated by translations parallel to its edges in three directions without rotating the unit cell creates the crystal lattice **[15]**. The three directions in which the unit cell translates within the crystal also define its axis: the x-axis, y-axis, and z-axis are created in the a, b, and c directions respectively. Furthermore, the unit cell can have internal symmetry when two or more identical structures are related by axes or planes of symmetry. When these planes coincide with those of the crystal, the unit cell is said to have crystallographic symmetry. The individual unit that yields the entire unit cell when is repeated an integral number of times is called the asymmetric unit. Asymmetric units can crystallise in 65 different lattices, which are known as space groups **[15]**. The allowed geometrical translational operations of the

unit cell create 14 different types of lattices (Bravais lattices) which belong to 7 crystal systems.

Diffraction and Phase problem. The diffraction patterns break down the structure into discrete sine waves; any shape can be represented in 3 dimensions as the sum of sine waves of varying amplitudes and phases. The individual reflections of a diffraction pattern represent these waves, which have wavelength components in the three dimensions inversely proportional to their values of *h, k, l* (Miller indices).

The 3-dimensional object can be reconstructed by recombining the individual sine waves, but since X-rays cannot be focused only the diffraction pattern is recorded, and from this, the intensity and position of the reflected waves are registered and the space group of the lattice, the dimensions of the unit cell, the number of asymmetric units inside the unit cell as well as their volume can be determined. However, a 3-dimensional structure definition also requires their relative phases, which is known as the "Phase problem". There are different methods to solve it **[11]**:

- Isomorphous replacement: Phases are determined by combining diffraction data from a native crystal with data from other crystals containing the same protein packed in the same structure but marked with a heavy atom.

- Molecular replacement: previously known structures provide approximate phases that can be used to solve the new structure.

- Multiwavelength Anomalous Dispersion (MAD): Using a tuneable X-ray source (through a synchrotron) phase information can be obtained through the solution of crystal structures from measurements of the variation of the intensity distribution in the diffraction pattern over a range of wavelengths.

Model building, structure refinement, and validation. Once the diffraction pattern is recorded, and the phases of each of the reflections are determined, the unit cell structure is calculated through Fourier transformation. The electron density ρ at a point (x, y, z) in the unit cell is given by:

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{k} F\,(h,k,l)e^{i\alpha(h,k,l)}e^{-2\pi i(hx+ky+lz)}$$

Equation 3. Fourier synthesis calculation used to approach the calculation of the electron density map.

Where V is the volume of the unit cell, F(*h, k, l*) is the amplitude of the reflection with indices *h, k,* and *l,* and α*(h, k, l*)is its phase. As can be seen, the electron density at each point includes the amplitude and the phase of all the reflections of the diffraction pattern. The level of detail achieved in a crystal structure depends then on the number of reflections included in the summation of the Fourier transformation or synthesis; adding waves of decreasing wavelength (higher values of *h, k,* and *l*) will confer increasing detail to the density map, but in practice, it is necessary to limit the Fourier synthesis at some maximum *h, k,* and *l* values to consider only those reflections included in a sphere around the origin of the reciprocal lattice to limit the apparent detail **[15]**.

The quality of the electron density map will also depend on the regularity of the crystal. This last aspect is a critical limiting factor as the resulting electron density map is an average overall the unit cells of all the crystals and over the whole time the data is being collected, during which the X-rays may be damage the protein. This disorder in the crystal will be seen as a smearing of the electron density. Model building and model refining are connected. Building the initial model and real space fitting of the model into the electron density is continually alternated with restrained reciprocal space refinement of the positional parameters of the model **[17]**. This is generally considered the most labour-intensive part of the structure determination but is currently almost entirely made by semi-automated or fully automated model building computer programs.

Several specialised programs are available that can perform one or all of the building and refining processes as well as building a graphic model: ARP/wARP, CCP4, Phenix, Phaser, and a long list, however, proper parametrisation of the programs is required to achieve a model that makes good chemical and biological sense. Specific crystallographic methodologies and experimental procedures are described in the corresponding section (Results Chapter 1).

2.6.7 Electron paramagnetic resonance

Electron paramagnetic resonance (EPR) spectroscopy, is a highly sensitive spectroscopic technique, also known as electron spin resonance (ESR) spectroscopy or electron magnetic resonance (EMR) spectroscopy, used for the study of molecular systems

with unpaired electrons and the atomic nuclei located in their proximity. Although limited to substances with unpaired electron spins, analysis of the fundamental splitting of energy levels of spins concerning their orientation in an external magnetic field, interactions between paramagnetic spin systems and their local environments can be detected.

EPR spectra are highly sensitive to the local electronic structure, oxidation state, and the proximity of magnetic nuclei to the system in question, and therefore this resonance spectroscopy is the preferred technique for the determination of the electronic structures, dynamics, and the geometries or spatial distribution of species ranging from radicals, transition metal complexes, rare-earth materials, semiconductors and paramagnetic metal clusters in biomolecules **[18]**. Optical spectroscopic methods consist of a source of radiation, a sample, and a detector; magnetic spectroscopic techniques, such as EPR, Nuclear Magnetic Resonance, Magnetic Circular Dichroism, or Mossbauer spectroscopy also require an external magnetic field. However, only in EPR, the magnetic field is varied while keeping the EM radiation frequency at a constant value so that the amount of radiation absorbed by the sample is thus measured as a function of the magnetic field, as opposed to all other magnetic spectroscopic techniques where the magnetic field is kept constant and is the EM frequency is variable.

When an external magnetic field is applied to a molecule with unpaired electrons, it will cause the splitting of the electron spin moments into two populations: either parallel or antiparallel with the external field, each population having different energy states. The difference between the two energy states will be proportional to the external field and this relation can be written as:

$$\Delta E = h\nu = g\,\mu B\,B$$

Equation 4. Mathematical representation of the Zeeman splitting effect to describe the energy levels of unpaired electrons in presence of an external magnetic field.

Where ν is the spectrometer frequency, h is Plank's constant, μB is the Bohr's magneton, and the external magnetic field is B. The constant g is the free electron g-factor which is related to the total angular momentum; in addition to considering the total magnetic dipole moment of a paramagnetic species, the g-value takes into account the local environment of the spin system. The existence of local magnetic fields produced by other paramagnetic species, electric quadrupoles, magnetic nuclei, ligand fields (especially in the

case of transition metals) can change the effective magnetic field that the electron experiences and it is, therefore, the main goal of the EPR spectroscopy analysis.

In order to be able to determine the *g* value for a given system, EPR relies upon tuning the magnetic field *B* to match the energy difference ΔE, as it becomes equal to the energy of the incoming radiation *hv*, thus reaching a state of resonance. In the resonant state, the incident electromagnetic radiation is absorbed for the transition of particles from the lower to the higher state and eventually emitted when the particles relax from the higher to the lower state (stimulated emission). Normally, in an initial state more molecules are in the lower state (ground state) than in the higher state (excited state), and resonance will therefore result in net absorption of the radiation. EPR spectrometers use radiation in the Giga Hertz range (GHz is 109 Hz), and the most common type of spectrophotometer operates with radiation in the X-band of microwaves in the range of 9–10 GHz. Samples were prepared for EPR as isolated in 300 µL aliquots at 4 $mgmL^{-1}$ in the appropriate working buffer solution and transferred into 4 mm Suprasil quartz EPR tubes (Wilmad, USA) under anaerobic conditions. The samples were directly frozen and stored in liquid nitrogen. EPR experiments were conducted using the parameters as follows: microwave power 0.5 mW, modulation frequency 100 kHz, modulation amplitude 5 G, temperature 30 K. Spectra were obtained using a Bruker ELEXSYS E500 spectrometer, Super high Q resonator (ER 4122SHQ), Oxford Instruments ESR900 cryostat and ITC503 temperature controller.

2.7 *In silico* methodology

2.7.1 PSI-BLAST

PSI-BLAST search (Position Specific Iterated Basic Local Alignment Search Tool) against the NCBI (RefSeq) database (ncbi.nlm.nih.gov), using the translated sequence of the gene of interest (ssRdhA) was made. PSI-BLAST relies on a PSSM (Position Specific Scoring Matrix) built automatically from a multiple sequence alignment dataset produced after an initial BLAST search. We obtained significative alignments of 100 sequences after 10 iterations, presented according to the parameters of E (that reflects the probability of two sequences being similar to each other by chance and not because these are evolutionary related) and score (that indicates the coincidence between the aa residues between a pair of sequences).

The compositional matrix used was BLOSUM62, with the following algorithm parameters: Gap 10, Existence 11, Extension 1, and PSI-BLAST threshold 0.001. Sequence

selection was done considering the quantitative values of similitude and identity generated during the alignment, that indicate the conservation or presence of identical sites on the aligned sequences, and the similarity of the sequences depending on the physicochemical properties of the aa residues, respectively. We recovered the sequences that presented identity of more than 15% with respect to the query and presented E values tending to 0. The sequences were downloaded in FASTA format and curated by hand, eliminating the sequences from the same genera **[19-21]**.

2.7.2 Modeling and molecular dynamics, docking, and additional programs

The homology-models were generated using the I-TASSER server, designed for automated structure-function prediction that constructs the models using a threading alignment approach based on templates from the PDB **[22]**. The output file contains information about the coordinates, predictions of contacts in the 3D structure, intermediary modelling states, and information about the putative function, based on information from the public databases. To correct the anomalies on the interatomic bond lengths and angles but respecting the connectivity of the 3D structure while restricting the movement of the backbone to prevent drastic changes in the folding. Our original I-TASSER model was subjected to energy minimisation according to the following strategy: 1) Restrict the movement from the side groups until the backbone is minimised to convergence. 2) Minimise the side chains while limiting the movement of the backbone until 50 cycles of minimisation have been reached. This procedure was repeated iteratively until the system converged and then, an energy minimisation with no restrictions was made to adjust the final geometry.

We used the Amber forcefield in HyperChem 8.0 (Hypercube). After that, we performed a molecular dynamic simulation (MD). MD consists of predicting the movement of an atom or group of atoms, under a defined potential given by a force-field. The simulation was performed for 5 ns using GROMACS 2018.8, in a box with octahedral geometry with 1.5 nm of distance to the wall, using water as the solvent and a fixed concentration of 100 mM of NaCl at 303 K. Control of pressure and temperature was attained by the coupled Berendsen thermostat and barostat. We used the *leap-frog* algorithms for the integration of the classic movement equations. We used a restricted simulation where the backbone and side chains of the protein were fixed to certain coordinates inside the box by the application

of an harmonic force of 1000 N in x,y,z directions. This is done to favour the approximation of solvent molecules to the protein to achieve a solvation layer, this simulation proceeded for 200 ps and the neighbours list was updated every 20 fs. After that, we performed the productive simulation without restrictions. The more populated conformer at the end of the simulation was obtained by a clustering algorithm according to its occurrence during the simulation. The conformer was then minimised again for 5000 steps. During the 5 ns of MD important differences were observed in the last conformer in comparison with the initial I-TASSER.

These new models were then scored using the Rosetta design-HMMer (Rd.HMM). In this protocol, a 3D homology model is considered to be close to an equilibrium structure and certain energy values are assigned to it; if it retrieves its original primary sequence from the databases, with a score close to 0.6 times the length of its amino acid sequence, and the alignment produced by the Rd.HMM does not show gaps, nor frame-shifts, the model will be considered acceptable and no energy penalisation will occur. The score of the original amino acid sequence should be amongst the top scores (or ideally the first), and the sequences in this group should present high sequence similarity amongst them (usually above 90% identity for close orthologous, such as the NpRdhA). Rd.HMMer was performed using 15 intermediates with randomised sequences and each was reconstructed 10 times. The searches were done against the RefSeq-protein sequence database at NCBI.

Note:

Highly specialised protocols are included again in the respective results chapters as part of a Journal format.

References

1. T7 RNA Polymerase Expression System Handbook, https://www.mobitec.com/media/mobitec/category-preview/vector-systems/pdf/1/T7_RNA_Polymerase_Expression_System-Handbook.pdf, (accessed October 2020).
2. pET System Vectors and Hosts Instruction Manual, https://www.agilent.com/cs/library/usermanuals/Public/211521.pdf, (accessed October 2020).
3. Ion Exchange Chromatography Principles and Methods, https://cdn.cytivalifesciences.com/dmm3bwsv3/AssetStream.aspx?mediaformatid=10061&destinationid=10016&assetid=13101, (accessed October 2020).

4. E. Gasteiger, A. Gattiker, C. Hoogland, I. Ivanyi, R. D. Appel and A. Bairoch, *Nucleic Acid Research*, 2003, **31**, pp. 3784–3788.
5. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy (Second Edition)*, Springer, New York, USA, 2004.
6. MicroCal ITC Range, https://www.malvernpanalytical.com/en/products/product-range/microcal-range/microcal-itc-range, (accessed 20 October 2020).
7. M. M. Pierce, C. S. Raman and B. T. Nall, *Methods*, 1999, **19**, pp. 213-221.
8. J. L. Cole, J. W. Lary, T. Moody and T. M. Laue, *Methods Cell Biol.*, 2008, **84**, pp. 143-179.
9. D. J. Scott, S. E. Harding and A. J. Rowe, in *Analytical Ultracentrifugation: Techniques and Methods*, eds. D. Scott, S. E. Harding and A. Rowe, Royal Society of Chemistry, London, 2005, pp. 1-25.
10. A. Furst and D. N. Taulbee, in *Encyclopedia of Analytical Science*, eds. P. Worsfold, A. Townshend and C. Poole, Elsevier, Amsterdam, 2005, pp. 463-469.
11. A. M. Lesk, *Introduction to Protein Science*, Oxford University Press, Oxford, 2004.
12. G. Tsiminis, E. P. Schartner, J. L. Brooks and M. R. Hutchinson, *Appl Spectrosc Rev*, 2017, **52**, pp. 439-455.
13. G. Tsiminis, E. P. Schartner, J. L. Brooks and M. R. Hutchinson, *Measurements of vitamin B12 in human blood serum using resonance Raman spectroscopy*, IEEE, Waikoloa, HI, 2016.
14. G. l. Büyükköroğlu, D. D. Dora, F. Özdemir and C. Hizel, in *Omics technologies and bio-engineering : towards improving quality of life*, eds. D. Barh and V. Azevedo, Academic Press, London, 2018, pp. 317-351.
15. T. E. Creighton, in *Proteins. Structures and Molecular Properties*, W. H. Freeman and Co, New York, 1993, pp. 202-214.
16. J. R. Luft, E. H. Snell and G. T. Detitta, *Expert Opin Drug Discov*, 2011, **6**, pp. 456-480.
17. B. Rupp, in *Biomolecular Crystallography. Principles, Practice, and Application to Structural Biology*, Garland Science, Abingdon, 2010, pp. 1-22.
18. P. Kaur, P. Oyala and Y. Guo, EPR: Theory, https://chem.libretexts.org/Bookshelves/Physical_and_Theoretical_Chemistry_Textbook_Maps/Supplemental_Modules_(Physical_and_Theoretical_Chemistry)/Spectroscopy/Magnetic_Resonance_Spectroscopies/Electron_Paramagnetic_Resonance/EPR%3A_Theory, (accessed October 2020).
19. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, pp. 1189-1191.
20. S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Mol Biol Evol*, 2018, **35**, pp. 1547-1549.
21. R. D. M. Page and E. C. Holmes, *Molecular Evolution. A Phylogenetic Approach*, Blackwell Science, Cambridge, 1998.
22. A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, pp. 725-738.

3.0 Chapter Three. Results Chapter I

Crystal structures of the MarR-type regulator $RdhR_{CbdbA1625}$ provide insight into sensing of chloroaromatic compounds

Running title: Structures of a MarR-type organohalide sensor.

Samantha Gaytán[1], Carolina P. Quezada[2], Thomas Jowitt[3], Mark Dunstan[3], Sam Hay[1], David Leys[1]*

[1] Manchester Institute of Biotechnology, Faculty of Life Sciences, The University of Manchester, Manchester M1 7DN, United Kingdom

[2] Centro Intregativo de Biología y Química Aplicada (CIBQA), Universidad Bernardo O'Higgins, General Gana 1702, Santiago PC 8370993, Chile

[3] SynBioChem, Faculty of Life Sciences, The University of Manchester, Manchester M1 7DN, United Kingdom

* To whom correspondence may be addressed: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. Tel.: +44-161-306-5150; E-mail: david.leys@manchester.ac.uk

Manuscript in preparation for submission to the Journal of Biological Chemistry (JBC).

Keywords: *Dehalococcoides*, organohalide respiration, reductive dehalogenation, MarR-type transcriptional regulation, crystal structure, ligand-binding protein, DNA-binding domain, binding protein, microbiology.

3.1 Abstract

Reductive dehalogenases or RdhAs, are encoded in the genome of all organohalide respiring bacteria (OHRBs) known to date. The expression of the structural genes *rdhA* and *rdhB* is regulated in response to the presence of organohalides that serve as terminal electron acceptors during respiration. The responsible transcriptional regulators have distinct evolutionary origins and include members of the Cpr-Fnr and MarR (Multiple

Antibiotic Resistance Regulators) protein families. While the mechanism of the organohalide sensor CprK is relatively well understood, a detailed description of the RdhR (a multiple antibiotic resistance regulator or MarR-type regulator) is missing. Here, we present the crystal structure of the $RdhR_{CbdbA1625}$ from *Dehalococcoides mccartyi* strain CBDB1, both in the ligand-free state and in complex with various chlorophenol ligands. These reveal the presence of three distinct chlorine-binding sites. Solution studies suggest that the relative position of the ligand chlorine substituents on trichlorobenzene (TCB) compounds determines the affinity of the $RdhR_{CbdbA1625}$ for the ligands. Tight-binding of the ligands was observed to affect the stability of the protein-DNA complex, thus suggesting the physiological role for the $RdhR_{CbdbA1625}$ as a transcriptional repressor.

3.2 Introduction

The recent developments in metagenomic sciences have led to a substantial increase in the number of microorganisms found capable of using organohalides as final electron acceptors during organohalide respiration. These microorganisms are generally grouped as "organohalide respiring bacteria" (OHRBs) [1-3]. Although relatively limited information about the biochemistry and catalytic mechanisms of dehalogenation in both aerobic and anaerobic OHRBs is available, it is clear that the single unifying feature in these Bacteria is the presence of the tightly regulated genes that encode reductive dehalogenases (RdhAs) in their genomes [4].

The *Dehaloccoidetes* class of OHRBs are irregular coccoid-like bacteria [5] that belong to the obligate anaerobic phylum Chloroflexi, known to live in highly specific ecological niches, thriving in aquifers, river sediments, industrial and sewage sludges [6]. They rely on organohalides as final electron acceptors in energy conservation pathways with H_2 as the sole electron donor and either acetate or CO_2 as the only carbon sources [7, 8]. *Dehalococcoides mccartyi* strain CBDB1 is one a model microorganism specialised in reductive dehalogenation, known for its ability to reductively dehalogenate a wide range of both aliphatic and aromatic chlorinated compounds. These include 1,2,3-trichlorobenzene and 1,2,4-trichlorobenzene, but also two of the most toxic environmental pollutants: pentachlorophenol (PCP) and polychlorinated dibenzo-p-dioxins (PCDD) [6, 9, 10]. *D. mccartyi* strain CBDB1 has a genome of about 1.39 Mbp, considered a rather small genome for a free-living bacterium that interestingly, encodes a high number of putative reductive dehalogenase paralogous genes (32 *rdhAs*) [9]. It is precisely the abundance of structural

genes encoding multiple reductive dehalogenases that highlights its potential to reductively dehalogenate multiple substrates [11], and it is what makes *Dehalococcoides* bacteria one of the most widely used OHRBs in large-scale bioremediation process *in situ* [12, 13]. Although there is a high degree of conservation in the *Dehalococcoides spp*. genome, a remarkable variability in the regions that contain the *rdh* genes has been noted. The presence of flanking mobile elements in some of the *rdhAB* clusters suggests that there has been great evolutionary pressure in these genomic areas, which might include mechanisms of gene duplication and transfer, as is implied by the presence of multiple paralogous genes within the genome of various OHRBs [9, 14].

It has been determined that most reductive dehalogenase genes are organised in bicistronic operons, composed mainly of two genes (*rdhA* and *rdhB)*, encoding the catalytic subunit of the respiratory reductive dehalogenase RdhA, a B_{12}-dependent enzyme that also depends on two cubane type iron-sulfur clusters (4Fe-4S) for its activity, and a putative membrane-anchor protein RdhB [3, 15]. Other accessory sequences encoding proteins involved in the maturation of the RdhA such as the chaperone RdhT and a group of translocases are also encountered in the gene clusters, together with multiple transcription factors [16]. Up to date, very few of the *Dehalococcoides* spp. RdhAs have been studied, due to their low expression yields and the complexity of studying O_2-sensitive membrane proteins. In *D. mccartyi* strains 195 and CBDB1, only the PceA and TceA (that dechlorinate of PCE and TCE to vinyl chloride (VC)) and the CbrA (that dechlorinates polychlorinated benzenes), have been characterised at a basic biochemical level [4, 8, 10]. For the past years, research has focused on the organohalide sensing mechanisms of the transcriptional regulators, or *rdhR* genes, intending to determine their specific physiological effectors. This in turn would suggest the substrate specificity of the corresponding structural *rdhA* genes that they regulate, and serve as an alternative approach to challenging expression and characterisation RdhAs systems [17]. Transcriptional regulation of *rdhAB* pairs can be associated with members of the Cpr-Fnr family, the two-component systems and the MarR family of transcriptional regulators [18]. In *D. mccartyi* most of the structural *rdhA* genes are clustered with genes that encode either the two-component signal transduction systems or the MarR transcriptional regulators. [14, 19].

The initial identification of the MarR family of transcriptional regulators began with the recognition of the *marR* phenotype in *Escherichia coli* strain K-12, that confers resistance to antibiotics [20] and is encoded by the *marRAB* operon [21]. Transcription factors of this superfamily control a wide variety of biochemical and physiological

responses to several biotic and abiotic stresses. These include oxidative stress, pathogen infection, virulence factor production and catabolism of aromatic compounds used as disinfectants, organic solvents and some of the most widely used antibiotics, hence its relevance for the environmental sensing of aromatic compounds [22, 23].

A large number of MarR-type family proteins has been found in the phylogenetic domains of Eubacteria and Archaea at the moment of writing, more than 34 527 protein sequences that exhibit the typical domain architecture have been annotated in the databases although the biochemical and structural characterisation has only been reported for a few of them [24, 25]. MarR-type regulators are small proteins of 17-22 kDa, that present an overall α/β fold that exhibits six α-helixes and two β-strands. The residues in both N-terminal and C-terminal domains contribute to the interface of dimerisation. These transcription factors form triangular homodimers in solution that can interact with specific double-stranded DNA (dsDNA) binding sites, that frequently overlap with the -35 and/or -10 promoter elements of the regulated genes, while the second often overlapping with the ribosome binding site (RBS). This suggests that repression of the target gene expression occurs by competing with the RNA polymerase and ending just before the initial codon of the *rdhR* gene. The intergenic regions are located between divergent promoters, consequently, the MarR regulators control the expression of all the genes encoded in their operon and frequently, its expression. Both binding sites are palindromic, with two 5 bp half-sites separated by 2 bp, according to foot-printing experiments [22, 25, 26]. The MarR family of proteins mediates the interaction with DNA via their conserved winged helix-turn-helix (winged HtH)[1] motif that permits the recognition of the target sequence, a dimerisation domain, and frequently more than one ligand binding site per dimer for their putative physiological effectors. The interaction between the dsDNA and the transcription factor *in vivo* is controlled or regulated by the presence of certain anionic hydrophobic molecules, often aromatic compounds, for example, salicylates, benzoates, and m-chlorophenylhydrazone derivatives. Some MarR homologues have also been reported to respond to oxidative stress by binding Cu^{2+} [26, 28, 29].

Although most of the MarR transcriptional regulators act as repressors, it has been reported that some can also act as activators or as both, as was shown in the case of the SlyA

[1] Helix-turn-helix motifs are included in many DNA binding prokaryotic proteins that recognise and bind the specific regulatory regions of DNA. The two α helixes have the same orientation relative to each other and are connected by a loop region similar in all HtH motifs [27].

from *Enterococcus faecalis* [30, 31]. The competition with the RNA polymerase is not the only possible repressing mechanism, and some regulators prevent the transcription by affecting the elongation process, by preventing the promoter escape [32] or by competing with transcription activators for binding at the promoter sites. In all cases, the binding of the ligand has a negative allosteric effect on dsDNA-binding, thus relieving the repression.

The role of $RdhR_{CbdbA1625}$ as a transcriptional regulator has previously been studied, suggesting that it acts as a repressor *in vivo* [8, 32-34]. However, the exact signal molecule to which $RdhR_{CbdbA1625}$ responds remains unknown. Here, we present the crystal structure of the $RdhR_{CbdbA1625}$, the first crystal structure of an organohalide-sensing MarR-regulator, both in the ligand-free state and bound to dichlorophenolic and trichlorophenolic ligands. Initial binding of dichlorophenolic compounds allowed a fragment-merging approach to suggest high affinity for 1,2,3-trichlorobenzene compounds. Indeed, tight binding of 2,3,4-trichlorophenol could be observed in solution by fluorescence spectroscopy and found to affect the binding of the $RdhR_{CbdbA1625}$ to the DNA. We discuss the implications of these findings for our understanding of the transcriptional regulation of the reductive dehalogenases by organohalide respiring bacteria.

3.3 Experimental procedures

3.3.1 Cloning of *Dehalococcoides mccartyi* CBDB $RdhR_{CbdbA1625}$

The codon optimised *rdhR* (cbdb_A1625) gene (GenBank accession number WP_011309983.1) from *Dehalococcoides mccartyi* strain CBDB1 was amplified from pASK-IBA5plus-$RdhR_{CbdbA1625}$ (a generous gift from Anke Wagner, Martin Luther University of Halle-Wittenberg, Germany) using Phusion polymerase (New England Biolabs) and the following primers: 5'-AAGTTCTGTTTCAGGGCCCGAACGAATTTGAGACTCTGG-3' (forward) and 5'-ATGGTCTAGAAAGCTTTAAATGTATTTAAACGGGGTCAGGG-3' (reverse). The PCR product was cloned into pOPINF vector encoding an N-terminal hexahistidine tag (OPPF Oxford, UK) using In-Fusion HD (Clontech). The r*dhR* (cbdb_A1625) gene was also cloned into pEt30a (+) vector using Q5 polymerase (NEB) and the following primers: 5'-AAGGAGATATACATATGAACGAATTTGAGACTCTGGAACC-3' (forward) and 5'-GGTGGTGGTGCTCGAGTTATTTTTCGAACTGCGGGTGGCTCCAAGCGCTAATGTATTTAAACGG GGTCAGGGT -3' (reverse) that encode a streptavidin tag. The correct insertion of the gene was checked by DNA sequencing (MWG Eurofins). The plasmid was introduced via transformation into *E. coli* BL21 (DE3) pLysS cells (NEB) for protein overexpression.

3.3.2 His-tag RdhR$_{CbdbA1625}$ heterologous expression and purification

Transformed *E. coli* BL21 (DE3) colonies were selected and inoculated in 50 mL of LB media (Formedia) for over-night cultures. LB media was supplemented with ampicillin 100 μgmL^{-1} and chloramphenicol 34 μgmL^{-1}. Next day, 2 mL of the over-night growth were used to inoculate 12x1 L of LB media and incubated at 37 °C until an OD^{600} of 0.6 was reached. To induce RdhR$_{CbdbA1625}$ overexpression, IPTG 0.2 mM was added and the temperature was lowered to 20 °C, at constant agitation (180 rpm). After an overnight expression, cells were harvested by centrifugation and resuspended in Tris-HCl 50 mM pH 7.5, 300 mM NaCl and supplemented with Complete EDTA-free protease inhibitors (Sigma), RNAse and DNAse I (Sigma). Resuspended cells were lysed using a cell disruptor (Constant Systems). The lysate was clarified by centrifugation at 40,000 rpm for 1 hour at 4 °C in an Optima CE-80K ultracentrifuge (Beckman-Coulter). The supernatant was loaded onto a 5 ml of Ni-TDE Agarose resin (Macherey-Nagel) preequilibrated with the buffer solution, after passing all the supernatant through the column, the resin was washed with 50 mL of the working buffer and then with 25 mL of Tris-HCl 50 mM pH 7.5, NaCl 300 mM and imidazole 10 mM. Elution of the His-tag RdhR$_{CbdbA1625}$ protein was achieved by adding 25 mL of Tris-HCl 50 mM, NaCl 300 mM, and imidazole 100 mM, and a single elution fraction was recovered. All purification steps were performed at 4 °C. Samples of each fraction were loaded onto a 10% Run Blue precast SDS-PAGE gel cassette, ran at 150 V, and stained with Instant Blue (Expedeon). The combined elution fractions were transferred to a desalting-column (Bio-Rad), into HEPES-KOH 50 mM pH 7.5, NaCl 300 mM. The elution fractions that contained the pure protein were further purified using a 5 mL Hi-Trap Heparin column (GE Healthcare). The elution of the protein was achieved at an approximate NaCl concentration of NaCl. RdhR$_{CbdbA1625}$ was concentrated using a Vivaspin 10 kDa cut-off spin concentrator (GE) to a final concentration of 15 mgmL^{-1} (as estimated by UV-Visible at 280 nm using a Cary 60 spectrophotometer (Agilent)) and fast-frozen at -80 °C.

3.3.3 Strep-tag RdhR$_{CbdbA1625}$ heterologous expression and purification

Single transformed *E. coli* BL21 (DE3) colonies were selected and inoculated in 50 mL of LB media (Formedia) for over-night cultures. LB media was supplemented with kanamycin 50 μgmL^{-1}. The next day, 2 mL of the over-night growth were used to inoculate 12x1 L of LB media and incubated at 37 °C until an OD^{600} of 0.6 was reached. To induce RdhR$_{CbdbA1625}$ overexpression, IPTG 0.2 mM was added and the temperature was lowered to

20 °C, at constant agitation (180 rpm). After growing overnight for approximately 18 h, cells were harvested by centrifugation at 6000 rpm. The pellet was resuspended in Tris-HCl 50 mM pH 7.5, NaCl 300 mM supplemented with Complete EDTA-free protease inhibitors (Sigma), RNAse, and DNase I (Sigma). Resuspended cells were lysed using a cell disruptor. The lysate was clarified by centrifugation at 40,000 rpm for 1 hour at 4 °C in an Optima CE-80K ultracentrifuge (Beckman-Coulter). The supernatant was loaded onto 20 mL of the preequilibrated Streptactin resin (IBA) and then washed with 100 mL of the same buffer solution. Elution of the samples was achieved by washing with 6x10 mL of the Tris-HCl 50 mM pH 7.5, NaCl 300 mM buffer supplemented with d-desthiobiotin 2.5 mM. Samples of each fraction were loaded onto a 10% Run Blue precast SDS-PAGE gel cassette, ran at 150 V, and stained with Instant Blue (Expedeon). The elution fractions that contained the pure protein were further purified with a 5 mL Hi-Trap Heparin column (GE) in the ÄKTA pure protein purification system (GE). The elution of the protein was achieved at approximately a concentration of NaCl 600 mM. All purification steps were performed at 4° C. $RdhR_{CbdbA1625}$ was concentrated using a Vivaspin 10 kDa cut-off spin concentrator (GE) to a final concentration of 10 $mgmL^{-1}$ (as estimated by UV-Visible absorbance at 280 nm using a Cary 60 spectrophotometer (Agilent)) and fast-frozen and stored at -80 °C, except for crystallisation trials, where the protein was used fresh.

3.3.4 Trp Fluorescence Spectroscopy: Intrinsic fluorescence quenching

Trp fluorescence quenching experiments were performed using a FLS920 series fluorometer (Edinburgh Instruments Ltd.) equipped with a 450 W Xe900 Xenon arc. A 1.0 mL cuvette with a 10 mm path-length quartz fluorescence cuvette from Hellma Analytics was used. Alternatively, a 0.3 mL cuvette with a 3 mm path-length quartz cuvette from Starna Scientific form was used also to reduce the inner filter effects during the titration when required. A sample of $RdhR_{CbdbA1625}$ (final concentration 0.25 μM) was titrated with increasing concentrations of three possible physiological effectors 2,3,4-trichlorophenol, 2,3,5-trichlorophenol, and 3,4,5-trichlorophenol (Sigma) until saturation was reached.

The same protocol was followed in case of the different DCPs tested: 2,3-DCP, 2,4-DCP, 2,6-DCP, and 3,4-DCP (Sigma and Fluorochem), but the $RdhR_{CbdbA1625}$ final concentration during the assay was 1.0 μM. All the compounds were solubilised in analytic grade MeOH (Sigma), suitable for spectroscopy. The buffer solution HEPES-KOH 50 mM, pH 7.5, and NaCl 500 mM] was filtered and degassed. The fluorometer was started 30 minutes

before use. Excitation and emission bandwidths were set at 2.0 and 5.0 nm, respectively. Spectra were recorded with 0.5 nm steps and dwelling times between 0.2 and 0.5 s. Emission scans were performed by exciting the Trp residues at 280 nm and monitoring the intrinsic fluorescence between 290 nm and 420 nm. Excitation scans were performed by monitoring the emission fluorescence at 333 nm (fluorescence maximum) by scanning between 250 nm to 320 nm. All experiments were performed at 25 °C. Dilution control experiments were made in all cases. The emission, excitation, and UV-Visible absorption scans were recorded at each titration point. The area behind the curve in each spectral dataset was normalised and fitted to the quadratic equation for tight binding or irreversible inhibitors (**Equation 1**), in the case of the trichlorophenol ligands and to a non-linear single-site binding model for the dichlorophenols. For the DCPs a non-linear function of the in-built fitting tool NLFit from OriginPro 9.1 ® [35] similar to a Michaelis-Menten model was used. In case of the TCPs, the area under the curve for each concentration of ligand added to the cuvette was normalised and then fitted to a quadratic velocity model, by manually defining the function (**Equation 1**. Morrison equation for tight-binding ligands) and declaring the variables using the Function Builder (OriginPro 9.1 [35]) until convergence was reached after a number of iterations, according to the software manual [35]. Corrections for inner filter effects were made as described elsewhere [36]. Five independent experiments were performed for each compound tested.

$$v = \frac{([E_T] + [S_T] + K_D) - \sqrt{([E_T] + [S_T] + K_D)^2 - 4[E_T][S_T]}}{2E_T}$$

Equation 1. Morrison or quadratic velocity equation.

A sample of $RdhR_{CbdbA1625}$ (final concentration of 2.5 µM) in the buffer solution HEPES-KOH 50 mM, pH 7.5 and NaCl 500 mM was filtered, degassed and then titrated with growing concentrations of the 42 bp DNA palindrome 5'-GTATATTAGTCTATATGGACTAGTCCATATAGACTAATATAC-3' (Tm 65.5° C) synthesised and HPLC purified by MWG Eurofins, until saturation conditions. Before the titration experiment, the DNA was incubated at 60 °C for 5 min using the T100 Thermal Cycler and left to cool down slowly for hybridisation (1 °Cmin^{-1}) (Bio-Rad). Stocks and buffers were prepared with nuclease-free water. DNA concentration was determined using a Nanodrop 2000c; protein concentration was determined by UV-Visible spectroscopy according to the A^{280} measurements using a Cary 60 spectrophotometer (Agilent).

A quartz cuvette of 3 mm path length from Starna Scientific was used to perform the experiment to reduce the inner filter effects caused by the DNA. Spectra were recorded with 0.5 nm steps and dwelling times of 0.5 s. Emission scans were performed by exciting the Trp residues at 280 nm and monitoring the intrinsic fluorescence between 290 nm and 420 nm. Excitation spectra were obtained by measuring the emission at 333 nm and 360 nm between 250 nm to 320 nm and 250 to 350 nm. Dilution control experiments were made in all cases. The emission, excitation, and UV-Visible absorption scans were recorded at each titration point. An inner filter effect correction, as described by [36, 37] was made for each titration point. The data was fitted to a non-linear regression model using Origin Lab software [35]. Both mathematical models were applied under the assumption of rapid equilibrium, but in case of the tight-binding model, the free-ligand approximation that states that the free-ligand concentration (or substrate in the Michaelis-Menten model) is equivalent to its total concentration in solution does not apply under steady-state conditions when the substrate's K_M (or K_D) is equal or lower than the total enzyme concentration [38].

3.3.5 Agarose electrophoretic mobility shift assays (EMSAs)

For the EMSA assays, we followed the general protocol of Davis *et al.* [34]. DNA concentration was determined using a Nanodrop 2000c®; protein concentration was determined by UV-Visible spectroscopy according to the A^{280} measurements using a Cary 60 spectrophotometer (Agilent). The palindromic DNA and the random DNA were resuspended to a 100 µM concentration using nuclease-free water and then incubated at 60 °C for 5 min and then left to cool down slowly to ensure the formation of the DNA duplex (1 °Cmin^{-1}) using a T100 Thermal Cycler (Bio-Rad). The reaction mixtures were prepared using the binding buffer 5X Tris-HCl 50 mM, KCl 50 mM, $MgCl_2$ 5 mM, EDTA 1.0 mM Glycerol 5% (v/v), poly (dI-dC) 5 ngμL^{-1}, DTT 1mM, in a final volume of 10 µL. The dsDNA probes were added to a final concentration of 25 nM to the $RdhR_{CbdbA1625}$ (2.5 µM). The effect of the TCP ligands on the binding of $RdhR_{CbdbA1625}$ to the dsDNA probes was tested by adding the compounds to a preformed protein-DNA complex, after mixing and incubating the samples for 15 min at 25 °C.

The compounds 2,3,4-trichlorophenol, 2,3,5-trichlorophenol, and 3,4,5-trichlorophenol (Sigma) were dissolved in analytical grade MeOH (Sigma). The ligands were added to the reaction mixtures to a final concentration of 10 mM, including the phenol control. The reactions were further incubated at 30 °C for 15 min. Before loading into a 1.0

mm/ 10 well Novex® TBE precast 6% DNA retardation gel (Invitrogen) and Orange Dye 1X (Licor) were added. The running buffer Tris-borate 5X (TBE) (Invitrogen) was diluted to 1X and ran at 100 V (8-10 mA) for 1 h. The gel tank was placed at 4 °C in the dark. Finally, the gel was treated with SYBR Gold DNA Stain (Invitrogen) 1X for 15 min and further washed with nuclease-free Milli-Q water. The gel was kept protected from light. Gels were visualised using the Typhoon FLA6000.

3.3.6 Isothermal titration calorimetry ITC

To determine the affinity and binding parameters (K_D) of the $RdhR_{CbdbA1625}$ to its palindromic DNA sequence **(Table 1)**, isothermal titration calorimetry (ITC) experiments were made in a Microcal PEAQ-ITC calorimeter (Malvern Instruments) by titrating a solution of freshly purified protein by using gel-filtration chromatography with an analytical column S-200 (GE); the sample was further dialysed for 1 h at room temperature. Approximately 60 µM of the $RdhR_{CbdbA1625}$ were loaded into the sample cell to be titrated with increasing concentrations of a 200 µM stock of the target 42 bp dsDNA. Both the palindromic and the randomised DNA used as binding control were synthesised and HPLC purified by MWG Eurofins; the lyophilised samples were resuspended and further dialysed in the same working buffer solution of Tris-HCl 50 mM pH 7.5 and NaCl 300 mM, prepared in nuclease-free Milli-Q water. The buffer solution was thoroughly filtered and degassed previously to the assays and the tubing and material to be used during the titration were cleaned prior to each experiment with nuclease free-water Milli-Q water with DCPEP 0.1% (Sigma). A range of concentrations of NaCl (200-500 mM) were tried until the appropriate conditions to prevent the protein aggregation were found. The ITC experiments were performed in triplicate, each one with 20 injections of 10 µL with constant stirring at 25 °C. Controls to measure the change in enthalpy (ΔH) because of sample dilution and mixing effects were made. The data obtained was analysed using the Microcal Origin 9 software [35]. The baseline was adjusted using the "Auto Baseline" and then all the peaks were integrated using the "Integrate All Peaks" routine. The data can be normalised in different ways using the "DeltaH" window, which plots the data in $kcalmol^{-1}$ of injection versus the molar ratio of the ligand (in this case dsDNA/protein). The integrated data can then be adjusted to different mathematical models, we used the in-built non-linear regression model of "One Set of Sites". The fitting parameters can be changed using the dialog box for editing functions typical in the Origin Lab® programs, as is described elsewhere [39].

Palindromic DNA	5'-GTATATTAGTCTATATGGACTAGTCCATATAGACTAATATAC-3'
Randomised DNA	5'-CATATAATCAGATATACCTGATCAGGTATATCTGATTATATG-3'

Table 1. $RdhR_{CbdbA1625}$ palindromic DNA sequence used for ITC experiments.

3.3.7 Analytical ultracentrifugation and equilibrium dialysis

$RdhR_{CbdbA1625}$ was buffer exchanged into Tris-HCl 10 mM pH 7.4, NaCl 300 mM using gel-filtration (S-200 10/30 GE). Sedimentation velocity was carried out using a XL-A model centrifuge at 45000 RPM at 20 °C in 450 µL double sector cells. The sedimentation boundary was monitored every 90 seconds using two different wavelengths of 280 nm or 230 nm for a total of 200 scans. The analysis of protein and DNA interactions was first performed using between 8 to 20 µM protein concentrations and with varying ratios of DNA/protein from 10:1 to 2:1 in total. The data was interpreted with a model-based distribution of the Lamm equation analytical solutions $C_{(s)}$ using the software Sedfit [40]. The Lamm equation is a partial differential equation that describes the evolution of macromolecular concentration distribution in a sector-shaped solution column under the influence of a centrifugal field. For an ideal macromolecule the concentration profile *(χ)* is a function of the distance from the center of rotation *(r)*, and time *(t)*, following the **Equation 2**, where the sedimentation coefficient is *s* and the diffusion coefficient *D* in a sector-shaped volume and *w* the angular velocity of the rotor [41].

$$\frac{\partial \chi}{\partial t} = \frac{1}{r}\frac{\partial}{\partial r}\left[rD\frac{\partial \chi}{\partial r} - s\omega^2\gamma^2\chi\right]$$

Equation 2. Lamm differential equation.

Apparent sedimentation coefficients were obtained by integration of the peak and the hydrodynamic radius and frictional ratios (f/f_o) for the sedimenting dimer were calculated in the program Sednterp [42] using the mass obtained from MALLS. Sedimentation equilibrium experiments were performed in 6-sector cells using different protein concentrations ranging from 50 µM, 30 µM, 20 µM, 10 µM, 2.5 µM and 1 µM. The samples were centrifuged at 10, 15, and 20 krpm with monitoring at both 280 nm and 230 nm at 4-hour intervals until the last two scans overlapped when equilibrium was reached approximately after 16 hours. The data was analysed using the program HeteroAnalysis [43] using either a single species approximation or a monomer-dimer with a fixed monomer mass of 19 330 Da (buoyant mass of 5262 Da). HeteroAnalysis is widely used for analysing

sedimentation equilibrium data from AUC experiments. The program performs global analysis of multiple phenomena, including complex association, by implementing global non-linear least-squares fitting of the sedimentation equilibria to pre-programmed or inbuilt model functions [44].

3.3.8 Crystallisation procedures, data collection and structure determination by X-ray crystallography

Preparations of ~10 mgmL^{-1} of the purified RdhR$_{CbdbA1625}$ (both His and Strep constructs) in Tris-HCl in 25 mM, pH7.5, NaCl 500 mM were used for crystallisation screening. A sitting-drop vapour diffusion method was used and 300 nL protein solution plus 300 nL mother liquor drops were equilibrated against 30 µL mother liquor in a 96-well format using a high-throughput liquid-handling robot (Mosquito MD11-11, Molecular Dimensions). Crystals grew in multiple conditions from the Morpheus and SG1 commercial screen (Molecular Dimensions) within 2 days of incubation at 4 °C. Optimisation from the screen SG1 (Molecular Dimensions) using the robot Dragonfly (TTP LabTech) was performed using freshly prepared solutions of Tris-HCl 100 mM pH 7.5, sodium acetate 300 mM, and PEG-8000 15% w/v. Crystals were flash-cooled in liquid nitrogen following cryo-protection in mother liquor supplemented with PEG-200 10%. For soaking crystals with heavy atoms, 1 mM stock solution of potassium tetrachloroplatinate (II) K_2PtCl_4 or mercury (II) acetate $Hg(OAc)_2$ (Sigma) was added into a drop of mother liquor containing the crystals to a final concentration of ~100 µM.

Crystals were incubated for 5 min before flash-cooling. A similar procedure was performed to soak RdhR$_{CbdbA1625}$ crystals with various chloroaromatic ligands. Data was collected at Diamond beamlines at 100 K and integrated and scaled using XDS. The RdhR$_{CbdbA1625}$ structure was solved by Multiple Isomorphous Replacement with Anomalous Scattering (MIRAS) using a Pt and Hg derivative. Heavy atom sites and initial phases were calculated using Autorickshaw [45], and the initial model building completed using ARP/wARP [46]. The 2,4-DCP, 3,4-DCP, 2,3-DCP, 2,3,4-TCP bound RdhR$_{CbdbA1625}$ structures were solved by molecular replacement using Phaser MR [47] and refined using Phenix [48] and CCP4 [49] by difference Fourier methods along with real-space refinement in COOT [50]. The crystal structures images were generated using Chimera [51] and QtMG (CCP4MG) [52]. The data collection tables can be found in **Supplementary Material, Figure 8**.

3.4. Results

3.4.1 Interaction of $RdhR_{CbdbA1625}$ with organohalide ligands

To identify possible physiological effectors for the $RdhR_{CbdbA1625}$ we performed intrinsic fluorescence quenching experiments to assess the binding of several candidate organohalide compounds. In proteins, fluorescence is an intrinsic property of aromatic amino acid residues and can offer information on ligand binding and conformational changes due to variations in electronic microenvironments within the protein. Fluorescence quenching consists in the reduction of the fluorescence intensity due to multiple phenomena, such as complex-formation in the ground-state, charge-transfer processes, and collisional quenching. Collisional or dynamic quenching occurs when the fluorophore intermittently encounters and collides with a quencher molecule. Static quenching in contrast involves the formation of a complex between the fluorophore and the quencher [36]. A combination of these two types of quenching is often a complicating factor during experiments.[2] The $RdhR_{CbdbA1625}$ transcription factor contains Trp 14 and Trp 28 per monomer, which a fluorescence maximum at 333 nm.

As was demonstrated before by Wagner *et al* and Krasper *et al* [8, 9], the expression of the $RdhA_{CbdbA1624}$ from *Dehalococcoides mccartyi* strain CBDB1 is induced in the presence of 1,2,4-trichlorobenzene. Unfortunately, due to its low solubility in aqueous solutions, the titration of the protein $RdhR_{CbdbA1625}$ with these compounds could not be achieved. We opted for more soluble chlorophenol compounds as alternatives and observed quenching of the Trp (or intrinsic) fluorescence in response to the binding of dichlorophenolic ligands, particularly the 2,4-DCP and 3,4-DCP. The DCPs titration data was fitted to a non-linear fitting function [35]. We performed titrations with the regioisomers 2,3,4-TCP, 2,3,5-TCP, and 3,4,5-TCP, and derived approximate K_D values for the three ligands following the analysing of the intrinsic fluorescence quenching after titration of the $RdhR_{CbdbA1625}$ sample (0.25 μM) with each regioisomer. The area under the curve in each spectral dataset was normalised and fitted to a quadratic velocity model (**Equation 1**. Morrison equation for tight-binding ligands). Inner filter effects and dilution effects were dismissed as the reason for the Trp fluorescence quenching and good correlation values (R^2 = 0.98-0.99) were obtained after the fitting, indicating that the functions describe the binding accurately. The mixed contribution of dynamic quenching and static quenching of the fluorescence is

[2] Apparent quenching also occurs because of the optical properties of the sample, such as turbidity.

probably one of the reasons of the sloping or deviations observed **(Supplementary Material, Figure 3)**. The K_D values are in the μM range and comparable to those reported for other MarR homologous upon binding aromatic molecules, like salicylates [53]. According to these binding experiments, the relative position of the phenolate hydroxyl group does not affect the affinity, given that the real effector probably is benzene. The numerical values of the binding constants show that the transcription factor $RdhR_{CbdbA1625}$ binds both the dichlorinated and trichlorinated ligands under the experimental conditions used **(Figure 1 and 2)**.

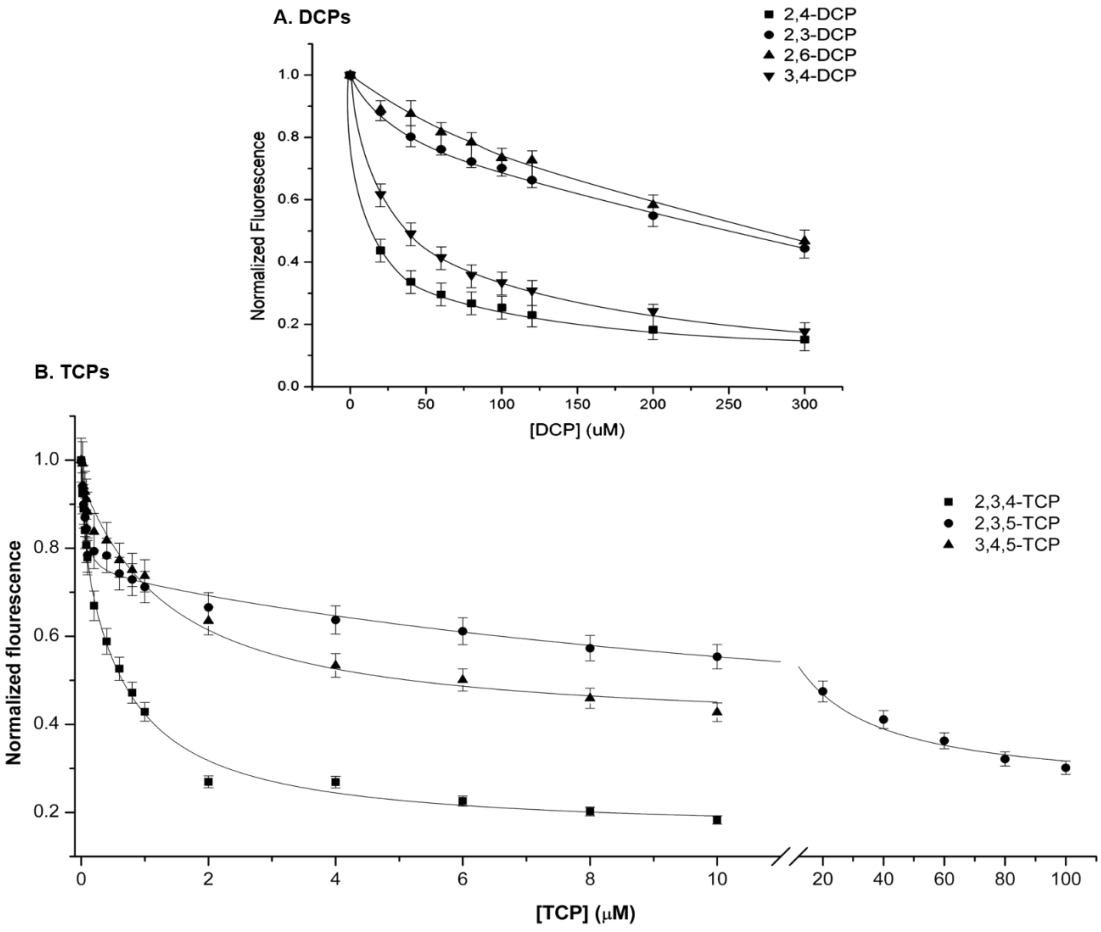


Figure 1. Ligand-binding saturation curves of the $RdhR_{CbdbA1625}$ titrated with organohalides. Trp fluorescence quenching in response to the titration of $RdhR_{CbdbA1625}$ with the dichlorophenols 2,4-DCP; 3,4-DCP; 2,3-DCP and 2,6-DCP (A) and the trichlorophenols 2,3,4-TCP, 2,3,5 and 3,4,5-TCP (B) regioisomers. In the case of the DCPs, the data was fitted to a non-linear single binding site model, while the TCPs data was fitted to a quadratic model after integrating the fluorescence emission for each data-point of the titration using OriginPro 9.1® [35]. The 2,3,5-TCP presents two different orientations, and its binding is described by two quadratic

equations. The sharpness of the inflection in each saturation curve can be attributed to differences in affinity of each tight-binding effector.

In contrast, distinct binding affinities for the various dichlorophenolic conformational isomers suggest that the relative position of the chlorine atoms is a key factor in determining the binding specificity (2,4-DCP and 3,4-DCP). Weaker binding to other dichlorophenolic compounds by $RdhR_{CbdbA1625}$ was also observed, as with the 2,6-DCP. In the case of the trichlorophenolic ligands **(Figure 1)**, it was determined that the TCP regioisomers bind with almost nanomolar affinity to $RdhR_{CbdbA1625}$, with the 2,3,4-TCP having the lowest value of K_D of 0.34 ± 0.1 μM. The 3,4,5-TCP ligand binds with similar affinity (K_D of 1.23 ± 0.2 μM), so we conclude the relative position of the hydroxyl group towards the halogen substituents does not affect the recognition of the ligand significantly. Interestingly, our data suggests that $RdhR_{CbdbA1625}$ can bind the 2,3,5-TCP ligand (that mimics the physiological ligand 1,2,4-TCB *in vivo*), in two different orientations, as two K_D values were calculated (K_{D1} of 5.2 ± 1.6 μM and K_{D2} of 15.7 ± 3.1 μM), revealing distinctive affinities for each binding mode. We summarised the binding parameters for all the organohalide ligands tested according to the binding affinity in the form of dissociation constants at μM values **(Figure 2)**.

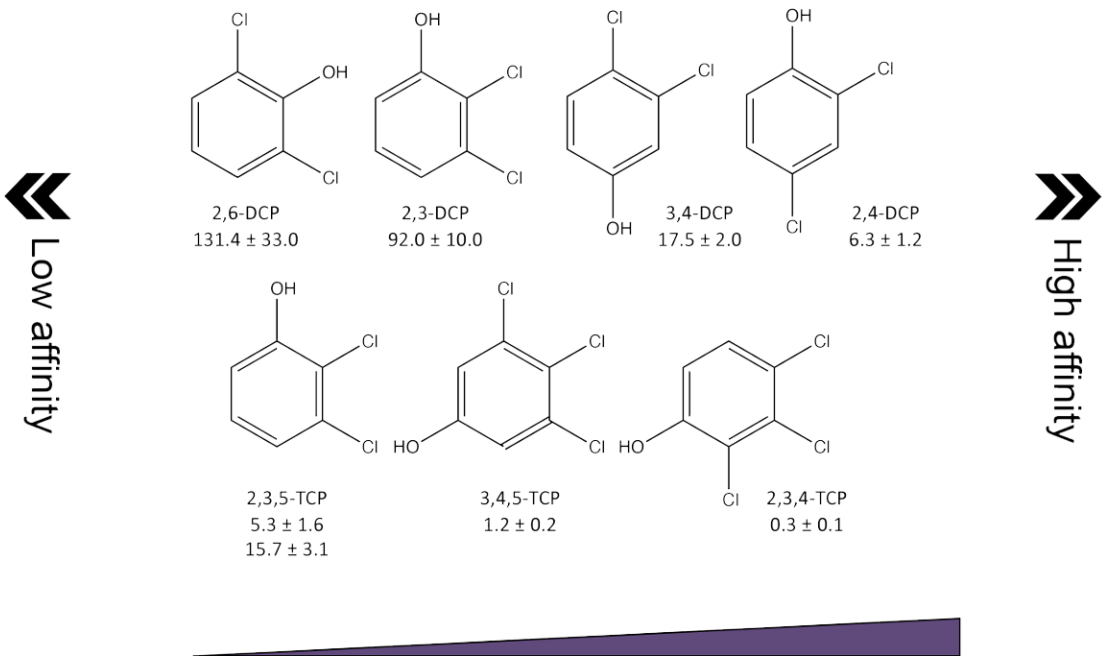


Figure 2. The chloroaromatic compounds tested for binding to $RdhR_{CbdbA1625}$ using Trp fluorescence quenching titration studies. The DCP and TCP regioisomers are shown according to the numeric K_D values (μM) obtained after fitting to non-linear binding models, ordered from low affinity to high affinity. The standard error is presented for each dissociation constant.

3.4.2 *In vitro* binding of the RdhR$_{CbdbA1625}$ to the 42 bp palindromic sequence from the *rdhA*$_{cbdbA1624}$-*rdhR*$_{CbdbA1625}$ intergenic region

The analysis of the 246 bp *rdhA*$_{cbdbA1624}$-*rdhR*$_{cbdbA1625}$ intergenic region revealed six small palindromic sites with 4-6 bp half-sites separated by 1-6 bp and a larger palindrome of 42 bp (each half-site with 21 bp) that comprises five small palindromes. These palindromic sites are localised in the putative transcriptional start sites, overlapping the -10 and -35 regions corresponding to the structural gene *rdhA*$_{cbdbA1624}$ and the transcriptional regulator gene *rdhR*$_{cbdbA1625}$. Previous work done by Smidt and Wagner [9] using PCR products of sub-fragments of the complete intergenic region showed that the binding site for RdhR$_{CbdbA1625}$ is a 107 bp region, that includes the 42 bp palindromic sequence presented in **Figure 3**.

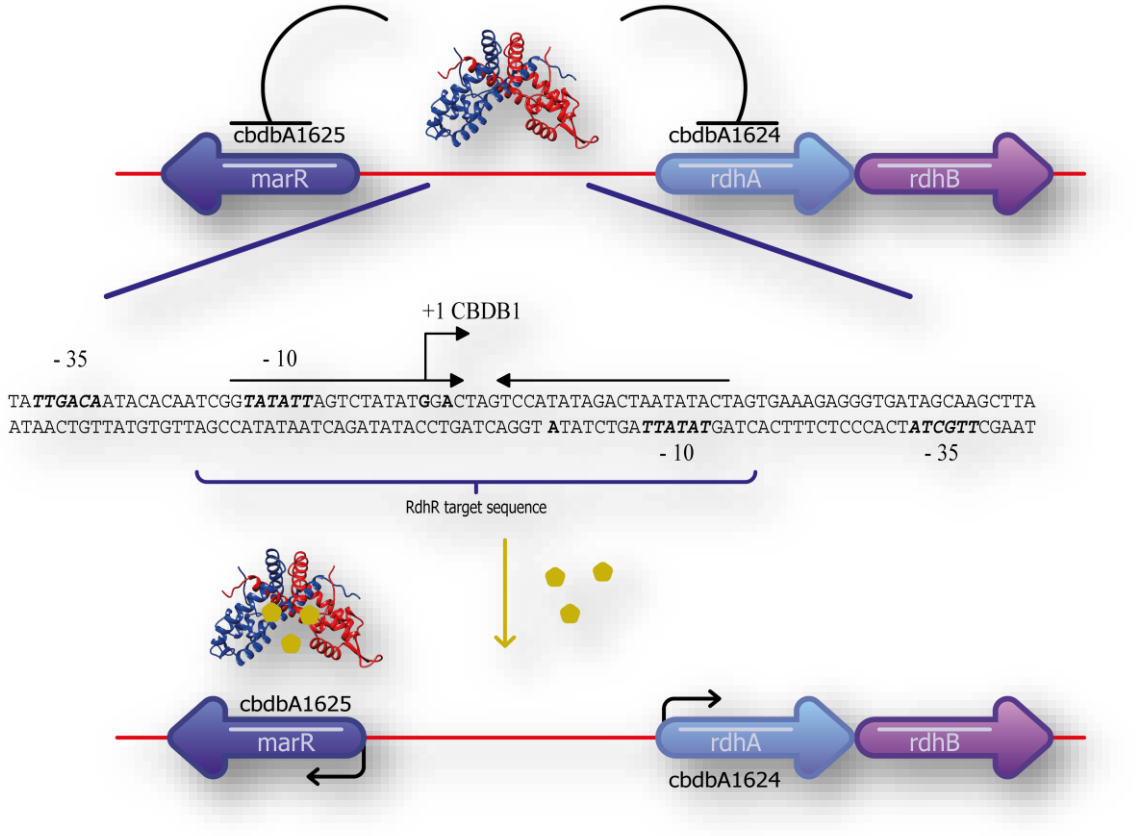


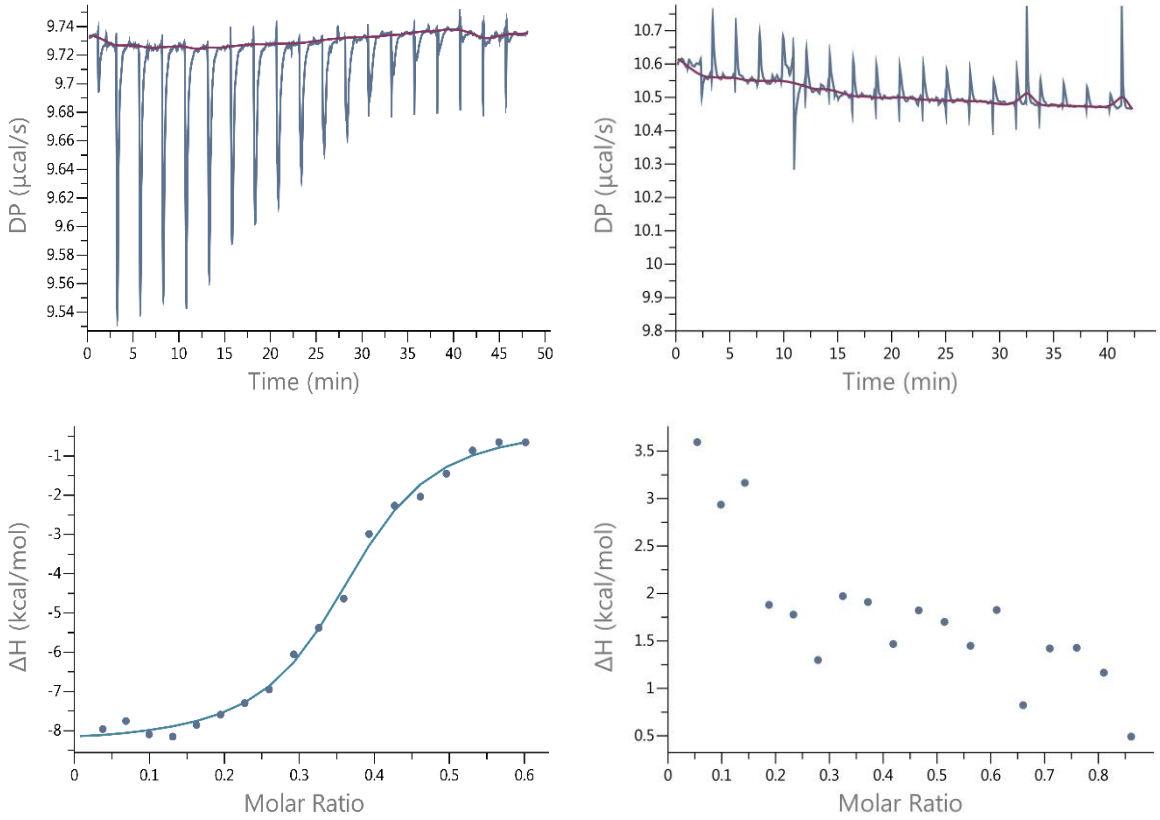
Figure 3. A model for RdhR$_{CbdbA1625}$ transcriptional regulation. Schematic representation of the palindromic sites in the *rdhA*$_{CbdbA1624}$-*rdhR*$_{CbdbA1625}$ intergenic region and the proposed physiological role of the RdhR$_{CbdbA1625}$ as repressor of the transcription. The 42 bp palindromic sequence selected for our experiments is represented inside the horizontal blue bracket. The angle arrows indicate the transcriptional start sites for each gene, including the *rdhA* gene CbdbA1624 and the *marR* (or *rdhR* gene CbdbA1625). Each half-site palindrome is shown with a horizontal arrow matching the other half-site [9]. The ligands that act as physiological effectors are represented in yellow rounded pentagons. This figure can be used as a graphical abstract.

To evaluate if some of the chloroaromatic regioisomers act as allosteric physiological effectors of the $RdhR_{CbdbA1625}$ we first performed a series of preliminary experiments to determine the binding parameters of the transcription factor to its 42 bp dsDNA palindrome (shown inside the blue bracket in **Figure 3**). Isothermal calorimetric experiments were done by titrating the $RdhR_{CbdbA1625}$ with increasing concentrations of the target sequence. Additional titrations with a randomised dsDNA of the same length and GC content as the target DNA were made as control. A K_D of 524 ± 11 nM was obtained for the target sequence, but no binding was observed in case of the random dsDNA.

The affinity for the correct sequence fragment is in the sub-micromolar range, as opposed to the random dsDNA, that shows no interaction **(Figure 4)** and this indicates that the recognition of the dsDNA by the $RdhR_{CbdbA1625}$ is sequence-specific. The K_D value obtained for the $RdhR_{CbdbA1625}$-DNA complex is comparable to those reported for other MarR-type regulators, such as the *Escherichia coli* MarR (K_D = 1 nM) and the transcription factor St1710 from *Sulfolobus tokadaii* (K_D =100 nM) [19]. The thermodynamic parameters obtained after fitting the data to a non-competitive model for one-set of equivalent binding sites using the MicroCal Tools (Origin 9.1® [35]) indicate that the interaction of the $RdhR_{CbdbA1625}$ with its cognate sequence is an exothermic process.

Our ITC data indicates an approximate stoichiometry of two dimers of $RdhR_{CbdbA1625}$ per 42 bp dsDNA molecule. The stoichiometry obtained after the ITC fitting can only make sense also if more than one binding region is present in the 42 bp palindrome, so that two $RdhR_{CbdbA1625}$ dimers interact with one dsDNA molecule, suggesting cooperative binding. The concentration-dependent equilibria between the monomeric and dimeric states of the $RdhR_{CbdbA1625}$ in solution was first observed after the purification of the sample by size exclusion chromatography **(Supplementary Information, Figure 2)** and later proven by sedimentation velocity experiments by AUC. According to the equilibrium sedimentation experiments **(Supplementary Information, Figure 4)**, the dimerisation constant for $RdhR_{CbdbA1625}$ is approximately 2.62 µM. The numerical value of this dissociation constant indicates the concentration limit at which the protein exists mainly as a dimeric species and is thus capable of interacting with the major groove of dsDNA molecule, as most winged HtH proteins do [25].

*



Non-competitive binding model	
N (sites)	0.471
K_D (M)	$524 \times 10^{-9} \pm 11 \times 10^{-9}$
ΔH(kcalmol^{-1})	-8.88 ± 0.29
ΔG(kcalmol^{-1})	-10.24
-TΔS(kcalmol^{-1})	-1.362
Offset (kcalmol^{-1})	-0.233 ± 0.19
Chi-square	4.3×10^{-2}
T= 25 °C	

Figure 4. ITC binding studies of RdhR$_{CbdbA1625}$-DNA binding. A total of 60 µM of protein were titrated with increasing concentrations of its palindromic 42 bp dsDNA (left) or a randomised dsDNA sequence of the same length and GC content (right). The thermodynamic parameters obtained for the DNA binding were obtained after fitting the data to a non-competitive model for one-set of equivalent binding sites (MicroCal Tools, Origin 9.1® [35]). A K_D of 0.524 ± 11 nM was obtained, indicating tight binding for its cognate dsDNA (left). An N value (or binding ratio) indicates an approximate stoichiometry of 2:1 molecules of protein (dimer) per dsDNA. The N<1 value might also indicate cooperative binding. The negative ΔG and ΔH values indicate that the binding process is exothermic and occurs spontaneously. No binding was observed in case of the randomised sequence (right).

Additionally, we performed fluorescence quenching experiments to determine if the binding of the 2,3,4-TCP would influence the stability of the $RdhR_{CbdbA1625}$-DNA complex by monitoring the excitation spectra of the protein. First, we titrated the protein with increasing concentrations of dsDNA until saturation was achieved. The titration curve was obtained after normalising the fluorescence using the emission maximum of the free $RdhR_{CbdbA1625}$ as reference and fitting the data to a quadratic model using OriginPro 9.1® [35] **(Figure 5. Detail)**. We obtained a K_D of a comparable magnitude (0.91 ± 0.33 µM) to that obtained by ITC. The excitation spectra were recorded at 360 nm, at dsDNA saturating conditions and after a single addition of the 2,3,4-TCP to a final concentration of 20 µM. Our data shows the formation of an excimer after the addition of dsDNA that is representative of the $RdhR_{CbdbA1625}$-dsDNA complex (with a maximum fluorescence of 300 nm). The change in the fluorescence excitation spectra when the 2,3,4-TCP is added to the $RdhR_{CbdbA1625}$-dsDNA, the quenching of the excited dimer fluorescence and gradual shift to the 280 nm, indicating the loss of the Protein/DNA/L interactions **(Figure 5)**, suggesting that the dsDNA complex was destabilised by the binding of the chlorinated ligand, and the role of the $RdhR_{CbdbA1625}$ is a transcriptional repressor.

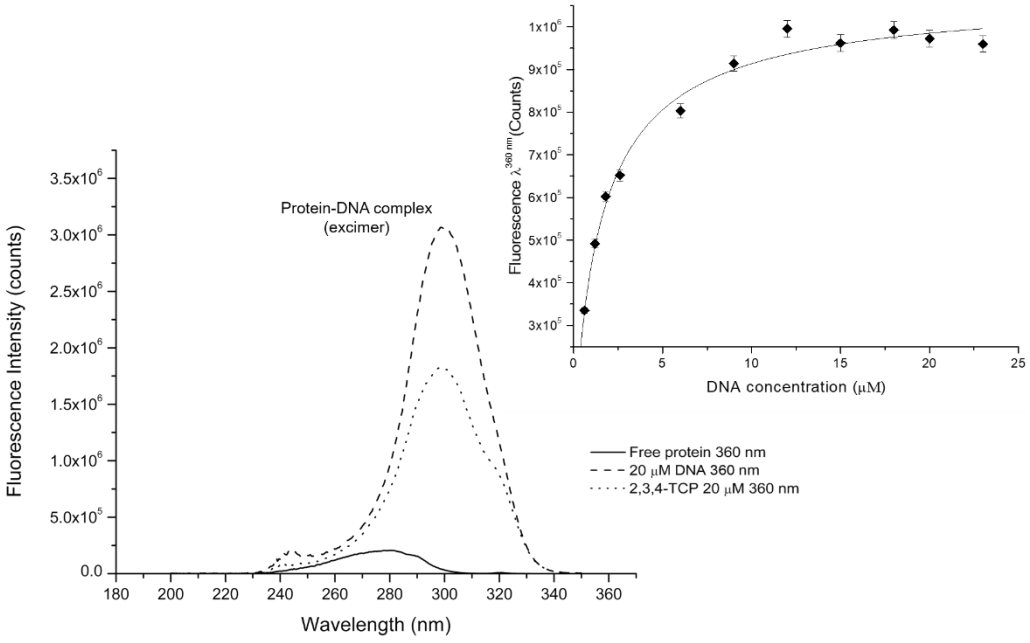


Figure 5. DNA binding experiments monitored by Trp fluorescence quenching. The data was fitted to a quadratic equation, giving a comparable value to the K_D obtained by ITC (0.91 ± 0.33 µM (R^2 = 0.97) (insert). We analysed the fluorescence excitation spectra in absence and presence of the 42 bp dsDNA at 360 nm. A stable $RdhR_{CbdbA1625}$-DNA complex was formed, as shown by the presence of an excimer. The addition of the ligand leads to the gradual disassembly of the $RdhR_{CbdbA1625}$-DNA complex (main figure).

With that in mind, we decided to further test this hypothesis by alternative means. Analytic ultracentrifugation was used to monitor $RdhR_{CbdbA1625}$ complex formation and initial studies were done at different $RdhR_{CbdbA1625}$-DNA ratios ranging from 2:1, 4:1, and 10:1, to determine the optimal concentration range for chlorinated ligand interference studies **(Figure 6)**. Our results suggest the formation of higher-order protein-DNA complexes, specifically a tetramer of $RdhR_{CbdbA1625}$ bound to the 42 bp dsDNA at approximately 6 Svedberg (S). This could be due to sequential binding of two independent protein dimers, at the same or the opposite face of the dsDNA, as has been shown before in case of other MarR homologues depending on the length of the palindrome [25, 34]. At lower ratios, the formation of a species consistent with a $RdhR_{CbdbA1625}$ dimer interacting with the dsDNA sequence was also observed. These results are in good agreement with the ITC stoichiometric information.

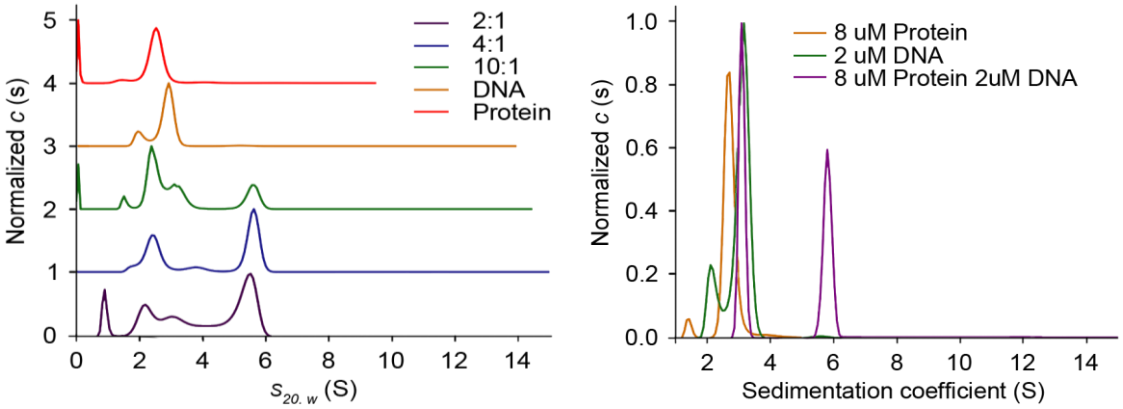


Figure 6. Analytic ultracentrifugation data showing the formation of higher-order protein-DNA complexes. The complexes form at different concentration ratios, going from 2:1 (purple, lane 0), 4:1 (blue, lane 1) and 10:1 (green, lane 2). The dsDNA fragment is shown as a single peak with a small shoulder between 2 and 3 S (yellow, lane 3), which might be the result of dynamic behaviour. The free protein is shown as a dimer at 2.4 S (red, lane 4), according to the calculated mass. The formation of a dimer of dimers or a tetramer was observed at higher concentration ratios, as a single peak at approximately 5 S (A, lanes 0,1, and 2). The tetramer formation at 6 S (in purple) was also observed during the preliminary sedimentation experiments using a 4:1 ratio, where the peaks relative to the free protein (yellow) and free DNA (green) species are also shown (B). The 42 bp palindromic sequence used contains four binding subsites that might have different binding affinities so that theoretically two dimers of the $RdhR_{CbdbA1625}$ can bound to the dsDNA palindrome used in our experiments given its length.

To study the effect of the tightest binding trichlorophenol ligand on the formation of the $RdhR_{CbdbA1625}$-DNA complex by analytical ultracentrifugation interference experiments, the higher-order protein-DNA complexes formed at a ratio of 4:1 (2 equivalents of dimeric RdhR per 1 equivalent of dsDNA) were exposed to a saturating

concentration of the 2,3,4-TCP (2 mM). The addition of the 2,3,4-TCP affected the protein-DNA interactions, as the higher-order species observed at 5 S disappeared after the TCP addition and the peak relative to the protein migrated to 2.7 S, suggesting the complete release of the dsDNA molecule (**Figure 7**). As controls we performed runs with both the protein and the dsDNA in the presence of 2,3,4-TCP.

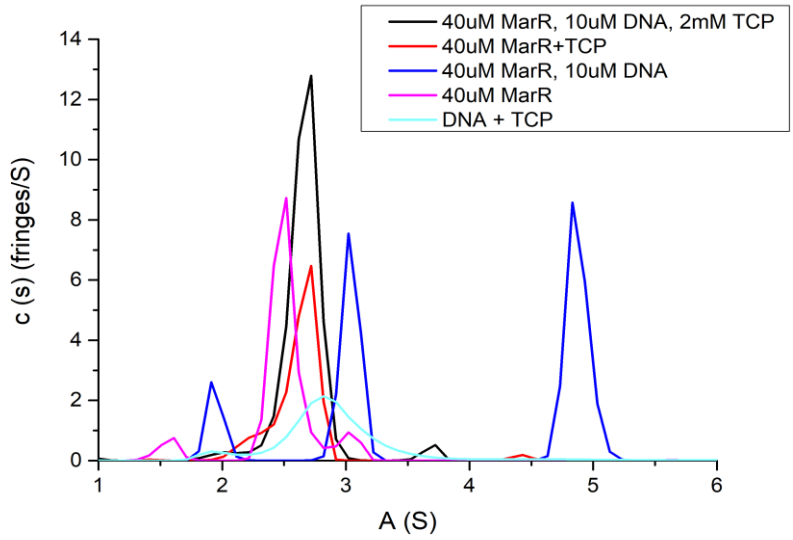


Figure 7. Analytic ultracentrifugation data suggest the physiological role of the RdhR$_{CbdbA1625}$ as a transcriptional repressor *in vitro*. AUC. The addition of the 2,3,4-TCP ligand to the tetramer formed with the 42 bp dsDNA (observed at 5 S (dark blue)) leads to the disassembly of the protein-dsDNA complex when saturating concentrations of the ligand (2 mM) are reached (as is indicated by the peak with shoulders at 2.5 S (black) that overlaps with the RdhR-ligand complex peak (red). As can be observed, this peak partially overlaps with the signals corresponding to the free RdhR (pink) and the free DNA (light blue).

Electrophoretic mobility shifts assays (EMSAs) have been used to determine interactions between proteins and nucleic acids. We performed gel-shift assays to test if the TCPs regioisomers and not only 2,3,4-TCP can act as negative effectors on the RdhR$_{CbdbA1625}$, and that in consequence, this protein may act as a repressor of the transcription of the structural *rdhA*$_{CbdbA1624}$ gene. Our results, shown in **Figure 8**, suggest that effectively the addition of the TCPs ligands and the 2,3-DCP disrupt the interaction between the RdhR$_{CbdbA1625}$-dsDNA (as there is no migration or shift). Additionally, we added phenol as a control to demonstrate that the chlorine substituents are necessary for the recognition of the ligands. When a random dsDNA sequence with the same length and GC content was used in the EMSA, no shift was observed, supporting the results obtained by ITC. The disruption of the interactions with DNA may depend on allosterism so that the binding of the effector might induce conformational changes on the recognition helices or by disrupting the oligomerisation of the RdhR$_{CbdbA1625}$, rendering the protein unable to bind dsDNA.

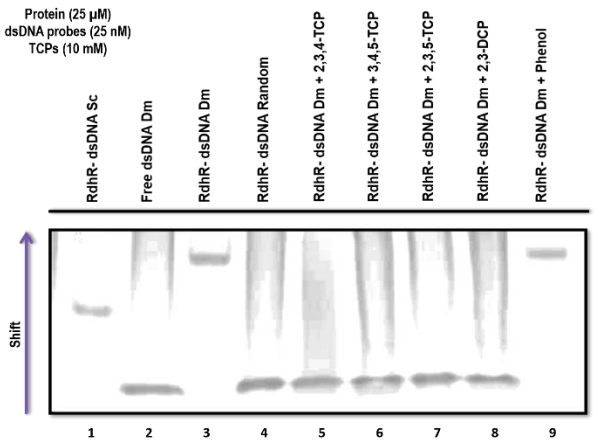


Figure 8. Electrophoretic mobility shift assays. The free dsDNA from *D. mccartyi* shows no migration during the assay (2), but when the RdhR$_{CbdbA1625}$ protein was added a shift is observed, indicating the formation of a protein-dsDNA complex (lane 3). The addition of the tight-binding ligands to the samples show the dissociation of the RdhR$_{CbdbA1625}$-dsDNA complex, as no migration was observed (lanes 5 to 8). Our results show that although the RdhR$_{CbdbA1625}$ can bind the semi-palindromic DNA sequence of the PcaV transcriptional regulator of *S. coelicolor* (lane 1), there is no indication of the formation of a complex with the randomised dsDNA sequence (lane 4). The addition of phenol does not lead to the disassembly of the complex, indicating that the chlorine substituents are necessary for the binding of the aromatic ligands (lane 9). The additional controls made for the EMSAs are presented in **Supplementary Material, Figure 5**.

PcaV is another MarR-type regulator that controls the transcription of proteins that participate in lignin degradation metabolism in response to protocatechuate, like the β-ketoadipate pathway [34]. When using the 32 bp palindromic sequence (5'-TTGACTATACTCAGTGCCCTGACTATGATACT-3') of PcaV from *Streptomyces coelicolor* kindly provided by Leopoldo Machado and Professor Neil Dixon, we proved that it also interacts with our RdhR$_{CbdbA1625}$, as shown in **Figure 8.** Interestingly, during the purification of the RdhR$_{CbdbA1625}$ we observed by UV-Visible spectroscopy that the protein eluted was bound to DNA, implying that it could also interact with some palindromic dsDNA native to *Escherichia coli* BL21 DE3 **(Supplementary Material, Figure 1)**. This suggests indirect readout and that this could play a significant role in the formation of RdhR$_{CbdbA1625}$-DNA complexes.

3.4.3 Crystal structure of the ligand-free RdhR$_{CbdbA1625}$

To provide a structural basis for the observed ligand specificity, we determined the crystal structure of the ligand-free RdhR$_{CbdbA1625}$ (His-tag 1.4 Å and Strep-tag 2.2 Å) **(Figure 9)**. The crystals of Step-tagged protein contain a pyramidal dimer of RdhR$_{CbdbA1625}$ per

asymmetric unit, as opposed to the His-tag $RdhR_{CbdbA1625}$, that contains a monomer **(Supplementary Material, Figure 6)**. The overall three-dimensional structure of $RdhR_{CbdbA1625}$ is similar to other MarR-type regulators that exist as homodimers in solution [18, 22, 27]. The secondary structure elements of the $RdhR_{CbdbA1625}$ dimer comprise 6 α-helices and 3 β-strands per monomer, arranged as a winged-helix DNA-binding domain: H1 (α2)-S1 (β1)-H2 (α3)-H3 (α4)-S2 (β2)-W1 (loop)-S3 (β3), where the winged region consists of small β sheets connected by a small loop of amino acid residues. These motifs are common for the winged helix-turn-helix DNA binding proteins including the MarR-superfamily of transcriptional regulators [23, 25]. The α1, α5, and α6 helices (located at the N and C terminal regions) constitute the dimerisation domain that determines the association between monomers via hydrophobic interactions, while the α3 and α4 helices form HTH motif, in which the DNA-recognition helix is present **(Figure 9)**.

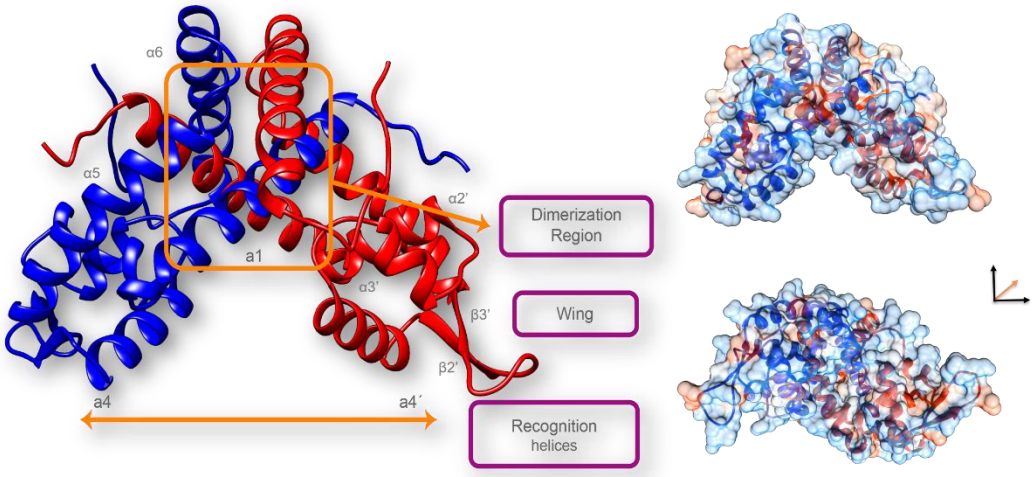


Figure 9. Crystal structure of $RdhR_{CbdbA1625}$. Structure of the ligand-free $RdhR_{CbdbA1625}$ solved at 2.2 Å, that crystallised as a dimer in the asymmetric unit, with each monomer related by a two-fold symmetry. Each monomer is shown in different colours. The space group was $P2_12_12_1$. The N-terminal Strep-tagged protein possess the typical pyramidal MarR-fold. The position of the ligand-binding pocket is located near the dimerisation region. Details of the secondary structure elements are shown in the figure, as the recognition helices α4 located in the winged helix-turn-helix domains that mediate its interaction with dsDNA. Structures were visualised in Chimera [51].

SlyA is a response regulator of virulence in *Salmonella enterica,* similar to the MarR superfamily that is known to act both as a repressor and activator of the transcription. Its crystal structure in the presence of salicylate and DNA was reported by Wu *et al* [31]. We generated a sequence alignment using CLUSTALW [54] and ESPript 3.0 [55] including the $RdhR_{CbdbA1625}$ and other homologous proteins phylogenetically related to the MarR/SlyA

families[23, 31], as well as other paralogous RdhR proteins found within the genome of *D. mccartyi* CBDB1. As can be observed in **Figure 10,** the similarities between the group are restricted to the α4 recognition helix, a segment of the α1 N-terminal and α6 C-terminal helices and the winged region, but low identity is observed on the rest of the positions of the alignment.

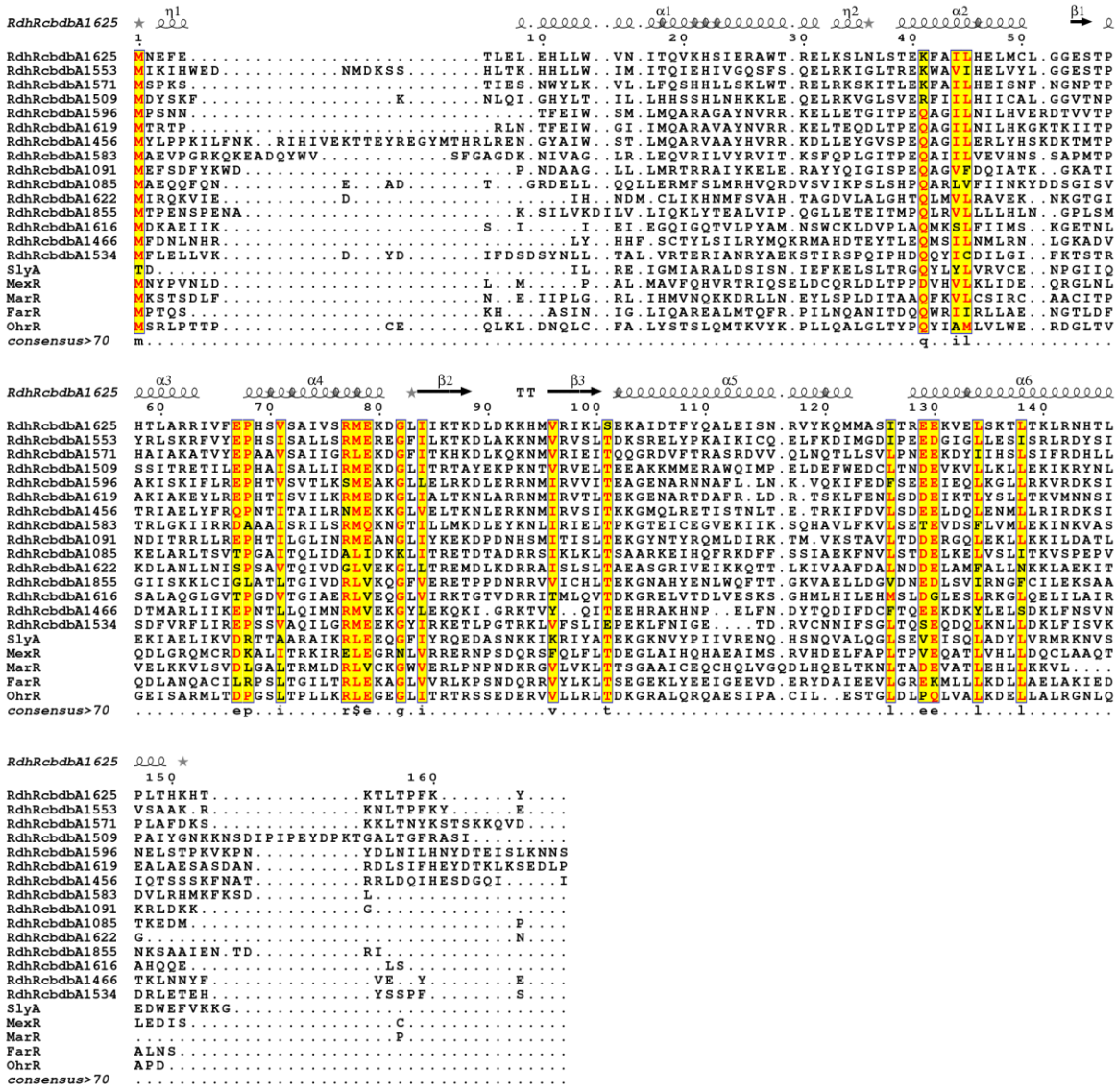


Figure 10. ClustalW sequence alignment of RdhR$_{CbdbA1625}$ with other paralogous MarR-type regulators in *Dehalococcoides mccartyi* strain CBDB1, as well as the homologous SlyA, MexR, MarR, FarR, and OhrR sequences. RdhR$_{CbdbA1625}$ secondary structure elements are shown on top of the alignment. The conserved residues are depicted in yellow colour. Black stars or dots highlight the residues that are perfectly conserved, frequently those that result relevant for the interaction with DNA bases or the chlorophenolic ligands. From the alignment, it can be observed that there is low conservation between the different regulators within the genome of *Dehalococcoides mccartyi* strain CBDB1 and with those of the orthologous to SlyA. The alignment was generated using ESPript 3.0 using the BLOSUM 62 matrix colouring scheme [55].

The structural alignment of $RdhR_{CbdbA1625}$ with SlyA and the complex SlyA-DNA (PDB ID 3Q5F) reveals that both proteins present similar folded structures and that the overall secondary structure motifs of the MarR superfamily are retained, such as the winged HtH domain, the recognition helices and the dimerisation domain **(Figure 11)**, despite the low identity and similarity values confirmed by the EMBOSS NEEDLE alignment algorithm [56] for over 140 residues **(Table 2)**. This implies that the structure and the function are retained between the MarR/SlyA family members despite evolutionary divergence. SlyA recognises AT-rich pseudo-palindromic sequences of about 12–17 bp in the promoter region of the target genes. One of the identified sequences is a 12-bp high-affinity binding site: TTAGCAAGCTAA, located downstream of the *slyA* transcriptional start site. In case of the SlyA, most of the ligand-binding induced changes occur in α4 and the winged domain [25].

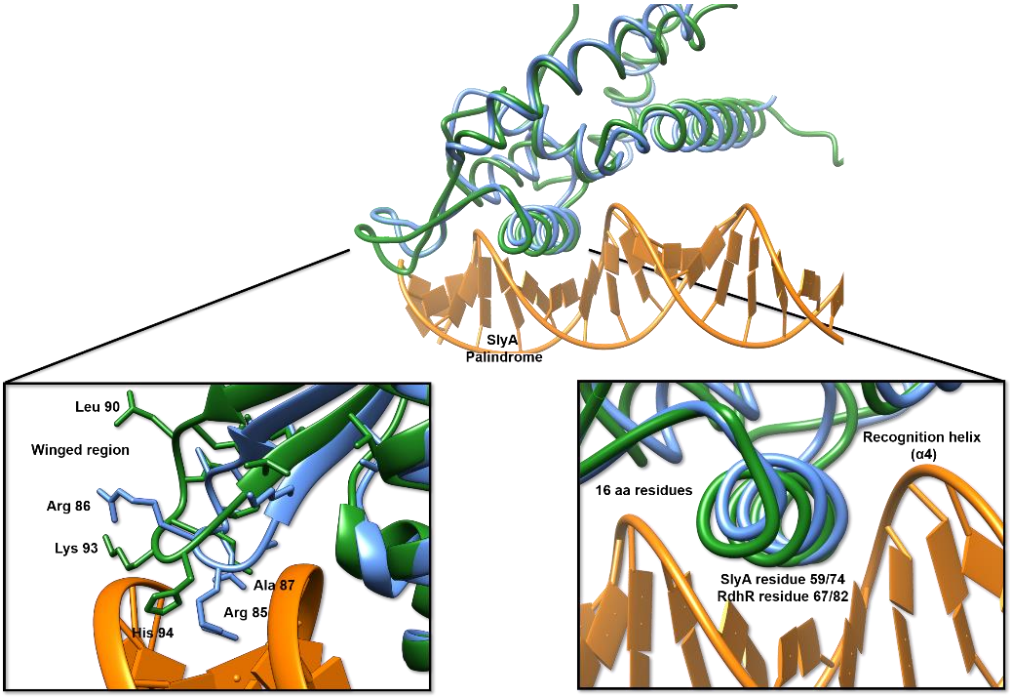


Figure 11. Comparison of $RdhR_{CbdbA1625}$ with other MarR regulators. Superposed structures of $RdhR_{CbdbA1625}$ in its ligand-free state (forest green) and the SlyA-DNA complex (where SlyA is shown in cold blue and the DNA in orange (PDP ID 3Q5F)). The DNA corresponds to 22 bp of the SlyA target sequence (ATAACTTAGCAAGCTAATTATA), that does not share complete identity with the $RdhR_{CbdbA1625}$ 42 bp palindrome from *D. mccartyi.* The winged region with the explicit side chain of the residues that participate in the interaction with dsDNA is shown in the left close-up and the recognition helixes of both SlyA and $RdhR_{CbdbA1625}$ (α4) is shown on the right one. The structural alignment for fully populated columns rendered an

RMSD of 1.736 Å, which increased when only the HtH domains were compared (RMSD 1.975 Å). Structures were aligned and visualised in Chimera ® [51].

Table 2. EMBOSS Needle alignment RdhR/SlyA		
Identity	26/154	16.95%
Similarity	49/154	31.80%
Gaps	50/154	32.50%
Length 154, Score 82, Gap penalty 10, Extended penalty 0.5, Matrix BLOSUM 62		

The α4 helices in the SlyA dimer recognise a specific DNA sequence, the direct contact with the DNA bases being provided by Arg 65 (**Figure 11**). This residue forms bidentate hydrogen bonds with guanine 14 in the pseudo-palindrome. In addition, Pro 61 and Ser 62 participate in Van der Waals contacts with the nucleotide bases. In $RdhR_{CbdbA1625}$, Ala 73 replaces Arg 65 and His 69 occurs instead of Pro 61. Only $RdhR_{CbdbA1625}$ Ser 70 and Glu 67 are conserved respectively with SlyA Ser 62 and Glu 59. The latter residue, located just outside the α4 helix, interacts with the DNA phosphate backbone in the SlyA-DNA complex. Other SlyA residues that participate in non-specific DNA contacts are not conserved in $RdhR_{CbdbA1625}$ (**Figure 11**). There is little similarity between the respective pseudo-palindromic SlyA and $RdhR_{CbdbA1625}$ target sequences, besides the presence of the TATA boxes. Nevertheless, one of the palindromes in the $RdhR_{CbdbA1625}$ target sequence (ATGGACTAGTCCAT) has similar characteristics to the SlyA high-affinity site, presenting a 5 bp segment in each half-site separated by 2 bp.

3.4.4 Fragment-merging approach uncovers high-affinity for 1,2,3-TCB compounds

By soaking the $RdhR_{CbdbA1625}$ crystals with DCP regioisomers we obtained ligand-bound structures that reveal a single organohalide binding site per monomer of protein, located at the dimer interface, as opposed to other MarR-type proteins that exhibit multiple superficial binding sites [23, 26]. All share a common binding site for one chlorine atom, bound near the Ile 64, Val 65, and Phe 66 residues (**Figure 12**). Two additional chlorine-binding sites were found adjacent to the first binding pocket, based on the information obtained from the 2,4-DCP, 2,3-DCP and 3,4-DCP ligands. The relative position of the chlorine-binding sites implies that the chlorine binding to each pocket is not mutually exclusive. A fragment-merging approach done by overlapping the DCP ligands correctly suggests that the $RdhR_{CbdbA162}$ can bind 1,2,3-TCB regioisomers with high affinity, while it

explains the relatively weak binding observed for the 2,5-DCP, 2,6-DCP, and 3,5-DCP ligands, as was demonstrated before by fluorescence spectroscopy.

The 2,4-DCP is observed to bind in two conformations related by a 180° rotation, where the hydroxyl moiety can bind in two different positions, a fact that suggest that the phenolate group is not key to the $RdhR_{CbdbA1625}$-ligand interaction. In the case of the 3,4-DCP complex structure, the ligand is observed in a single conformation. In this case, the DCP hydroxyl group is positioned close to A His 11 and B His 46, similar to one of the conformations observed for 2,4-DCP. The presence of chlorine in position 3 induces multiple rotations of the A Trp 14 and A Thr 18, some of which can establish hydrogen bonds with the ligand. In the 3,4-DCP complex structure, the ligand is observed in one conformation. The hydroxyl moiety is close to A His 11 and A His 46 B. The chlorines are positioned similarly to the 2,3-DCP. Compared to the other DCP ligands, only a few direct polar interactions between the 2,3-DCP ligand and the protein can be observed, especially with the hydroxyl group and none with the chlorine substituents, likely contributing to the relatively weak binding affinity for 2,3-DCP, when compared to 2,4-DCP and 3,4-DCP **(Supplementary Material, Figure 7)**.

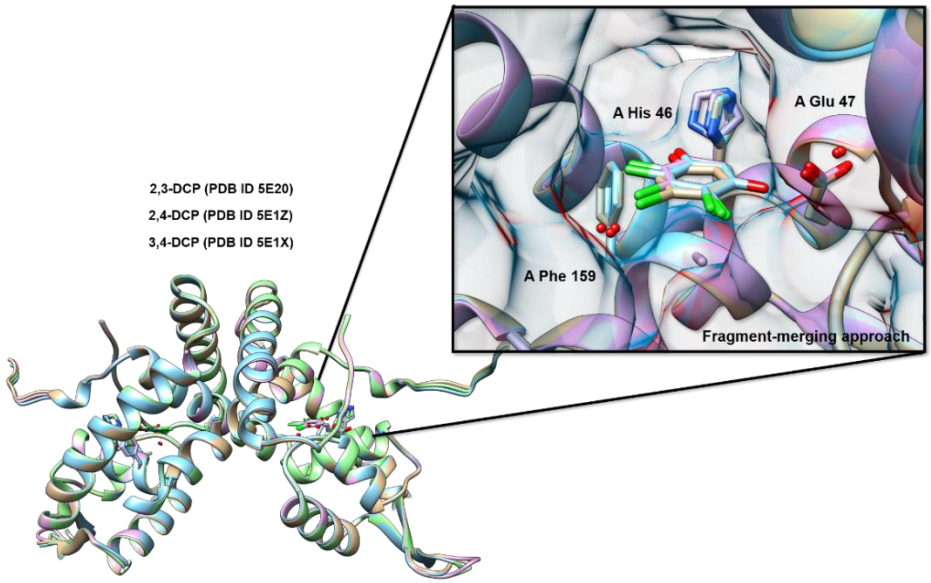


Figure 12. Detailed view of the $RdhR_{CbdbA1625}$-DCP binding site. An overlay of the 2,4-DCP bound $RdhR_{CbdbA1625}$, 3,4-DCP bound $RdhR_{CbdbA1625}$, and the 2,3-DCP bound $RdhR_{CbdbA1625}$ crystal structures are shown. In the inset, the spatial overlap of all the DCP regioisomers suggests that certain TCPs isomers are tight-binding. Structures were aligned and visualised in Chimera ® [51].

3.4.5 Crystal structure of the RdhR$_{CbdbA1625}$ with 2,3,4-TCP

The TCP-bound structure revealed a single binding site per monomer of Strep-tag RdhR$_{CbdbA1625}$ too and no significant over-all conformational changes were observed with respect to the ligand-free structures **(Figure 13)**. Interestingly, both RdhR$_{CbdbA1625}$ Trp residues are located nearby of the ligand-binding sites, consistent with our Trp quenching fluorescence experiments to the presence of DCP and TCP regioisomers. The 2,3,4-TCPs ligand is positioned in the same hydrophobic pocket formed at the dimer interface lined with residues A Trp 14, A Val 15 from α1, B Phe 42, and B Ala 43 from α2, B Ile 64 and B Phe 66 **(Supplementary Material, Figure 8)**.

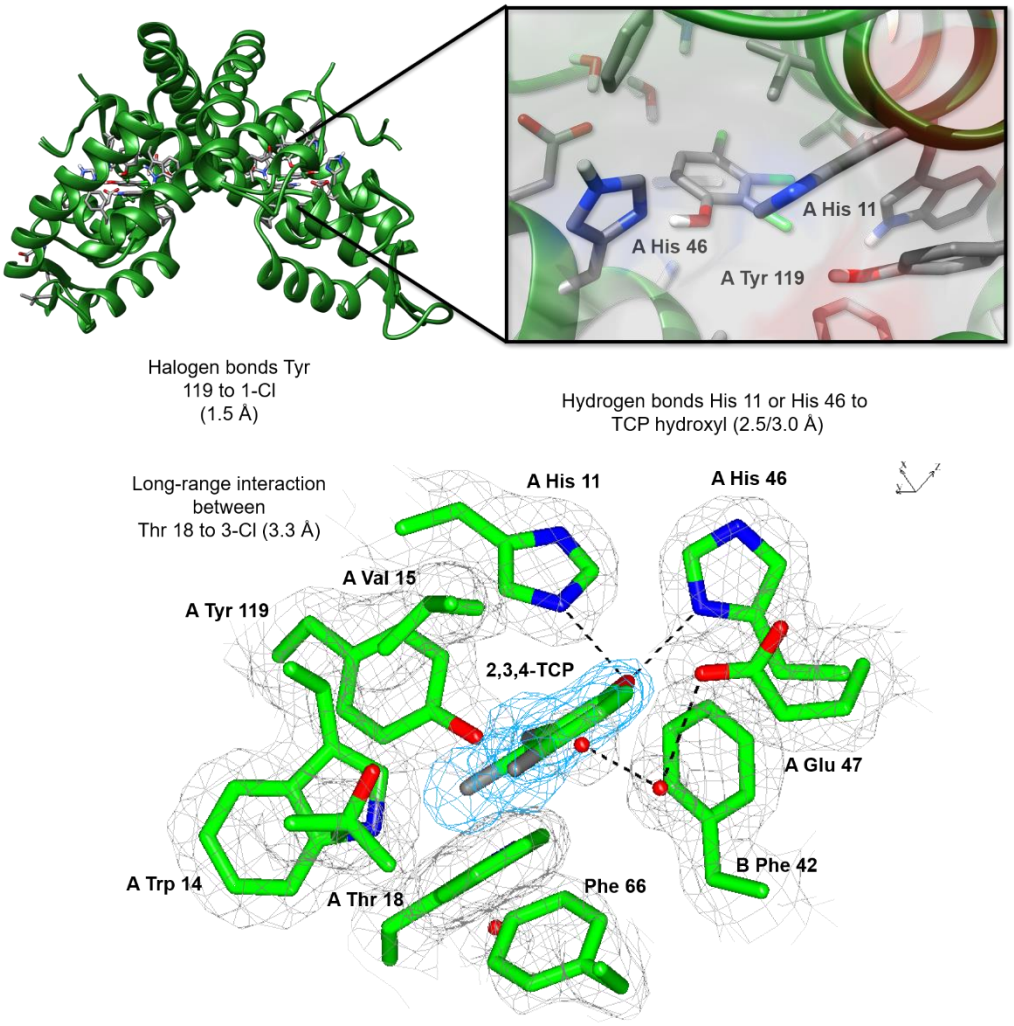


Figure 13. RdhR$_{CbdbA1625}$ binds 2,3,4-TCP with high affinity. A miniature overlay shows the ligand pocket surface. The $2F_O$-F_C electron density map is shown at 1 σ around hey aa residues, the ligand (blue), and the H_2O molecules that are close enough to support H-bonding, halogen bonding, and hydrophobic interactions. The crystals were structurally aligned and visualised in Chimera ® [51] and QtMG ® [52].

The 2,3,4-DCP also bind in two conformations as the 2,4-DCP, leading to different positions of the hydroxyl group proving once more that this substituent is not necessary for the interaction to the $RdhR_{CbdbA1625}$. The three chlorine atoms were found in the same binding subsites observed in case of the DCP ligands when overlapped, stabilised by hydrogen bonds with nearby His residues A His 11, B His 46, and B Glu 47, chlorine-hydrogen halogen bonds with A Tyr 119 (a distance of 2.9 Å), and hydrophobic interactions with A Trp 14, A Val 15 from α1, B Phe 42, and B Ala 43 from α2, B Ile 64, and B Phe 66, as shown in **Figure 13**. Chlorine-hydrogen bonds have been previously reported, even if organic chlorines are not considered to be good acceptors. These hard-donor/soft-acceptor types of interactions might be important for stabilising the ligand and could contribute to the difference in affinity of $RdhR_{CbdbA1625}$ for various TCP regioisomers. While the side chains of some residues adopt distinct conformations upon ligand binding, no wide-ranging structural effects that could explain an allosteric effect on the binding of dsDNA were observed.

3.5 Discussion

Despite the recent elucidation of two reductive dehalogenases structures [57] there is still a very limited insight into the substrate specificity of the highly complex RdhAs using *in silico* approaches. The characterisation of the transcriptional regulators associated with OHR gene clusters could reveal the likely physiological effectors, and thus provide a more facile experimental route to rapid determination of the substrate scope of the corresponding RdhA enzymes.

Many research groups have measured the transcriptional response of the *rdhR* genes to the presence of different organohalides in diverse OHRB to identify the possible substrates for the enzymes under transcriptional control [9, 58-61]. *Dehalococcoides mccartyi* CBDB1 contains a wide range of tightly-regulated paralogous structural genes *rdhA* in OHR related operons, thus being able to dechlorinate different types of aromatic organohalides. This fact reflects the diversity of potential organohalide terminal electron acceptors that can be reduced during respiration [17]. It seems plausible that each paralogous protein may sense a specific organohalide compound that corresponds to the substrate specificity of the associated RdhA enzyme. The $RdhA_{CbdbA1624}$ from *Dehalococcoides mccartyi* CBDB1 was shown to be upregulated in the presence of 1,2,4-TCB *in vivo*, suggesting that the associated $RdhR_{CbdbA1625}$ regulator is responsible for 1,2,4-TCB environmental sensing [33]. In the present work, we determined the organohalide ligand

binding properties and the crystal structures of the MarR-type regulator $RdhR_{CbdbA1625}$ from *D. mccartyi* CBDB1 to establish how it achieves the observed TCB dependent transcriptional regulation.

The $RdhR_{CbdbA1625}$ crystal structure corresponds to the typical 3D architecture found in other MarR transcription factors [26, 53, 62]. It forms a homodimer with a pyramidal shape, presenting a winged helix-turn-helix (wHTH) domain and a homodimerisation domain. While $RdhR_{CbdbA1625}$ is structurally similar to SlyA from *Dehalococcoides mccartyi* CBDB1, it only shares a moderate sequence similarity (31.80%). This significant diversity at the amino acid level suggests the evolutionary adaptability to recognize different environmental signalling molecules and diverse DNA targets [63]. As observed for SlyA [32], $RdhR_{CbdbA1625}$ binds the target DNA in the absence of any ligands, a fact that suggests it may act as a repressor requiring an allosteric effector to release the DNA. The 1,2,4-TCB was considered the best candidate as a possible effector for $RdhR_{CbdbA1625}$ [58]. However, we could not study the TCB binding *in vitro* due to its negligible solubility in water and the negative effects of the addition of organic solvents on $RdhR_{CbdbA1625}$ stability. Instead, we tested a range of more soluble chlorophenol ligands as putative analogues of 1,2,4-TCB.

Intrinsic fluorescence quenching studies revealed distinct binding affinities for the various chlorophenolic ligands, suggesting that the number and relative chlorine positions determine the binding affinity. Tight-binding affinities were observed for the 1,2,3-TCP and 1,2,4-TCP regioisomers tested, in comparison to 2,4-DCP and 3,4-DCP, while other isomers with different substitution patterns displayed a significantly weaker binding. The K_D values are comparable to those reported for other MarR-type regulators [26, 53] and are similar to the K_D values obtained for the organohalide sensing CprK and its corresponding ligand [19, 64].

Addition of the tightest-binding TCP ligand perturbed the $RdhR_{CbdbA1625}$-DNA complex *in vitro*, as established during interference studies by AUC and supported by fluorescence excitation experiments and EMSAs. Unfortunately, the molecular details of the repression mechanism or allosteric effects remain unknown, as a $RdhR_{CbdbA1625}$-DNA complex structure could not be obtained. The lack of organohalide ligand-binding induced structural changes in the crystal structures may reflect that the $RdhR_{CbdbA1625}$ ligand-free crystal structure is already in the low-affinity DNA-binding conformation.

A fragment-merging analysis suggested that the organohalide ligand-binding pocket should have a high affinity for ligands with 1,2,3-chlorosubstitution, while also explaining the relatively weak binding of the 2,5-DCP, 2,6-DCP, and 3,5-DCP conformers, as was proved by our fluorescence quenching experiments where nanomolar affinities were observed for the trichlorophenols. Interestingly, *D. mccartyi* CBDB1 can fully reduce the 2,3,4-TCP to 2,4-DCP, and 2,3-DCP *in vivo*, and then, the dichlorophenol products are partially transformed into monochlorophenols during respiration. On the other hand, the 2,3,5-TCP is transformed to 3,5-DCP, which is not further reduced [17].

In the $RdhR_{CbdbA1625}$ structures, the phenolate hydroxyl group of the various ligands is positioned within hydrogen-bonding distance either by the A His 11/A His 46 or to A Glu 47, but the lack of a specific hydroxyl binding pocket suggests that the physiological effectors of the $RdhR_{CbdbA1625}$ are not necessarily phenols, unlike the CprK regulators strict specificity for halophenolic substrates [19, 64-67]. This might also explain the observation that the $RdhA_{CbdbA1624}$ transcription was up-regulated in the presence of 1,2,4-TCB [58], even when a hydrogen atom occupies the third substitution position in the aromatic ring. From the fragment merging analysis, it was also speculated that $RdhR_{CbdbA1625}$ binding site could accommodate an additional chlorine atom instead of a hydroxyl moiety (or indeed a different substituent), suggesting the possibility of $RdhR_{CbdbA1625}$ binding 1,2,3,4-tetrachloroaromatic ligands too. Even though significant variation occurs in the amino acid residues implicated in the stabilisation of chlorine atoms in various *rdhR* genes from *Dehalococcoides mccartyi* CBDB1, our crystal structure in presence of the 2,3,4-TCP shows that the three chlorine atoms were found in the same binding subsites observed in case of the DCP ligands when overlapped, stabilised by hydrogen bonds, hydrophobic interactions and even a chlorine bond, a hard-donor/soft-acceptor interaction that might determine the differences in the affinity for each regioisomer. We speculate that each RdhR has a ligand-binding site that evolved to achieve high-affinity binding of organohalide ligands with specific substitution patterns over the aromatic ring.

Our data provide a template for future modelling of various RdhR orthologous proteins or other MarR-type regulators. This research reveals additional means to study the biochemistry of the MarR-type regulators and the type of ligand-sensing mechanisms that control the differential transcription and expression of the various enzymes involved in dehalogenase respiration. We also hope that our results would contribute towards the development of biosensors, specifically designed to detect the presence of halogenated organic compounds in nanomolar concentrations, on sites where industrial or agricultural

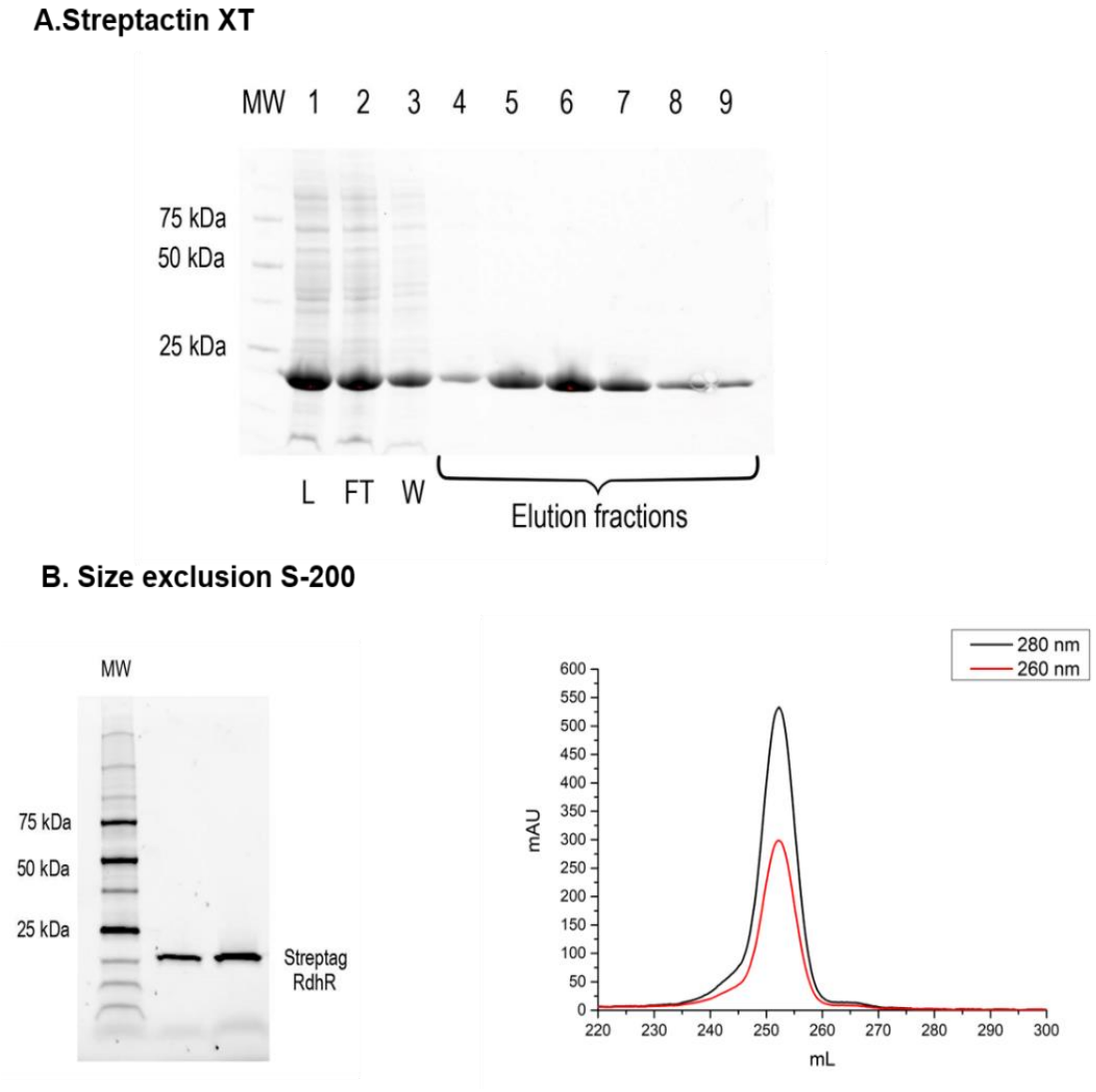
contamination is suspected, for example, in underground or superficial aquifers and river ecosystems, where later bioremediation strategies based in OHRBs cultures might take place.

The authors declared no conflict of interest. Samantha Gaytán and Carolina P. Quezada contributed equally to this publication. Mark Dunstan and Carolina P. Quezada cloned, expressed, and purified the His-tag-RdhR$_{CbdbA1625}$ and obtained the first crystal structures and initial binding data. Carolina P. Quezada and Mark Dunstan developed the groundwork with the His-tag-RdhR$_{CbdbA1625}$ regarding the binding-assays (not included in this work). Samantha Gaytán cloned, express, and purified the Steptag-RdhR$_{CbdbA1625}$, performed the ligand and DNA binding assays by fluorescence and ITC, EMSAs and solved the Strep-RdhR$_{CbdbA1625}$ crystals structures. Thomas Jowitt performed AUC and ITC assays. Sam Hay assisted with the interpretation of the fluorescence quenching binding studies. Samantha Gaytán, Carolina P. Quezada, and David Leys wrote the manuscript. David Leys designed and directed the research leading to this paper.

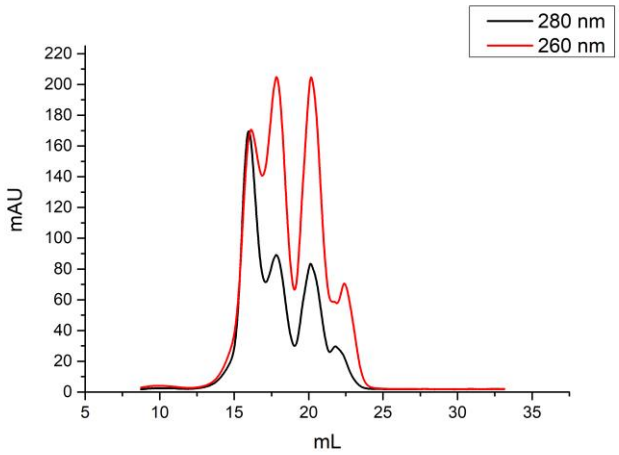
3.6 Acknowledgments

Samantha Gaytan was supported by a scholarship from CONACyT, Mexico (378154). Carolina P. Quezada was supported by FONDECYT, Chile (3170718). The atomic coordinates and structures factors (codes 5E1W, 5E1X, 5E1Z, and 5E20) have been deposited to the Protein Data Bank (http://www.pdb.org). We thank Anke Wagner and Ute Lechner, from the Martin Luther University of Halle-Wittenberg, Germany for providing the *rdhR*$_{CbdbA1625}$ gene. We also thank the Diamond Light Source for access that contributed to the results presented here (proposal numbers MX7146 and MX8997). The authors acknowledge the assistance given by the Biomolecular Analysis Core Facility and the Manchester Protein Structure Facility. Special thanks to Dr. Leopoldo Machado and Dr. Roger Kutta, for their valuable and insightful experimental recommendations and help. This work was supported by the European Research Council grant DEHALORES 206080 (to D.L.).

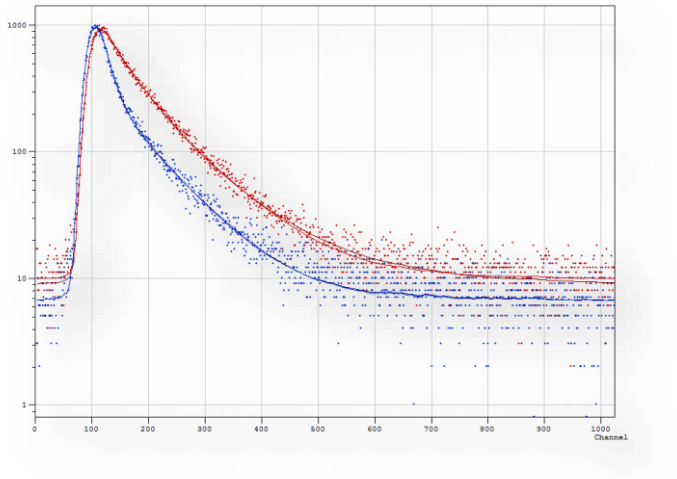
3.7 Supplementary material



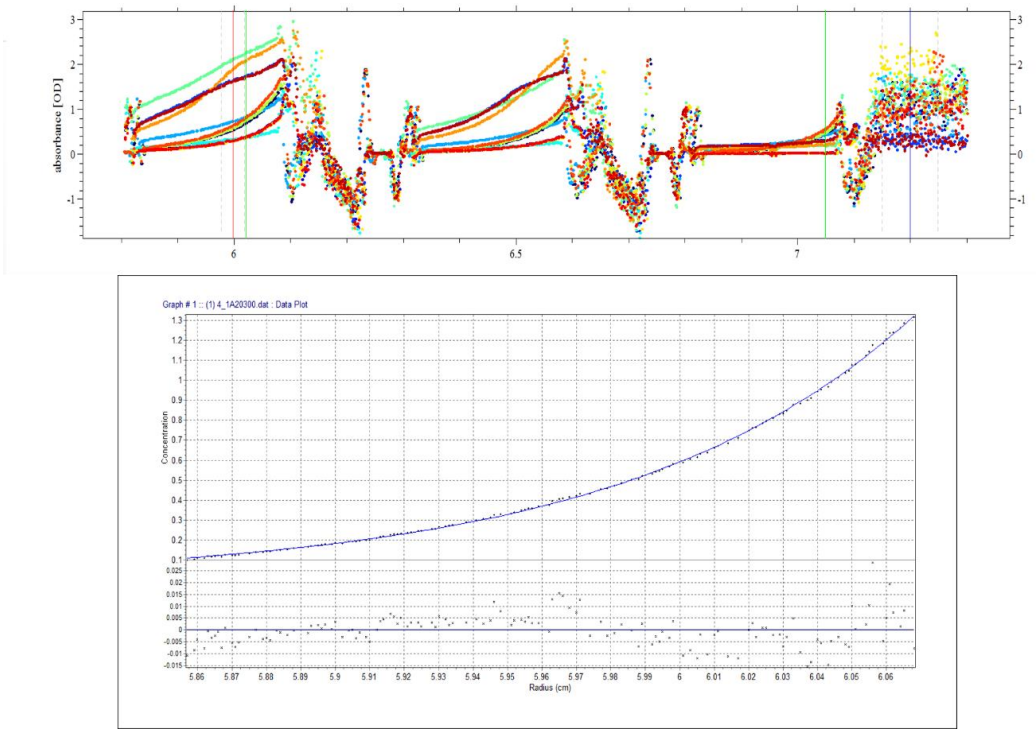
Supplementary Figure 1. **Purification of the RdhR$_{CbdbA1625}$ using a Streptactin XT resin (IBA)**. The protein of interest was eluted with biotin (2.5 mM), obtaining approximately 10 mgmL^{-1} of the pure protein (A). Fractions containing the pure Strep-RdhR$_{CbdbA1625}$ were further purified with a Hi-Trap Heparin HP (GE) using the ÅKTA Pure system (GE) to remove the dsDNA that co-purified with the RdhR$_{CbdbA1625}$. Elution was made with a linear gradient of NaCl and the RdhR$_{CbdbA1625}$ containing fractions eluted at approximately 700 mM (B and C). This suggests that RdhR$_{CbdbA1625}$ can interact with dsDNA from *E. coli* (BL21) when used as heterologous host for overexpression; we hypothesise that it is not a random double-stranded sequence and instead a palindrome that retains some degree of similarity with the RdhR$_{CbdbA1625}$ palindrome in *D. mccartyi* CBDB1 (such as the *marRAB* region).



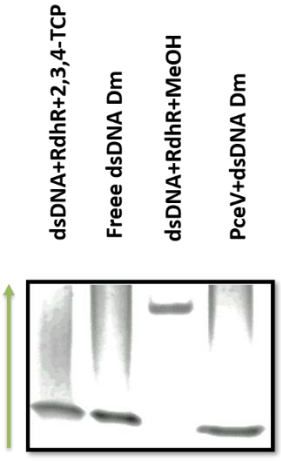
Supplementary Figure 2. Purification of the $RdhR_{CbdbA1625}$ by gel-filtration with the analytical S-200 column using ÄKTA Pure (GE). The spectra show the elution profile of the $RdhR_{CbdbA1625}$ when bound to dsDNA after the first purification step (Streptactin XT). Interestingly, the dynamic behaviour of the RdhR in solution can be observed as different oligomeric species are present.



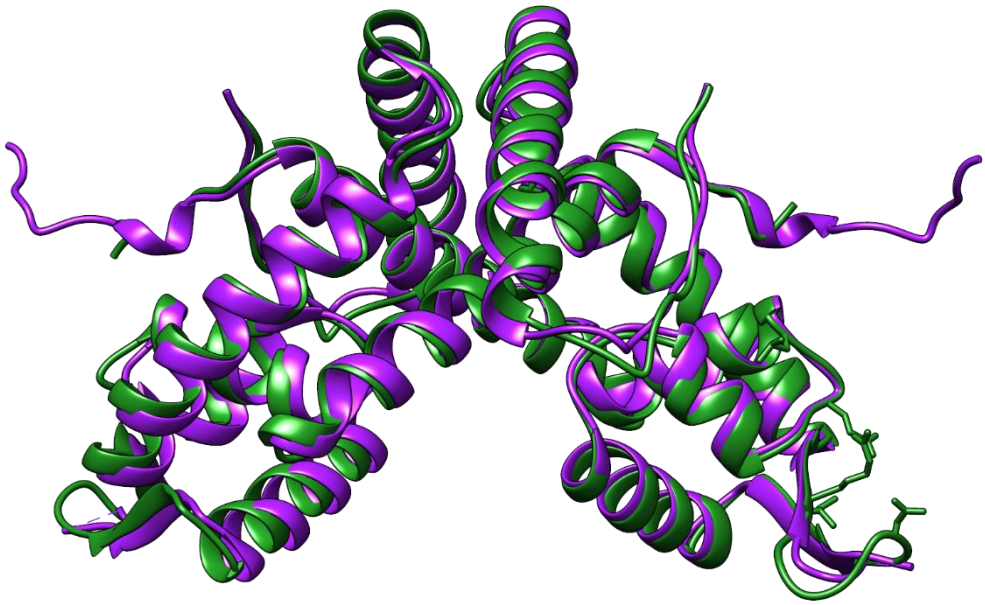
Supplementary Figure 3. Life-time decay of the free protein (red) and the RdhR-2,3,4-TCP complex. A biexponential decay suggests that quenching of Trp fluorescence occurs by more than one mechanism of de-excitation when the 2,3,4-TCP ligand is added, suggesting that the observed decrease in the fluorescence may occur both because of collisional and dynamic quenching.



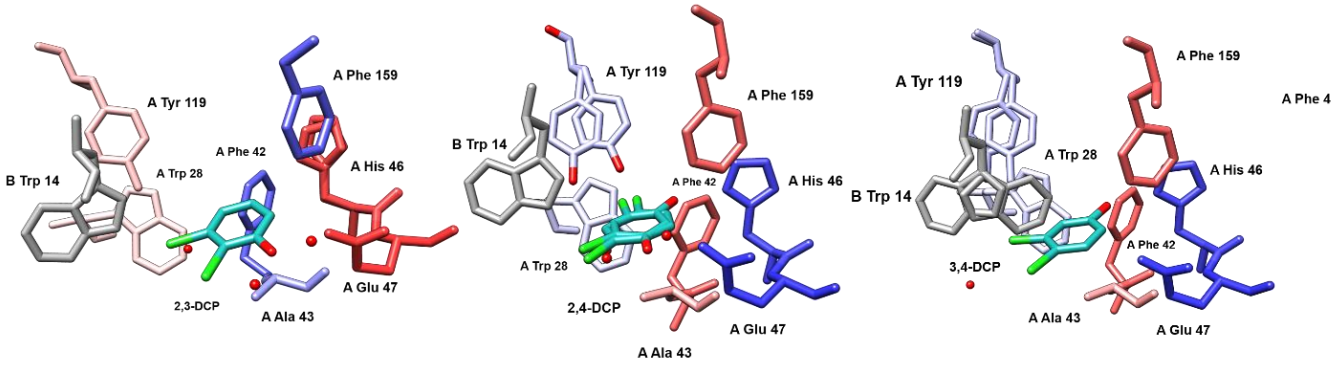
Supplementary Figure 4. **Equilibrium sedimentation experiments**. According to our experiments the dimerisation constant for $RdhR_{CbdbA1625}$ is given by the ln(k) 13.758. Indicating the concentration limit for the protein to exist mainly as a dimer in solution.



Supplementary Figure 5. **Control samples of the EMSA analysis**. The addition of MeOH to the RdhR-dsDNA complex (lane 3) does not lead to the disassembly of the protein-DNA complex, showing that the solvent used to prepare the ligand stocks does not affect the interactions between both species, as is observed after the addition of the tight-binding ligand 2,3,4-TCP (lane 1). The transcriptional regulator PcaV does not interact with the dsDNA palindromic sequence from *D. mccartyi* (lane 4), so there is no migration of the dsDNA, as shown in case of the free palindrome (lane 2).



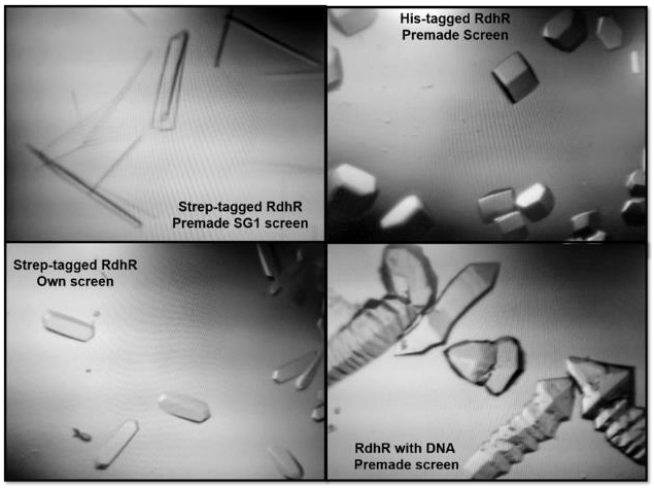
Supplementary Figure 6. Structural alignment of the His- and Strep-tagged RdhR$_{CbdbA1625}$. Structural alignment between the Strep-tag RdhR crystal, solved at 2.2 Å (forest green) and the crystal structure of the His-tag RdhR, solved at 1.6 Å (purple). The structural alignment was performed in Chimera [51], revealing an RMSD of 0.834 Å.



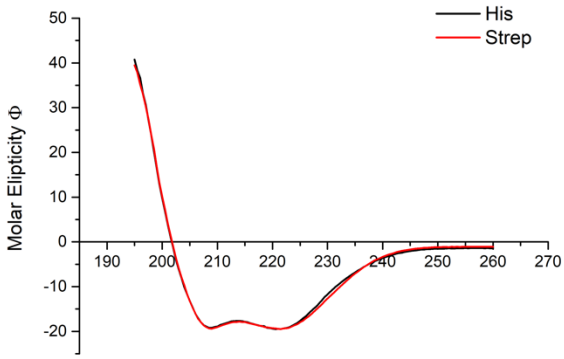
Supplementary Figure 7. Detailed view of the RdhR$_{CbdbA1625}$-DCP binding site with three different ligands. The 2,3-DCP bound RdhR$_{CbdbA1625}$, 2,4-DCP bound RdhR$_{CbdbA1625}$, and 3,4-DCP bound RdhR$_{CbdbA1625}$ are shown from left to right. Electronic density maps were omitted for clarity. The crystal structures are currently deposited in the PDB (IDs 5E20, 5E1Z, 5E1X).

	$RdhR_{CbdbA1625}$ (PDB 5E1W)	2,4-DCP-bound (PDB 5E1Z)	3,4-DCP-bound (PDB 5E1X)	2,3-DCP-bound (PDB 5E20)	2,3,4-TCP bound	RdhR free
Data collection						
Space group	$P3_121$	$P3_121$	$P3_121$	$P3_121$	$P\ 1\ 2_1\ 1$	$P\ 1\ 2_1\ 1$
Cell dimensions						
a,b,c (Å)	71.95, 71.95, 87.63	71.66, 71.66, 87.52	71.14, 71.14, 87.62	71.32, 71.32, 87.12	41.03,59.41, 90.12	61.36, 72.99, 90.55
R_{merge} (%)#	0.048 (0.666)	0.058 (0.682)	0.07 (1.074)	0.061 (0.543)	0.048 (0.666)	0.068 (0.790)
I/σI	19.4 (1.7)	23.5 (3.3)	16.8 (3.1)	24.4 (3.9)	?	?
Completeness (%)	99.8 (100)	99.99 (100)	99.87 (99.4)	99.9 (100)	90.72 (100)	89.99 (100)
Redundancy	6.7 (3.9)	9.8 (9.7)	8.2 (7.7)	9.8 (10.2)	6.7 (3.9)	9.8 (9.7)
$CC_{1/2}$					1.0(0.5)	?
Refinement						
Resolution (Å)	31.16-1.39	50.62-1.66	61.62-1.46	50.39-1.97	31.16-1.39	50.62-1.66
	(1.43-1.39)	(1.71-1.66)	(1.50-1.46)	(2.02-1.97)	(1.43-1.39)	(1.71-1.66)
No. reflections	50241 (3442)	29465 (2155)	42515 (3068)	17630 (1274)	28681 (721)	27465 (829)
R_{work}/R_{free} (%)	13.55/16.93 (19.7/23.6)	18.08/21.06	17.16/19.3 (23.4/22.3)	19.10/23.41 (24.7/26.4)	19.39(32.34)/23.75 (44.61)	21.67(24.56)/24.06
		(25.2/27.9)				22.79
Average B-factor					37.86	28.78
No. non-hydrogen atoms	1723	1664	1829	1543	2805	1883
Mean B factor (A^2)	19.195	23.884	23.537	32.702	37.86	30.88
RMS deviations						
Bond lengths (Å)	0.026	0.026	0.027	0.021	0.016	0.026
Bond angles (º)	2.35	2.43	2.77	2	1.7	2.3

Supplementary Figure 8. Crystallographic data collection and refinement statistics.



Supplementary Figure 9. **Morphology of the RdhR$_{CbdbA1625}$ crystals. The photographs were taken with a Sony Xperia XZ Premium**. Own Screen: Tris-HCl 100 mM pH 7.5, sodium acetate 300 mM, and PEG 8 000 15% w/v. Commercial or premade screen: SG1 from Molecular Dimensions ®.



Supplementary Figure 10. Circular dichroism spectra of the Strep-RdhR$_{CbdbA1625}$ and His- RdhR$_{CbdbA1625}$ (200 µM) obtained in Tris-HCl 50 mM, pH 7.5 and NaCl 300 mM (Chirascan®, Applied Photophysics). The observed spectra are typical of proteins with a high content of α-helixes as main secondary structure elements. This result implies that there are no structural changes within the RdhR$_{CbdbA1625}$ secondary motifs due to the presence of the His/Strep affinity tags.

3.8 References

1. D. Leys, L. Adrian and H. Smidt, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120316.
2. H. Smidt and W. M. d. Vos, *Annu. Rev. Microbiol.*, 2004, **58**, pp. 43-73.
3. T. Futagami, M. Goto and K. Furukawa, *Chem. Rec.*, 2008, **8**, pp. 1-12.
4. B.-E. Jugder, H. Ertan, M. Lee, M. Manefield and C. P. Marquis, *Trends Biotechnol.*, 2015, **33**, pp. 595-610.

5. X. Maymó-Gatell, Y.-t. Chien, J. M. Gossett and S. H. Zinder, *Science*, 1997, **276**, pp. 1568-1571.
6. C. Yang, A. Kublik, C. Weidauer, B. Seiwert and L. Adrian, *Environ. Sci. Technol.*, 2015, **49**, pp. 8497-8505.
7. M. Bunge, L. Adrian, A. Kraus, M. Opel, W. G. Lorenz, J. R. Andreesen, H. Görisch and U. Lechner, *Nature*, 2003, **421**, pp. 357–360.
8. L. Krasper, H. Lilie, A. Kublik, L. Adrian, R. Golbik and U. Lechner, *J. Bacteriol.*, **198**, pp. 3130-3141.
9. A. Wagner, L. Segler, S. Kleinsteuber, G. Sawers, H. Smidt and U. Lechner, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120317.
10. L. Adrian, U. Szewzyk, J. Wecke and H. Görisch, *Nature*, 2000, **408**, pp. 580–583.
11. N. Taş, M. H. A. V. Eekert, W. M. D. Vos and H. Smidt, *Microb. Biotechnol.*, 2010, **3**, pp. 389-402.
12. D. W. Major, M. L. McMaster, E. E. Cox, E. A. Edwards, S. M. Dworatzek, E. R. Hendrickson, M. G. Starr, J. A. Payne and L. W. Buonamici, *Environ. Sci. Technol.*, 2002, **36**, pp. 5106-5116.
13. W. W. Mohn and J. M. Tiedje, *Microbiol. Rev.*, 1992, **56**, pp. 482-507.
14. M. Kube, A. Beck, S. H. Zinder, H. Kuhl, R. Reinhardt and L. Adrian, *Nat. Biotechnol.*, 2005, **23**, pp. 1269–1273.
15. T. Schubert, L. Adrian, R. G. Sawers and G. Diekert, *FEMS Microbiol. Ecol.*, 2018, **94**. fiy035.
16. Y. Morita, T. Futagami, M. Goto and K. Furukawa, *Appl. Microbiol. Biotechnol.*, 2009, **83**, pp. 775–781.
17. L. Adrian, S. K. Hansen, J. M. Fung, H. Görisch and S. H. Zinder, *Environ. Sci. Technol.*, 2007, **41**, pp. 2318-2323.
18. R. E. Richardson, *Curr. Opin. Microbiol.*, 2013, **24**, pp. 498-505.
19. C. Levy, K. Pike, D. J. Heyes, M. G. Joyce, K. Gabor, H. Smidt, J. v. d. Oost and D. Leys, *Mol. Microbiol.*, 2008, **70**, pp. 151-167.
20. A. M. George and S. B. Levy, *J. Bacteriol.*, 1983, **155**, pp. 541-548.
21. S. P. Cohen, S. B. Levy, J. Foulds and J. L. Rosner, *J. Bacteriol.*, 1993, **175**, pp. 7856-7862.
22. T. Kumarevel, in *Antibiotic Resistant Bacteria - A Continuous Challenge in the New Millennium*, ed. M. Pana, Intech, London, 2012, ch. 16, pp. 403-418.
23. A. G. S.P. Wilkinson, *Curr. Issues Mol. Biol*, 2006, **8**, pp. 51–62.
24. A. G. I.C. Perera, *J. Mol. Cell Biol*, 2010, **2**, pp. 243–254.
25. A. Grove, *Curr. Biol.*, 2013, **23**, pp. R142-R143.
26. M. N. Alekshun, S. B. Levy, T. R. Mealy, B. A. Seaton and J. F. Head, *Nature*, 2001, **8**, pp. 710-714.
27. C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland Publishing, New York, USA, 2 edn., 1998.
28. Z. Hao, H. Lou, R. Zhu, J. Zhu, D. Zhang, B. S. Zhao, S. Zeng, X. Chen, J. Chan, C. He and P. R. Chen, *Nat. Chem. Biol.*, 2014, **10**, pp. 21-28.
29. R. Zhu, Z. Hao, H. Lou, Y. Song, J. Zhao, Y. Chen, J. Zhu and P. R. Chen, *J. Biol. Inorg. Chem.*, 2017, **22**, pp. 685-693.
30. E. D. K. Dolan, C. He, *J. Biol. Chem*, 2011, **286**, pp. 22178–22185.
31. R.-y. Wu, R.-g. Zhang, O. Zagnitko, I. Dementieva, N. Maltzev, J. D. Watson, R. Laskowski, P. Gornicki and A. Joachimiak, *J. Biol. Chem.*, 2003, **278**, pp. 20240-20244.
32. V. N. M. Stapleton, R. Read, J. Green, *J. Biol. Chem*, 2002, **277**, pp. 17630–17637.
33. L. S. A. Wagner, S. Kleinsteuber, G. Sawers, H. Smidt, U. Lechner *Philos. Trans. R. Soc. Lond. B. Biol. Sci*, 2013, **368**, p. 20120317.
34. J. R. Davis, B. L. Brown, R. Page and J. K. Sello, *Nucleic Acids Res.*, 2013, **41**, pp. 3888-3900.
35. OriginPro 9.1 (64-bit Sr2 b271), OriginLab Corporation, Northampton, MA, 2013
36. J. R. Lakowicz, *Principles of Fluorescence Spectroscopy* Springer, 3rd edn., 2006.

37. L. D. Ward, *Journal*, 1985, **117**, 400-414.
38. A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics*, Wiley-Blackwell, Weinheim, Germany, 2012.
39. MicroCal ITC analysis software using Origin™ User Manual, https://www.equipx.net/uploads/Malvern%20Instruments/Malvern-MicroCal-ITC-200-Software-User-Manual-English.pdf, (accessed February 2019).
40. P. Schuck, Sedfit (15.4 B), National Institutes of Health, Bethesda, Maryland, 2017
41. P. H. Brown and P. Schuck, *Comput. Phys. Commun*, 2008, **178**, pp. 105-120.
42. J. Philo, SEDNTERP (1.09), University of New Hampshire, John Philo & Thomas Laue, Durham, New Hampshire, 2019
43. J. L. Cole and J. W. Lary, HeteroAnalysis (1.1.0.57), University of Connecticut, Storrs, Connecticut, 2012
44. J. L. Cole and J. W. Lary, HeteroAnalysis: Help File, https://core.uconn.edu/files/auf/ha-help/HA-Help.htm, (accessed June 2020).
45. S. Panjikar, Auto-Rickshaw European Molecular Biology Laboratory, Hamburg, 2005
46. G. G. Langer, S. X. Cohen, V. S. Lamzin and A. Perrakis, *Nat. Protoc.*, 2008, **3**, pp. 1171-1179.
47. A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read, *J Appl Crystallogr*, 2007, **40**, pp. 658-674.
48. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L.-W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger and P. H. Zwart, *Acta Crystallogr. D*, 2010, **D66**, pp. 213-221.
49. E. Potterton, P. Briggs, M. Turkenburg and E. Dodson, *Acta Crystallogr. D*, 2003, **D59**, pp. 1131-1137.
50. P. Emsley and K. Cowtan, *Acta Crystallogr. D*, 2004, **D60**, pp. 2126-2132.
51. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, pp. 1605-1612.
52. S. McNicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, *Acta Crystallogr. D*, 2011, **D67**, pp. 386-394.
53. S. P. Wilkinson and A. Grove, *Curr. Issues Mol. Biol.*, 2006, **8**, pp. 51-62.
54. M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson and D. G. Higgins, *Bioinformatics*, 2007, **23**, pp. 2947-2948.
55. X. Robert and P. Gouet, *Nucleic Acids Res.*, 2014, **42**, pp. W320–W324.
56. F. Madeira, Y. m. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. N. Tivey, S. C. Potter, R. D. Finn and R. Lopez, *Nucleic Acids Res.*, 2019, **47**, pp. W636–W641.
57. M. Bommer, C. Kunze, J. Fesseler, T. Schubert, G. Diekert and H. Dobbek, *Science*, 2014, **346**, pp. 455-458.
58. A. Wagner, L. Adrian, S. Kleinsteuber, J. R. Andreesen and U. Lechner, *Appl Environ Microbiol.*, 2009, **75**, pp. 1876-1884.
59. A. Bisaillon, R. Beaudet, F. Lépine and R. Villemur, *Appl Environ Microbiol.*, 2011, **77**, pp. 6261-6264.
60. J. M. Fung, R. M. Morris, L. Adrian and S. H. Zinder, *Appl Environ Microbiol.*, 2007, **73**, pp. 4439-4445.
61. D. R. Johnson, E. L. Brodie, A. E. Hubbard, G. L. Andersen, S. H. Zinder and L. Alvarez-Cohen, *Appl Environ Microbiol.*, 2008, **74**, pp. 2864-2872.
62. M. I. Kim, M. U. Cho and M. Hong, *Biochem. Biophys. Res. Commun.*, 2015, **458**, pp. 644-649.
63. D. W. Ellison and V. L. Miller, *Curr. Opin. Microbiol.*, 2006, **9**, pp. 153-159.
64. M. G. Joyce, C. Levy, K. Gábor, S. M. Pop, B. D. Biehl, T. I. Doukov, J. M. Ryter, H. Mazon, H. Smidt, R. H. H. v. d. Heuvel, S. W. Ragsdale, J. v. d. Oost and D. Leys, *J. Biol. Chem.*, 2006, **281**, pp. 28318-28325.

65. L. R. Kemp, M. S. Dunstan, K. Fisher, J. Warwicker and D. Leys, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120323.
66. K. Gábor, C. S. Veríssimo, B. C. Cyran, P. t. Horst, N. P. Meijer, H. Smidt, W. M. d. Vos and J. v. d. Oost, *J. Bacteriol.*, 2006, **188**, pp. 2604–2613.
67. H. Mazon, K. Gábor, D. Leys, A. J. R. Heck, J. v. d. Oost and R. H. H. v. d. Heuvel, *J. Biol. Chem.*, 2007, **282**, pp. 11281-11290.

4.0 Chapter Four. Results Chapter II.

Heterologous expression of a self-sufficient catabolic reductive dehalogenase from marine Proteobacteria

Samantha Gaytán[1], Godwin Aleku[2], Karl Payne[1], Irina Gostimskaya[1],

Muralidharan Shanmugan[1], and David Leys *

[1] Manchester Institute of Biotechnology, Department of Chemistry, Faculty of Life Sciences,

The University of Manchester, Manchester M1 7DN, United Kingdom

[2] Department of Biochemistry, Sanger Building, University of Cambridge, 80 Tennis Court Road,

Cambridge, CB2 1GA, United Kingdom

* To whom correspondence may be addressed: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK.

Tel.: +44-161-306-5150; E-mail: david.leys@manchester.ac.uk

Keywords: Catabolic reductive dehalogenases, oxidoreductase, iron-sulfur protein, cobalamin, Proteobacteria, pesticides, NAD(P)H.

This manuscript is written as a Perspective/Communication and has not been submitted to any journal.

4.1 Abstract

Organohalide respiring bacteria (OHRBs) reduce highly substituted organohalide compounds considered major environmental pollutants during anaerobic energy conservation pathways and growth. Despite distinct phylogenetic backgrounds, all OHRBs depend on B_{12}-dependent enzymes called reductive dehalogenases (RdhAs) for respiration. In contrast, a subset of the RdhA family of enzymes features in aerobic catabolic pathways instead, some of which are fused to flavoprotein phthalate dioxygenase reductase (PDR) type domains. Here we present the heterologous production and the initial biochemical characterisation of a catabolic RdhA-PDR fusion from aerobic Proteobacteria. Both the xylose-inducible *Bacillus megaterium* and *Escherichia coli* expressing the B_{12}-uptake system (BtuB) were used to achieve its functional expression. Our results show that the RdhA-PDR fusion enzyme reduces *ortho*-halogenated phenols, while confirming the existing hypothesis that the PDR-like domain supports intramolecular transfer of electrons from NAD(P)H to the active site cobalamin, bypassing the need for an external reductase system, rendering these systems self-sufficient (ssRdhAs).

4.2 Introducing the catabolic reductive self-sufficient dehalogenases (ssRdhAs)

Microbial organohalide respiration is one of the strategies evolved by microorganisms to survive in O_2-depleted habitats; it involves a chemiosmotic gradient that couples the catabolism of organohalides to the production of ATP equivalents. Organohalide reduction has also been observed in microaerobic and aerobic conditions [1, 2] occurring co-metabolically as part of catabolic pathways where the removal of the halogen atom allows the carbon backbone to be fully degraded [3]. Organohalides are produced in Nature as part of the biogeochemical halogen cycle [4] but due to their numerous industrial and agricultural applications, these compounds now constitute a large proportion of the most toxic environmental pollutants [5].

Most microorganisms able to degrade organohalides by means of enzymatic catalysis are generally called “Organohalide Respiring Bacteria” (OHRBs) [6]. The dehalogenation reactions in OHRBs are carried out by reductive dehalogenases (RdhAs), a group of oxidoreductases (EC 1.97.1.8) that constitute a new subclass of B_{12}-dependent enzymes [7]. The RdhA enzyme family can be subdivided into respiratory and catabolic enzymes, according to their specific domain structure and putative metabolical roles [8].

Although homologous, these RdhAs present low sequence identity values that range between 20% and 30% [9]. The respiratory RdhAs couple the reduction of organohalides to the ATP synthesis via a chemiosmotic proton gradient. These enzymes function as terminal oxidoreductases of the respiratory chain as part of a protein complex of approximately 250 kDa localised in the outer face of the periplasmic membrane [10, 11]. The dimeric respiratory RdhAs are recruited to the complex by its direct interaction with the membrane anchor protein RdhB [12].

In contrast, the catabolic RdhAs, found mostly in aerobic organisms from the superphylum Proteobacteria, differ from the respiratory enzymes in many respects. The absence of a twin-arginine signal (TAT) peptide, required for translocation of the enzyme precursor through the membrane during maturation [7, 13], and the presence of a vestigial N-terminal B_{12}-binding domain, that might have originated after an ancestral duplication event and further evolutive divergence [8, 14] are but a few examples.

Metagenomic and bioinformatic analysis have shown that some catabolic RdhAs are fused to a C-terminal reductase domain, similar to the iron-sulfur (Fe-S) flavoprotein phthalate dioxygenase reductase (PDR). This likely serves to transfer electrons intramolecularly from pyridine nucleotides like NAD(P)H, without the need for external reductase systems rendering the RdhR-PDR fusion systems self-sufficient. The only characterised example of a self-sufficient RdhR is BhbA from *Comamonas* sp. 7D-2, that catalyses the aerobic debromination of the toxic herbicide bromoxynil to 4-carboxy-2-hydroxymuconate-6-semialdehyde [14].

Gene clusters related to organohalide catabolism are often found on individual plasmids. In addition to the reductive dehalogenase genes, the operons generally are composed of at least 5-6 genes, including putative enzymes that participate in the catabolism of aromatics. When the bromoxynil catabolism was first investigated *in vivo*, nitrilases, monooxygenases, and dioxygenases were identified as part of the same organohalide degradation pathway **(Supplementary Material, Figure 7)** [14, 15]. In some cases, the absence of a putative *rdhB* gene on the same operon as the structural *rdhA* has been noted, and combined with the lack of a twin-arginine-signal (TAT) peptide this suggests that some of these catabolic proteins might be localised in the cytosol instead [8].

The characterisation of the catabolic RdhA from *Nitratireductor pacificus* pht-3B (NpRdhA), an orthologous protein to BhbA that lacks the C-terminal PDR-like domain and the associated RdhB, revealed that NpRdhA is a 75 kDa monomeric protein with a globular α/β fold that catalyses the reduction of *ortho*-halogenated phenols with a marked selectivity

for 2,6-dihalogenated phenolic compounds under aerobic conditions [8]. NpRdhA was heterologously expressed in *Bacillus megaterium* as an active holo-protein, thus highlighting the possibility to use this organism for the expression of other B_{12}-dependent enzymes. Indeed, the large-scale production of RdhAs enzymes is challenging, and the production of apo-enzymes, that tend to aggregate into inclusion bodies as a result of the lack of cofactors or erroneous folding has been reported many times [16, 17].

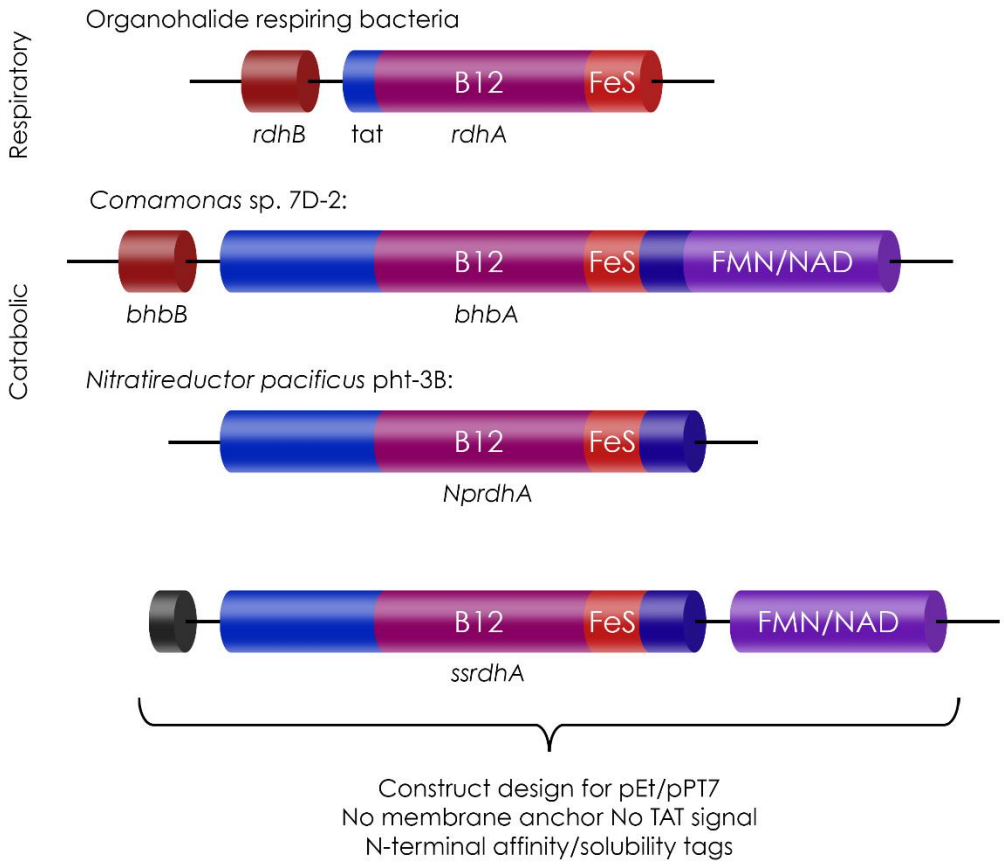


Figure 1. Representation of the domain structure of the respiratory and the catabolic reductive dehalogenases, including the ssRdhAs (Graphical abstract). In the scheme, the respiratory and catabolic enzymes are depicted according to their domain composition. The respiratory RdhAs, such as PceA from *Sulfospirillum multivorans*, are dimeric enzymes (monomers weight around 35 to 65 kDa), while the catabolic RdhAs are monomeric, like NpRdhA (75 kDa) and can contain additional domains such as BhbA from *Comamonas* sp. 7D-2 (120 kDa). The catabolic enzymes have a vestigial non-functional B_{12} domain that is likely a result of gene duplication. The TAT peptide (twin-arginine transit peptide), B_{12} (cobalamin-binding domain), Fe-S (iron-sulfur binding domain), FMN/NAD (reductase domain) are highlighted in different colours. The RdhB (equivalent to BhbB in the catabolic *Comamonas* sp. 7D-2) corresponds to the membrane anchor protein, that can be present or absent from the dehalogenase-related operons.

Thus, detailed studies of the RdhAs have been hampered for years because of the experimental complexity of working with enzymes requiring multiple cofactors, and with high sensitivity to oxidation (O_2-sensitive) [18], low expression yields, and poor solubility due to their association to membranes through its interaction with RdhB **(Figure 1)** [7, 19]. However, ssRdhAs homologous to BhbA are likely oxygen tolerant and cytoplasmatic. Furthermore, such systems offer the possibility of studying the electron transfer to the B_{12} without the constraints of utilising exogenous redox partners like flavodoxin reductase and ferredoxins to be able to turn over, as shown by Collins *et al.* [20]. Here we present initial heterologous expression of the full-length ssRdhAs and confirm efficient NAD(P)H driven organohalide reduction.

4.3 Results and discussion

The ssRdhAs can be expressed in heterologous hosts. Homologous proteins to BhbA are distributed in the superphylum Proteobacteria as either full-length PDR-fusions or lacking a NAD(P)H oxidoreductase domain, such as NpRdhA. For this study, we selected candidate sequences annotated as putative B_{12}/Fe-S oxidoreductases that exhibit the same domain structure as BhbA from *Comamonas* sp. 7D-2 and displayed between 50-70% sequence identity **(Supplementary material, Figure 1)**. We selected the homologous ssRdhA from *Ottowia thiooxydans* (WP_028602398), *Salinarimonas rosea* (WP_029029911), *Ruegeria pomeroyi* (WP_044027870), *Tropicibacter phthalicicus* (WP_099247216.1), *Pseudovibrio* sp. *FO-BEG* (WP_014287181) and *Comamonas* sp. 7D-2 (WP_015585978) for functional characterisation. The genes were codon optimised and synthesised by GeneArt®, and then cloned into pPT7 and various pEt vectors for heterologous expression.

The main objective of this work was to express heterologously an active form of at least one of the ssRdhA panel by using the same molecular cloning strategy as reported by Payne *et al*. and Judger *et al.* [7, 8] for RdhA heterologous expression in *B. megaterium* or *E. coli.* In case of *E. coli,* preliminary results obtained in our lab suggested that NpRdhA and other B_{12}-dependent proteins can be expressed in *E. coli* as fully complemented holo-enzymes when co-expressed with the BtuB transporter (B_{12}-uptake system). The latter is an outer membrane protein that actively transports cyanocobalamin across the membrane. The co-expression with the B_{12} transporter, a 66 kDa β-barrel, will increase the intracellular concentration of cobalamin during the logarithmic phase of growth and protein expression if cobalamin is supplemented in the media [21], so the rationale behind attempting the co-

expression of the transporter together with the reductive dehalogenases was clear. While *Escherichia coli* strains are cobalamin auxotrophic, the cells can acquire cobalamin via its native BtuB system. The expression of the transporter is heavily repressed in presence of B_{12} [22], but the independent expression of the BtuB via transformation of the *E. coli* host bypasses the mechanism, thus allowing the intracellular accumulation of cobalamin. We selected multiple T7 based promoter vectors for expression in *B. megaterium* and *E. coli* (under the control of xylose and IPTG as inducers, respectively) **(Supplementary Material, Figure 2A)**. Expression constructs contained a range of affinity tags, both His (6 to 10 aa) and/or Strep tags in one or both termini, depending on the vector used. The ssRdhA-pPT7/pEt constructs were cloned according to the vector specifications and the correct insertion of the *ssrdhA* genes was analysed by sequencing (Eurofins). The workflow methodology for the expression and purification of soluble ssRdhAs is summarised in **Figure 2**.

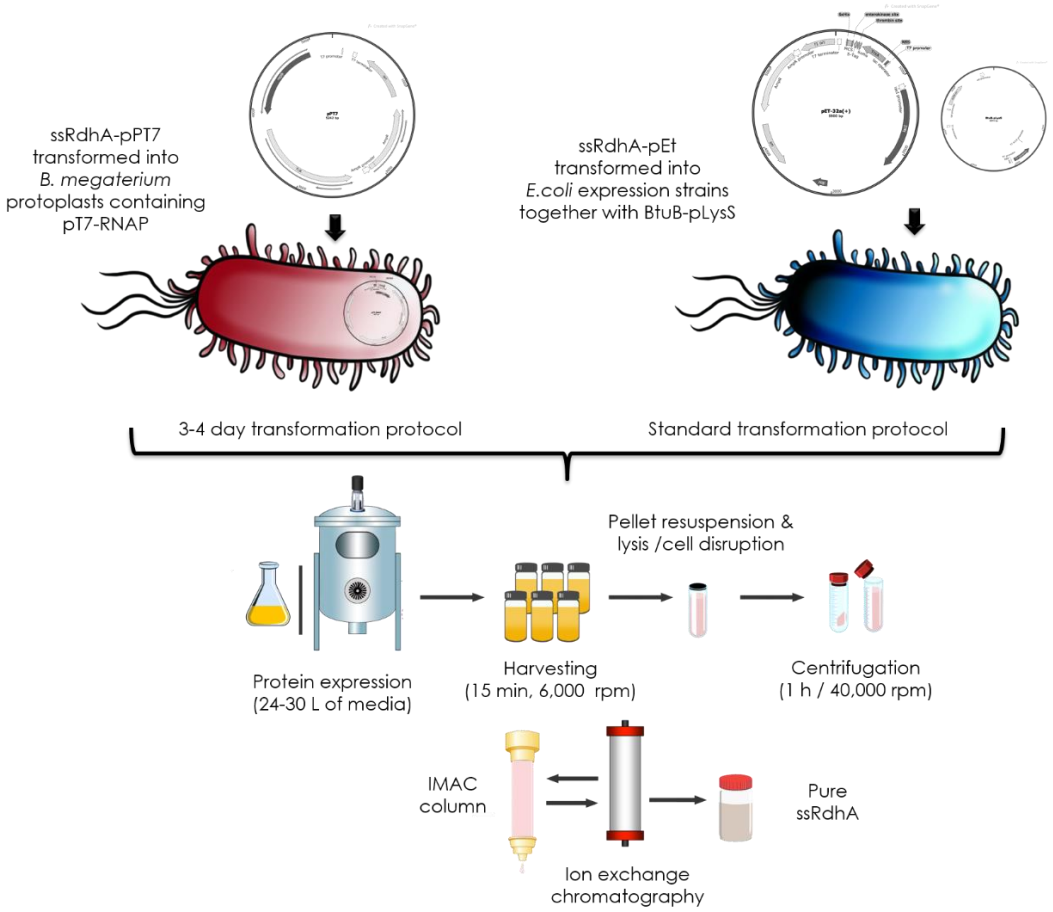


Figure 2. **Heterologous expression and purification of the ssRdhAs in both *B. megaterium* and *E. coli***. Here we schematise the transformation of *B. megaterium* MS941 protoplasts containing the pT7-RNAP vector before transformation with the ssRdhA constructs cloned into pPT7, and alternatively, the co-transformation of *E. coli* DE3 expression strains with the ssRdhA-pEt constructs and the B_{12} transporter BtuB-pLysS. The large-scale heterologous production and purification protocol are presented also.

B. megaterium MS941 protoplasts, that possess the pPT7-RNAP internal plasmid allowing the xylose-inducible expression were used for transformation with the ssRdhA-pPT7 constructs. Successful transformation was evaluated by colony PCR utilising the specific pPT7Fw and pPT7Rv primers and followed by small-scale expression trials. Western blot using an anti-6x His tag monoclonal antibody of small-scale extracts revealed the presence of two-bands at ~100 kDa in case of the 6 homologous reductive dehalogenases OtssRdhA, SrssRdhA, RpssRdhA, TpssRhhA, PvssRdhA, and BhbA cloned into pPT7. The double bands potentially suggest the presence of alternative transcription initiation sites **(Figure 3A)**.

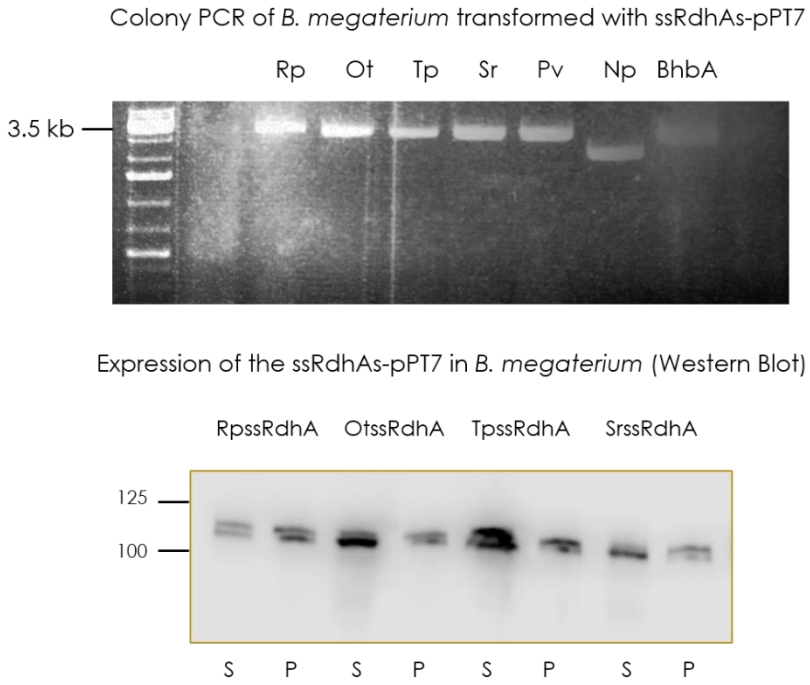


Figure 3A. Heterologous expression of the ssRdhA-pPT7 construct in *B. megaterium*. A) Colony PCR of the *B. megaterium* MS941 transformants, showing a single band corresponding to the *ssrdhA* inserts of all the homologues cloned into pPT7 and Western Blot showing a small-scale expression trials using the standard expression conditions (12 h, 17 °C, xylose 25%), revealing that different ssRdhAs present varying degrees of soluble expression.

Following detection of ssRdhA production on a small-scale, we attempted a large scale (24 L) expression of the ssRdhA using the standard conditions for *B. megaterium*, obtaining between 100 to 120 g of cells. We selected the OtssRdhA homologue to test if we could purify these enzymes in one-step via immobilised metal affinity chromatography (IMAC) under aerobic conditions. Our results showed that the OtssRdhA only marginally

interacted with the Ni-NTA resin, as opposed to NpRdhA-pPT7 construct (also tagged at the C-terminus), that can be purified in high yields after a single step of IMAC **(Figure 3B)**. We repeated these experiments using different orthologues, while also varying buffer solutions, and the concentration of NaCl, to establish if the affinity of the ssRdhA-His-tag for the IMAC could be increased. It is possible that the C-terminal His-tag adopts a conformation that does not allow for efficient interaction with the resin.

While the recombinant ssRdhA could not be purified to homogeneity by affinity chromatography, we demonstrate that the OtssRdhA (together with SrssRdhA and RpssRdhA homologues) could be partially purified using a preparative anion-exchange column (DEAE-Sepharose or Q-Sepharose Fast Flow). SDS-PAGE combined with MALDI-TOF showed that the band present at approximately 100 kDa contains the ssRdhA, however, other proteins (an elongation factor and DNA polymerase subunits) are also present in the same band at apparently higher yields **(Figure 3B)**.

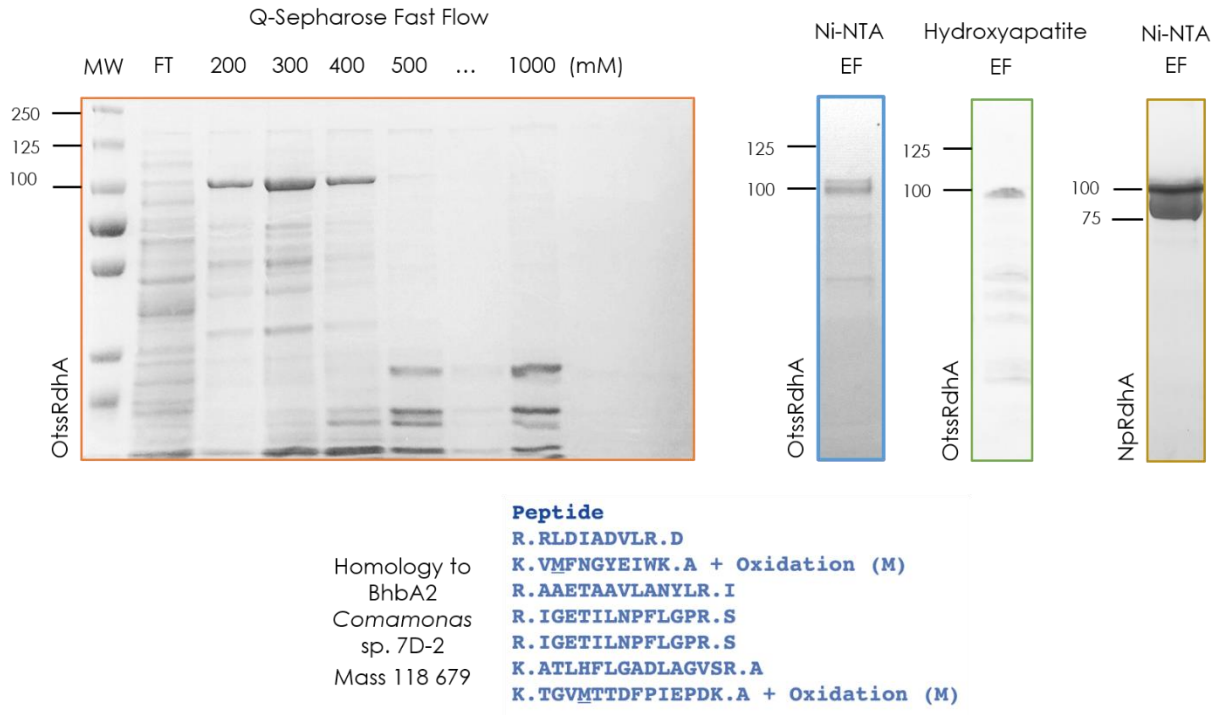


Figure 3B. Heterologous expression of the ssRdhA-pPT7 construct in *B. megaterium*. B) SDS-PAGE gels of the various purification steps attempted to isolate the ssRdhA (OtssRdhA). The anion exchange chromatography using the Q-FF resin, performed after a preparative step of soft anion exchange that allowed the concentration of the sample is presented first. The protein of interest eluted between 200 and 400 mM NaCl. This step was followed by IMAC (using a gravity-flow Ni-NTA column) or a ceramic hydroxyapatite purification step. We also show the MALDI-TOF results, showing the calculated mass and the peptide hits indicating homology with BhbA. The purification of NpRdhA-pPT7 (C-terminal His-tag) via IMAC, is shown as a reference.

Unfortunately, a change in parameters such as temperature (T=15 to 20 °C) and concentration of the inducer (25 to 50% xylose) did not improve the low ssRdhA yield, nor the use of an alternative expression vector for expression in *B. megaterium* (pNHis1622/pCHis1622) (data not shown). Hence, we concluded that the ssRdhA proteins could not be overexpressed in sufficient amounts to allow its biophysical characterisation in this host. However, we developed whole-cell assays that enabled us to prove that the expressed ssRdhAs were active regardless of the low yields, as will be shown later in this work **(Supplementary Material, Figure 3A)**.

Hence, we moved to analyse the constructs designed in pEt vectors for heterologous expression in *E. coli*. Following the co-transformation of *E. coli* expression strains with the *ssrdhA* genes and *btuB*-pLysS/pEt3a plasmids, a similar protocol for evaluation of small-scale expression was performed. In this case, we selected only a couple of ssRdhAs (the orthologues OtssRdhAs and TpssRdhAs) to develop a more complex grid evaluation of a range of expression experiments with the ssRdhA-pEt constructs.

The heterologous expression of respiratory RdhAs in *E. coli* has been reported before, but invariably yielded insoluble and inactive samples [12, 23, 24], likely due to its inability to synthesise cobamides *de novo*. Regardless, *E. coli* remains a convenient host for recombinant production, and it offers additional variables to test for soluble expression of the ssRdhAs, such as using different expression vectors and strains (i.e. DE3 strains, HMS174, BL21, soluBL21, and Arctic Express). The pEt expression systems support facile inclusion of both solubility enhancement tags and multiple affinity tags. Furthermore, the pEt vectors were chosen also to minimise leaky expression to prevent the accumulation of miss-folded protein.

It is known that overexpression in *E. coli* can lead to the aggregation of recombinant proteins into inclusion bodies, but *in vitro* reconstitution experiments to incorporate cobamides and Fe-S into the RdhA apo-enzymes recovered from the insoluble fraction have proved extremely challenging, and the refolding experiments have to be performed under strict anaerobic conditions, and limited success has only been reported for the respiratory VcrA of *Dehalococcoides mccartyi* CBDB1 [16]. With that in mind, the expression trials in this host were aimed to optimise the yield of soluble ssRdhAs, as the reconstitution of the protein from inclusion bodies was never considered an option. The design of the ssRdhA-pEt construct greatly depended on the intrinsic features from the vectors, but the observations previously done when working with the pPT7 constructs hinted at the fact that the most convenient way to tag the protein was at the N-terminus. Both ssRdhA-pEt28a(+)constructs and the ssRdhA-pPT7 ones were co-transformed with the *btuB* system

in *E. coli*. Analysis of cell extracts following transformation revealed that the corresponding ssRdhA protein was found both in the soluble and insoluble fractions and once again at low yields **(Figure 4A)**.

In an attempt to increase the expression levels of the ssRdhA as soluble holo-enzymes, the induction times were changed, in general, the expression was performed aerobically over-night (12 h), except when auto-induction media was utilised. The IPTG concentration was varied between 0.25 to 1.0 mM, and the expression proceeded at low temperatures (12-18 °C), depending on the *E. coli* strain used. Each large-scale production consisted of 30 L of media (M9, LB, and TB) distributed in 2 L flasks that were supplemented individually with cofactors after induction (the cobalamin was added as hydroxocobalamin and the Fe was supplemented in the form of soluble salts) **(Supplementary Material, Figure 3)**.

After harvesting, the cells were lysed aerobically in most of our purifications using a cell disruptor (Constant Systems) or anaerobically using a French press (Thermo IEC). Following ultracentrifugation, the clarified lysates were loaded into gravity flow IMAC columns, either under anaerobic conditions using a glove box (Belle Technologies) with an N_2 atmosphere, or under aerobic ones, as previous work on NpRdhA suggests the catabolic enzymes are not O_2-sensitive [8, 20, 25]. When IMAC chromatography was attempted, the OtssRdhA eluted at approximately 40 mM-100 mM imidazole suggesting that the affinity for the Ni^{2+} resin is low, resembling the results obtained when using *B. megaterium*. Although the yields were marginally better, the number of impurities that co-eluted with the protein after a single step purification will not allow its biophysical characterisation.

The addition of so-called solubility tags, when fused to the protein of interest, has been shown to increase the solubility and the yield of recombinant proteins [26]. This strategy has proved useful when producing soluble respiratory RdhAs, although those enzymes were inactive [24]. Thus, we decided to test both the inclusion of solubility tags and the optimisation of the affinity tags, by designing a construct that possess a 10x His-tag to increase IMAC affinity, and a Strep-tag, to allow a second affinity chromatography step if required. We cloned a construct that presented a thioredoxin tag (Trx) solubility tag (109 aa), utilising the pEt32(a) vector. Expression studies revealed this construct yielded protein that could readily be purified under aerobic conditions using the Ni-NTA resin in HEPES-KOH 50 mM, pH range 7.5 to 8.0, and NaCl 250 rendering a considerable amount of soluble protein **(Figure 4B)**. However, SDS-PAGE and Western Blot revealed that the protein undergoes considerable degradation, despite the addition of protease inhibitors cocktails, not only during the lysis but also directly into the buffer solution during purification. Sample

degradation was reduced by diminishing the expression time (6 h) and changing the temperature during growth and induction, avoiding a sudden drop in the temperature of the shakers (although cells were grown 37° C, before induction, we changed the temperature to 24° C for 1 h, and then the temperature was changed again to 15° C, or 12° C when using Arctic Express).

Furthermore, the addition of stabilisers (such as trehalose) to the buffer helped to diminish the degradation. Intriguingly, different preparations of the ssRdhA presented different degrees of degradation. This might be linked to the fact that both the pEt systems and the *btuB*-pLysS vector have the same origin of replication OriC, leading to genetic instability. We developed a more reliable strategy and cloned the *ssrdhA* or the *btuB* into vectors that either possessed distinct origins of replication or included both genes into the same plasmid with two multiple cloning sites; the ssRdhA-pEtDuet-1 was chosen for this reason. We repeated the large-scale production procedure described before with a double-tagged His/Strep-OtssRdhA construct cloned into pEtDuet-1 showing that the protein could be purified with an improved affinity to the Ni-NTA (as the protein eluted at 100 mM imidazole), however, the purity of the samples was still an issue. Additional purification steps were done to improve the purity of the ssRdhA (from either pEtDuet-1 or pEt32a(+) constructs), including anion-exchange chromatography using the ÄKTA Purifier (GE) (Res-Q and HiTrap Q), together with analytical Gel Filtration (Superdex-200 column (GE)). The elution fractions were analysed by SDS-PAGE, and Western Blot and MALDI-TOF confirmed the ID of the protein by tryptic digest **(Figure 4)**. It was clear that the presence of the solubility tag increased the yield of the protein, in comparison to the constructs cloned only with affinity tags. We also attempted the purification of this construct using the Strep-tag, however, the affinity for this resin was low and most of the sample was eluted during the washing step.

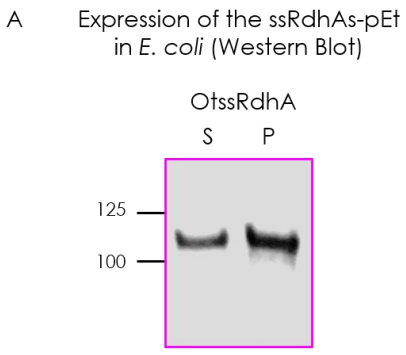


Figure 4A. Heterologous expression of the ssRdhA-pEt constructs in *E. coli*. A) Western blot of the small-scale expression of the OtssRdhA in a pEt system using standard expression conditions for this host.

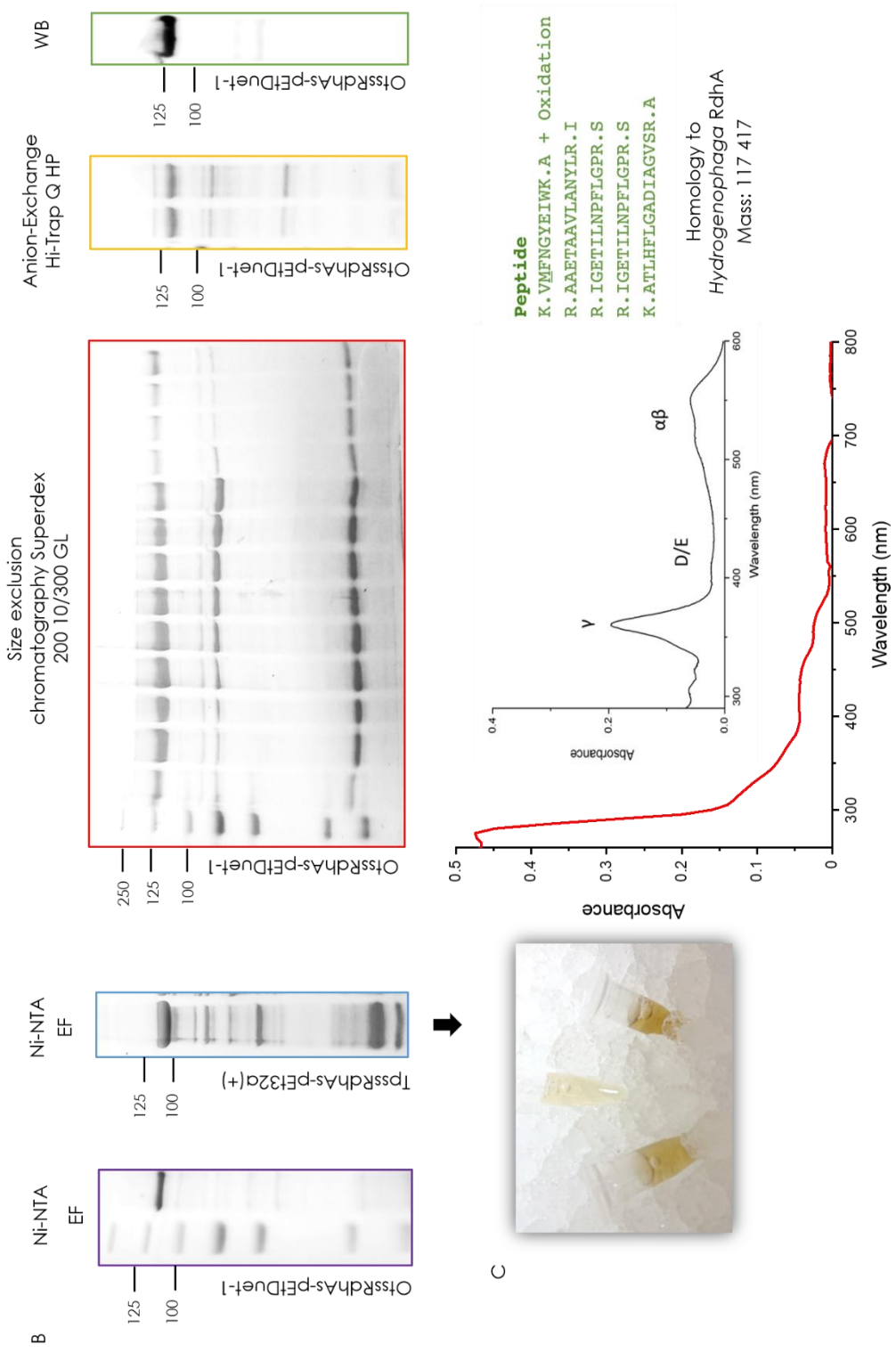


Figure 4B and 4C. Heterologous expression of the ssRdhA-pEt constructs in *E. coli*. B) SDS-PAGE of the IMAC (Ni-NTA) purification of OtssRdhA, followed by size-exclusion and anion-exchange purification steps using the ÄKTA Purifier (GE). C) UV/Visible spectra of the purification of the TpssRdhA, showing the cyanocobalamin determination by thermal denaturation of the sample after the addition of KCN (inset); the UV/Visible spectra revealed that the sample also contains FMN. MALDI-TOF results indicate homology to the reductive dehalogenase of *Hydrogenophaga*. Data was plotted using OriginPro ®.

The ssRdhAs are expressed as fully complemented holo-enzymes. UV/visible spectroscopy was used to determine if the cofactors were incorporated into the purified ssRdhAs. Again, we summarise these results using the *Tropicibacter phthalicus* and *Ottowia thiooxydans* ssRdhAs in different vectors (pPT7, pEt32a(+), and pEtDuet-1), as examples. The UV/Visible spectra from the heterologously expressed and purified ssRdhA samples show a broad absorbance between 300 and 600 nm indicative of the presence of Fe-S clusters and corrinoids in aerobic conditions, feature that is typical of all RdhAs and QueG [27], with an absorption maximum at approximately 420 nm **(Figure 5A)**. We also used the light-induced reduction under anaerobic conditions of the Co(II) to Co(I), using 5'-deazariboflavin and a 405 nm blue LED for 10 min, to quantify the amount of cobalt present per mole of protein.

Additionally, we used the thermal denaturation of the samples in presence of potassium cyanide, as a quick and inexpensive methodology to determine the presence of cobalamin. The recorded UV/Visible spectrum revealed the typical spectral features of cyano-cobalamin, showing well-defined absorption peaks known as γ (360 nm), D/E (400-420 nm), and the αβ (525-560 nm) [28] **(Figure 4C Detail)**. Alternatively, the RAMAN scatter spectrum of a powdered sample of the TpssRdhA semi pure enzyme, obtained using CaF and a LASER pulse of 532 nm, is also similar to those reported for free B_{12} **(Figure 5B)**, proving that this method, albeit destructive, can be used as an alternative for the qualitative determination of cobalamin.

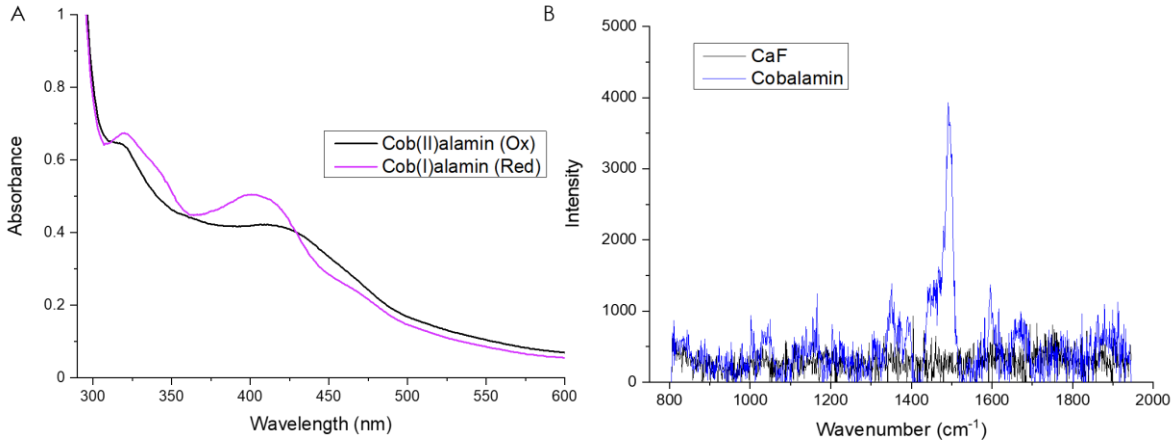


Figure 5. Spectroscopic characterisation of the TpssRdhA. A) UV-Visible spectrum of 75 µM of the Trx-TpssRdhA enzyme purified from *E. coli* under aerobic conditions and after the light-induced reduction of the metal center of the corrinoid to Co(I), using 5'-deazariboflavin, EDTA, and a 405 nm blue LED. B) Raman scatter spectrum of a powdered sample of the semi-purified enzyme from *B. megaterium*, using a 532 nm LASER. Data was plotted using OriginPro ®.

Although the data presented here are indicative of the presence of cofactors, the ssRdhA samples were highly heterogeneous and the purity of some protein samples was often lower than 70 %, complicating the analytic determination of the cofactors. In fact, some of the samples did not contain cob(II)alamin at detectable levels, although they contained Fe-S and FMN, as revealed by the acid thermal denaturation or EPR. We obtained the EPR spectrum for one of our constructs (OtssRdhAs-pEtDuet-1) in the presence/absence of a typical substrate 3,5-dibromo-4-hydroxybenzoic acid (although the cobalamin content in some of our samples accounted only for approximately 20 % of the total protein). The presence of the cobalamin is supported by the EPR active Co (II) hyperfine coupling (with its 8 spectral features with an I=$^{7/2}$ for ^{59}Co). As reported previously in the case of the NpRdhA, our spectra also show a direct interaction between the enzyme and the 3,5-dibromo-4-hydroxybenzoic acid, as can be assumed from the presence of super-hyperfine coupling. Interestingly, our results also show a species highly similar to a Co (III)-O_2^- radical **(Figure 6).**

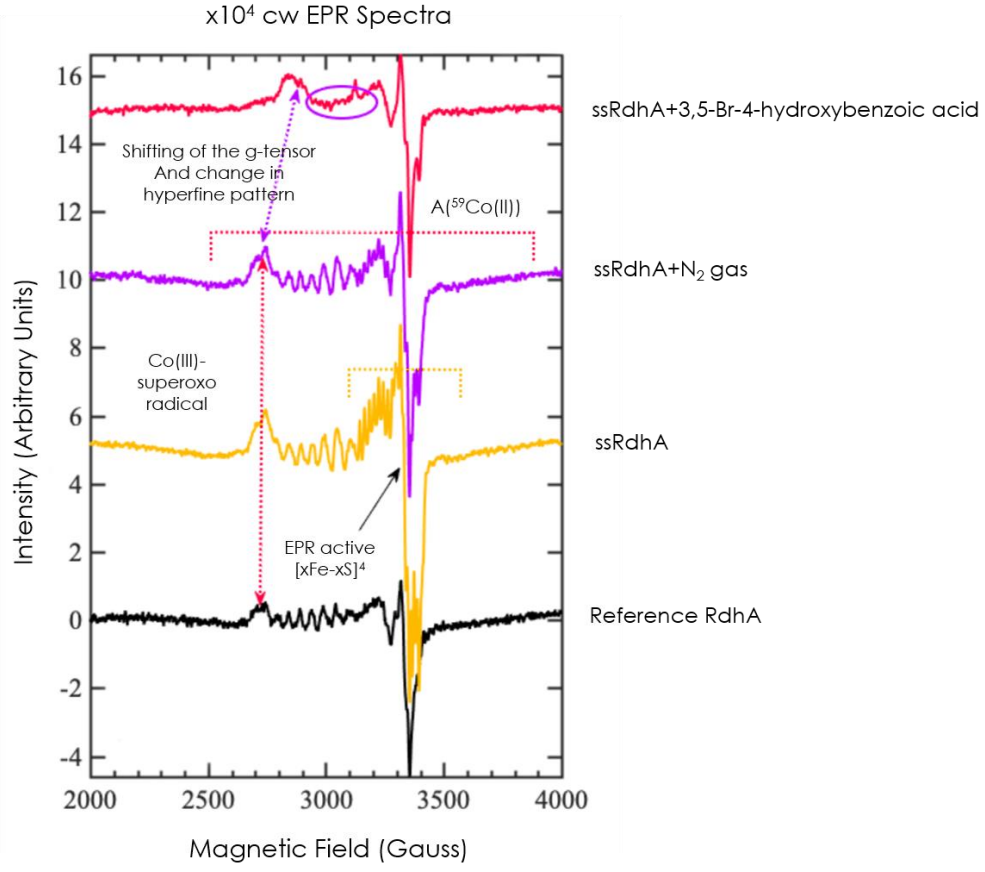


Figure 6. EPR spectroscopy of the OtssRdhA. X-band (9.5 GHz) continuous wave of 150 M of the ssRdhA. The presence of the cobalamin is supported by the EPR active Co (II) hyperfine coupling (with its 8 spectral features with an I=$^{7/2}$ for ^{59}Co). The spectrum shows a direct interaction between the Co (II) center in the cobalamin and one of the halogen substituents from the substrate 3,5-dibromo-4-hydroxybenzoic acid, according to the

changes in the super-hyperfine pattern. The appearance of a super-hyperfine quartet shows the coupling Metal-X. Our results also suggest the presence of a species resembling the spectrum of Co (III)-O_2^- radicals.

We also attempted the determination of the Fe content using a colourimetric bathophenantroline assay, revealing the presence of 8.67 ± 0.25 mol of Fe^{3+} per mol of protein, with the expected value of 10 for the ssRdhAs (given the presence of a simple [2Fe-2S] cluster, in addition to the two [4Fe-4S].

Although the inclusion of B_{12} is low in the ssRdhA enzyme preparations, the total protein yields are encouraging. Previously, the production of catabolic holo-enzymes was limited to NpRdhA, and even other homologous catabolic proteins, such as the *Burkholderiales joshi* RdhA, presented very low B_{12} or undetectable incorporation when expressed in both heterologous hosts [25]. It is suspected that the cobalamin binds to the active site as part of the folding process, and once the protein core is folded the binding of B_{12} is not possible, explaining why most reconstitution experiments have been unsuccessful with the notable exception of Parthasarathy *et al.* [16].

This might be of particular importance in the case of the catabolic RdhAs, which contain an additional vestigial B_{12} domain that might lead to enhanced folding in absence of cobalamin. In this context, the presence of bulky solubility tags included in the ssRdhA constructs to increase the expression yield and solubility of the enzymes, such as the Trx-tagged ssRdhA, might also affect the incorporation of the cobalamin, as was suspected by Sjuts *et al.* [24] in case of the PceA tagged with the trigger-factor and even have some effect in the activity of the enzyme.

It is also possible that the quantitative incorporation of the B_{12} cofactor might also depend on the type of B_{12} supplemented in the media during growth, as the respiratory RdhAs have been reported to possess specific cobalamin derivatives with differences in the nucleotide loop composition and additional modification on the ring system, such as the norpseudovitamin B_{12} [10]. It is unclear how the cofactor loading of the enzyme occurs in the native organisms and if there is specialised machinery involved, as is suspected in the case of the respiratory RdhAs that possess the TAT sequence [7, 19]. But fortunately, the incorporation of the Fe-S is not an issue when using *E. coli* as expression host, and the supplementation with soluble iron salts before induction allows the native machinery to assemble the Fe-S clusters, as shown by EPR and Fe extraction.

A more homogeneous incorporation of the cobalamin throughout the sample thus remains the main obstacle for the overexpression of active ssRdhAs, fact that emphasises

the relevance of optimising or developing not only a new methodology to express the BtuB before the expression of the reductive dehalogenase starts (by using, for example, two independent induction periods, started by the addition of a different inductor molecule), to guarantee that the intracellular concentration of the B_{12} is high enough before the synthesis of the ssRdhA).

The heterologously expressed ssRdhA are active holo-enzymes. We tested whether the expressed ssRdhAs were able to catalyse the dehalogenation of 3,5-dibromo-4-hydroxybenzoic acid to 3-bromo-4-hydroxybenzoic and/or 4-hydroxybenzoic acid, as is the case of BhbA [14]. The dibrominated benzoate is a degradation product of the herbicide bromoxynil, one of the most widely used pesticides in agriculture, that interestingly has also been reported in marine environments [29].

Initially, we tested the activity of the enzymes by following the conversion of the substrate to the partial and/or completely reduced products by reverse phase HPLC. Semi-pure preparations of the ssRdhAs (70 - 80%) coming from pEt32a(+) and pEtDuet-1 were assayed anaerobically to test the dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid substrate *in vitro*, in presence of excess NADPH for fixed time-periods at 25 °C. Our results reveal that after 1.5 h of reaction, the conversion to 3-bromo-4-hydroxybenzoic acid occurred almost completely **(Figure 7A)**. Interestingly, the full conversion to 4-hydroxybenzoic acid was not observed and only approximately 18% of the fully dibrominated product was formed, suggesting that this enzyme possess different affinities for the two substrates (the dibrominated and the nonbrominated species). The assays were done in presence and absence of the NADPH, and controls included adding an inactivated protein sample. Our results suggest that the addition of pyridine nucleotides was necessary for the quantitative dehalogenation of the substrate to proceed after short periods of incubation at 25° C, and that the reaction did not occur spontaneously, without the addition of the enzyme in the reaction mixture.

Additionally, we performed a substrate-dependent NAD(P)H oxidation experiment monitored by UV/Visible spectroscopy, but the results were inconclusive and therefore, the data was not included here. The reaction mixtures were prepared in presence of excess NAD(P)H, and then the reaction was initiated by the addition of the enzyme or the halogenated substrate. These experiments were performed anaerobically to keep the O_2-dependent NAD(P)H oxidation rate minimal. The reaction was followed for a period of 5

min and proved to be an alternative method that also suggests that the activity of the ssRdhA depends on pyridine nucleotides.

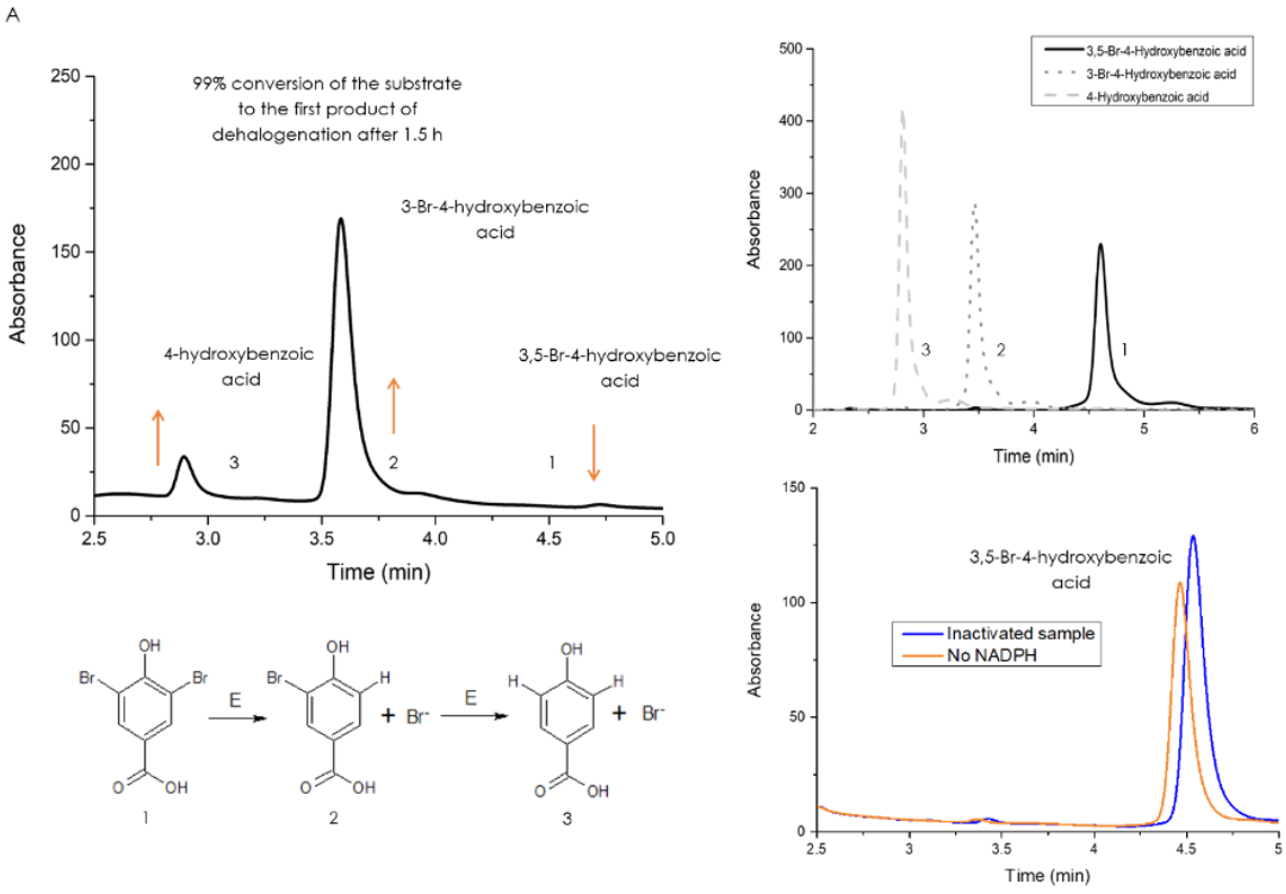


Figure 7. Activity of the purified full-length RdhA enzyme. A) HPLC chromatogram showing the *in vitro* formation of both products of dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid substrate after 1.5 h of reaction in presence of 5 mM NADPH, at 25°C, showing that the fist dehalogenation reaction occurs quantitatively. Additionally, we present controls containing an inactivated protein sample and a reaction mixture with no added NADPH.

As was mentioned before we also measured the dehalogenation activity in the clarified lysates (after centrifugation) and in whole-cell assays **(Supplementary Materials, Figure 4A)** utilising *B. megaterium*, where the protein could not be purified to homogeneity. Our results regarding the expression in *B. megaterium* are similar to those reported by Kunze *et al.* [30] when using *Shimwellia blattae,* to express the respiratory RdhA from *D. hafiense Y51* as the recombinant protein was expressed but could not be purified to homogeneity and dehalogenation activity was only detectable in crude extracts. Interestingly our results suggest that the dehalogenation reaction to generate the 3-bromo-4-hyrdoxybenzoic acid occurs, albeit less quantitatively, in comparison to the assays done with pure or semi-pure protein *in vitro*. If the full conversion to the 4-hydroxybenzoic acid

occurs, we suspect that the fully dehalogenated product does not accumulate, and instead is metabolised quickly so that its detection by HPLC is impossible. Our results suggest that the dehalogenation reaction might take place simultaneously to other equilibria *in vivo,* such as the decarboxylation of the substrates/products in whole-cell experiments.

Activity assays performed with sodium dithionite reduced methyl viologen under anaerobic conditions (instead of NADPH) indicate that the ssRdhA (particularly the OtssRdhA) might have similar kinetic parameters to NpRdhA, as revealed by the calculated kinetic parameters obtained after a non-linear fitting (Michalis-Menten model). However, this assay likely bypasses the reductase system (as the MV transfers the electrons directly to the cofactors on dehalogenase domain). Thus, this assay allowed us to test the integrity of this individual domain, in a similar way as the cytochrome C reduction experiments can be used to study the activity of the PDR-like reductase domain **(Supplementary Material, Figure 4B)**.

Although a proper steady-state characterisation of the reaction catalysed by the ssRdhA is lacking still, our preliminary results indicate the dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid substrate is dependent on NAD(P)H (although the NADH data is not shown). According to previous data obtained in case of NpRdhA, the presence of a hydroxyl group in the *ortho* position with respect to the halogen atom in aromatic organohalides is a strict requirement for the reductive dehalogenation reaction to occur, a fact that strongly suggests the relevance of this substituent during catalysis. This hypothesis can further be tested by using other halogenated species, that present different substitution patterns.

The self-sufficient ssRdhA architectures characterised here offer a direct possibility to study the mechanistic details of reductive dehalogenation **(Figure 8)** since there is no requirement for exogenous redox partners that limit turnover, as is the case of NpRdhA [20]. However, at the present stage of this project, it is impossible to contribute to the mechanistic discussion, but once a more robust strategy is established to express these enzymes in a standardised and reproducible manner, the ssRdhAs system will offer the possibility to study the electron transfer from pyridine nucleotides NAD(P)H via FMN, and the Fe-S clusters to the cobalamin cofactor. Interestingly, some of the key catalytic residues found in the catabolic NpRdhA and the respiratory PceA are conserved in all the orthologous ssRdhAs that we selected initially in this work **(Supplementary Material, Figure 5)**.

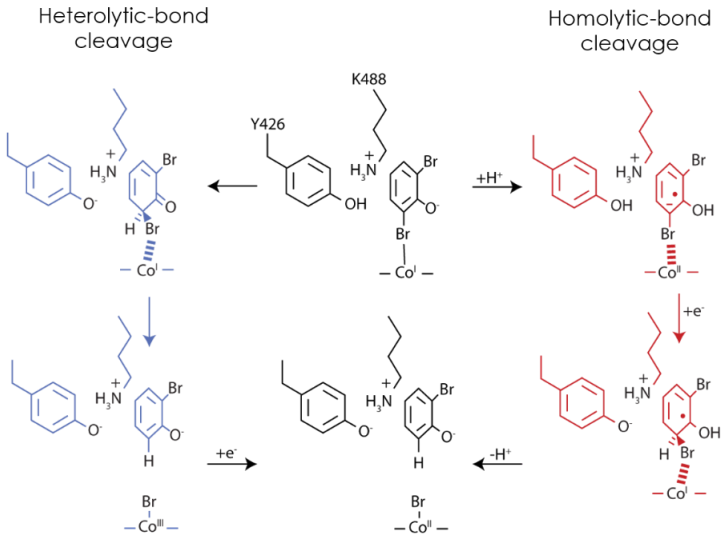


Figure 8. Although the precise reaction mechanism remains unknown, there are different proposals to explain the enzyme catalysed reductive dehalogenation and all of them depend on the initial formation of a super-nucleophilic species of Co(I) by direct electron transfer from the proximal Fe-S [8, 30], but the sequential steps of the catalysis remain one the biggest question on the field. In this context, achieving the soluble expression of the ssRdhAs in enough yield can help to elucidate the individual steps of the catalytic mechanism.

Homology models of the ssRdhA reveal interesting features and a dynamic nature. As shown by Judger, *et al.* [17], the removal of both the TAT signal peptide of the membrane-spanning domains (RdhB) from the respiratory RdhAs does not guarantee the soluble expression of the protein in heterologous hosts. Previous experience suggested that BhbA, when expressed in *B. megaterium* was either not expressed or expressed in the insoluble fraction, even when the membrane anchor was not included in the design of the constructs. The work presented here with various orthologous ssRdhA from α and β-Proteobacteria has confirmed the observation and we believe that the lack of cofactors (cobalamin, Fe-S, and FMN) might still be the main cause of aggregation, leading to the accumulation of miss-folded protein.

After a close inspection of the OtssRdhA homology model, generated by I-TASSER [31] and refined using a combined protocol of energy minimisations and molecular dynamics simulations with the Amber99 force-field **(Supplementary Material, Figure 6)** we identified disordered and dynamical regions in the ssRdhA structure that may increase the probability of aggregation during expression and purification, located between the typical B_{12}/Fe-S domain and the PDR-like reductase domain, that consists of a mobile linker region that adopts different conformations during a productive MD simulation of 5 ns **(Figure 9)**. These observations were further confirmed by the secondary structure

prediction software JPred [32], but experimental proof is required. An internal peptide of approximately 40 aa, that was identified in the non-functional B_{12} domain is not present in the naturally truncated NpRdhA, interestingly. Although its identity has not been determined it resembles a transit peptide or a region that might serve to mediate interactions with other proteins; as at least 6 residues within a segment of 37 have been predicted to act as functional residues for post-translational modifications, according to a neural-network algorithm (ConSurf) [33]. The question as to how the ssRdhAs interact with the membrane, by binding to the RdhB is still unsolved, so the presence of this region might be relevant for protein-protein interactions and have physiological implications.

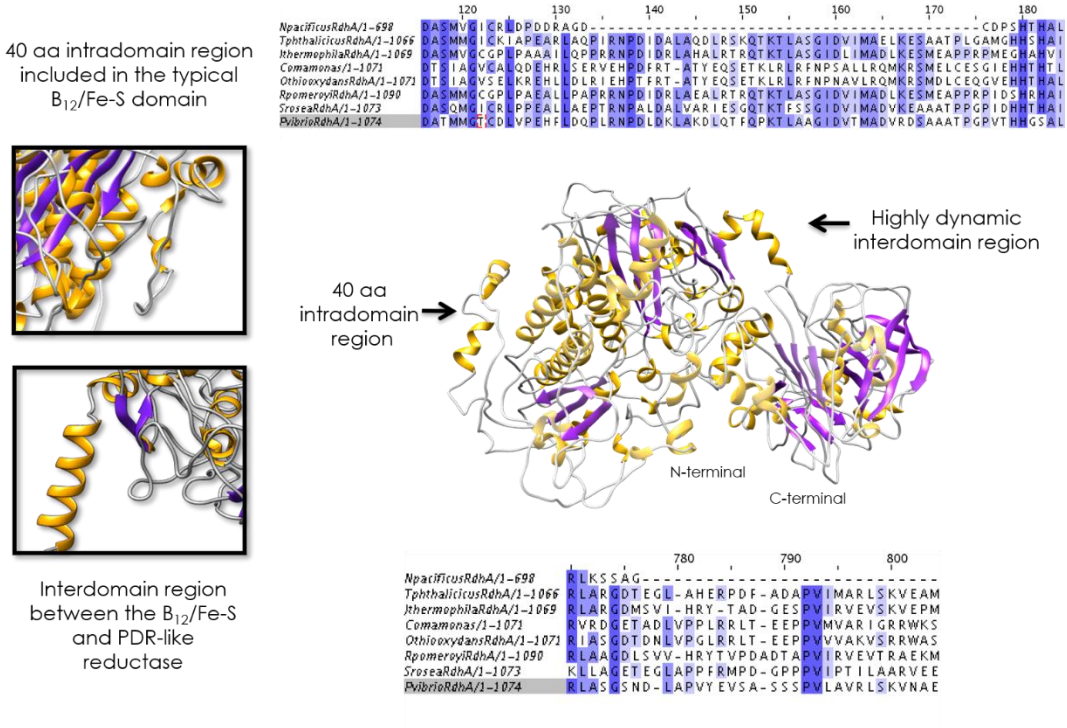


Figure 9. The OtssRdhA homology model reveals some structural features of the ssRdhAs. Highlighted here are the intrinsically disordered and dynamic regions in the OtssRdhA model. The primary sequence alignment indicates that the internal peptide (40 aa) is present in all the ssRdhA orthologous proteins, but not in NpRdhA. The highly dynamic interdomain region is composed of approximately 22 aa and it is speculated that it allows great conformational flexibility between domains during the catalysis. Alignment done using Jalview (MAFFT with defaults) [34].

4.4 Conclusion

In comparison to the NpRdhA and related B_{12}-dependent enzymes, such as epoxyqueuosine reductase (QueG) [27, 35], the heterologous expression of the self-sufficient enzymes is experimentally more challenging. Despite utilising the same

methodologies and bacterial hosts, the yield of soluble holo-enzymes was lower with inefficient incorporation of the cobalamin. Expression of a panel of candidate ssRdhAs *in B. megaterium,* using both the pPT7/pCHis1622 and pNHis1622 constructs, yielded protein that was expressed poorly regardless of the vector used or the specific orthologues selected. Hence, we concluded that the ssRdhAs cannot be expressed in this host in yields appropriate for further biochemical characterisation.

Previous attempts at expressing reductive dehalogenases in *E. coli* failed **[16, 17, 24]** likely due to the low levels of intracellular B_{12} in *E. coli,* that lacks the ability to synthesise this cofactor *de novo*. This was circumvented to an extent by the simultaneous overexpression of the vitamin B_{12} transporter BtuB. The BtuB co-transformed *E. coli* allows the incorporation of the B_{12}, yielding active protein, and its cofactor content can be determined qualitatively and quantitatively via colorimetric methods. However, the samples are heterogeneous and only a small percentage (ranging between 15 to 25 %) of the sample is expressed as active holo-enzyme. For that reason, the development of prototrophic *E. coli* strains that can synthesise cobamides *de novo*, as reported recently by Fang *et al.* **[36]**, may ultimately prove to be a better solution to achieve homogeneous production of B_{12}-dependent enzymes.

This work outlines that although the full-length ssRdhA proteins can be expressed without the solubility tags, the yields are noticeably lower when only the affinity tags are included in the construct. If the strategy to add solubility tags to increase the expression yields is maintained in future expression attempts, such as in the Trx-ssRdhA constructs, a compromise between the increased yield of soluble protein, the incorporation of $B_{12,}$ and the activity of the enzyme has to be found, as our preliminary results (not shown in this work) indicate that the solubility tags may not only affect the inclusion of the cofactor, but also the activity of the enzyme due to steric hindrance.

Even though the results presented here are a qualitative description of the ssRdhA rather than a quantitative characterisation, this work provides a starting point and outlines the future strategy to follow for the heterologous expression of the self-sufficient catabolic dehalogenases. Our results show that these enzymes can be heterologously expressed as soluble and active holo-enzymes if co-expressed with the BtuB transporter in *E. coli.* Optimisation of the expression and purification might allow a robust characterisation of the self-sufficient RdhAs both spectroscopically and kinetically, and this will hopefully enable a complete description of the chemical mechanism of dehalogenation, without the limitations of other systems that depend on external redox partners. Many questions remain open

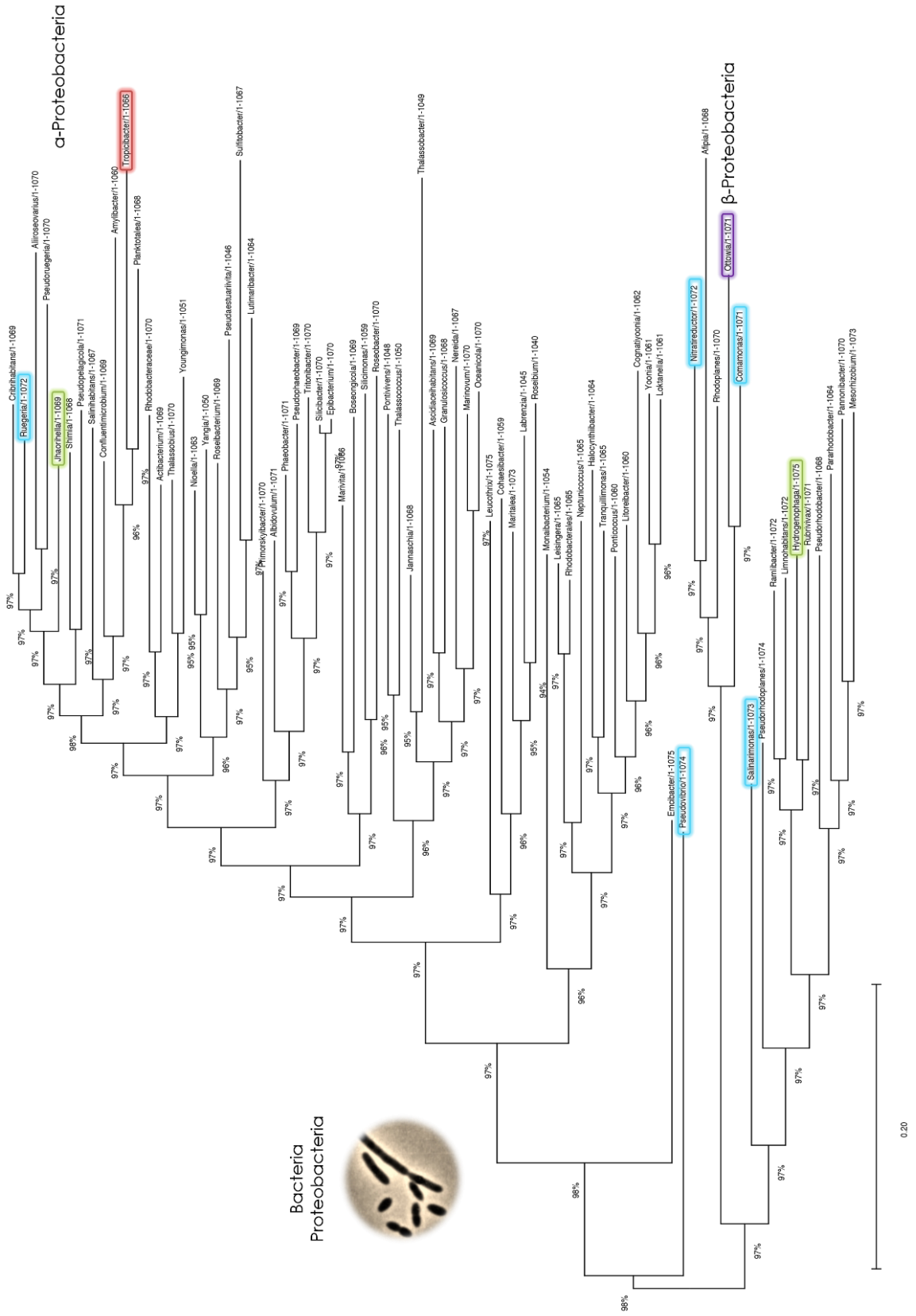
about the ssRdhAs, but we trust that in the future this architecture will provide definite answers for the long-standing questions in the field of the reductive dehalogenation.

The authors declared no conflict of interest. Samantha Gaytán (S.G) and Irina Gostimskaya cloned the ssRdhA constructs. S.G and Karl Payne transformed *B. megaterium*. S.G expressed and purified the ssRdhAs, performed metal-cofactor determination, UV/Visible spectroscopy, and prepared the samples for Raman spectroscopy, and EPR. S.G and Godwin Aleku measured the activity of the enzyme. Muralidharan Shanmugan performed EPR studies and assisted with the interpretation. S.G. performed the *in silico* analysis presented in this work and the phylogenetics. S.G. wrote the manuscript. David Leys designed and directed the research.

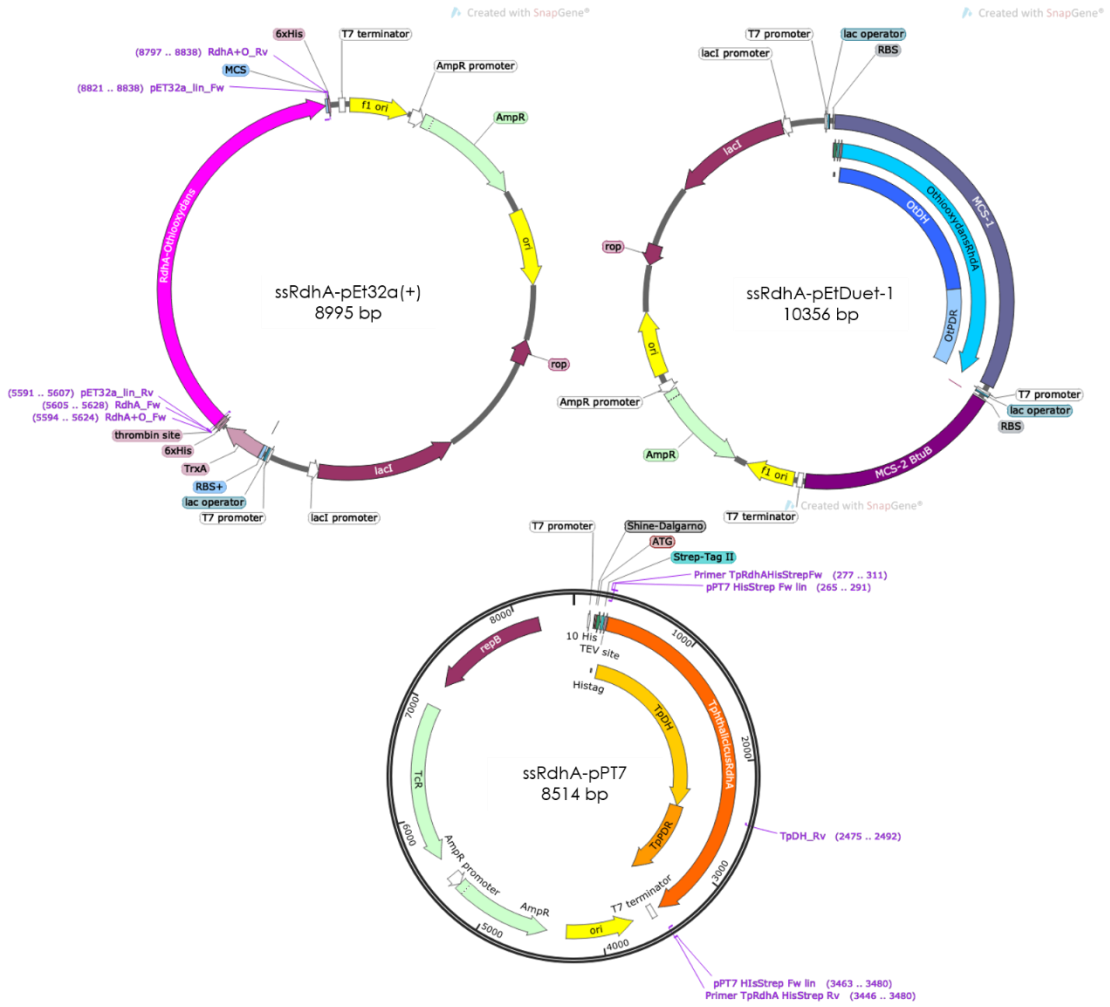
4.5 Acknowledgments

Many thanks to Dr. Bat-Erdene Judger for his valuable recommendations during the preparation of this manuscript. I want to say thanks also to Professor Rogelio Rodríguez (Facultad de Química, UNAM) for his help with the *in silico* part of this project. Thanks to Dr. Bethan Mcavan too for her help with the Raman spectroscopy. Special acknowledgment to Professor Martin Warren (University of Kent) for sharing the BtuB expression vector with the Leys group prior to publication. Samantha Gaytán received a scholarship from CONACyT (Mexico). D.L. is a Royal Society Wolfson Research Merit Award holder.

4.6 Supplementary Data



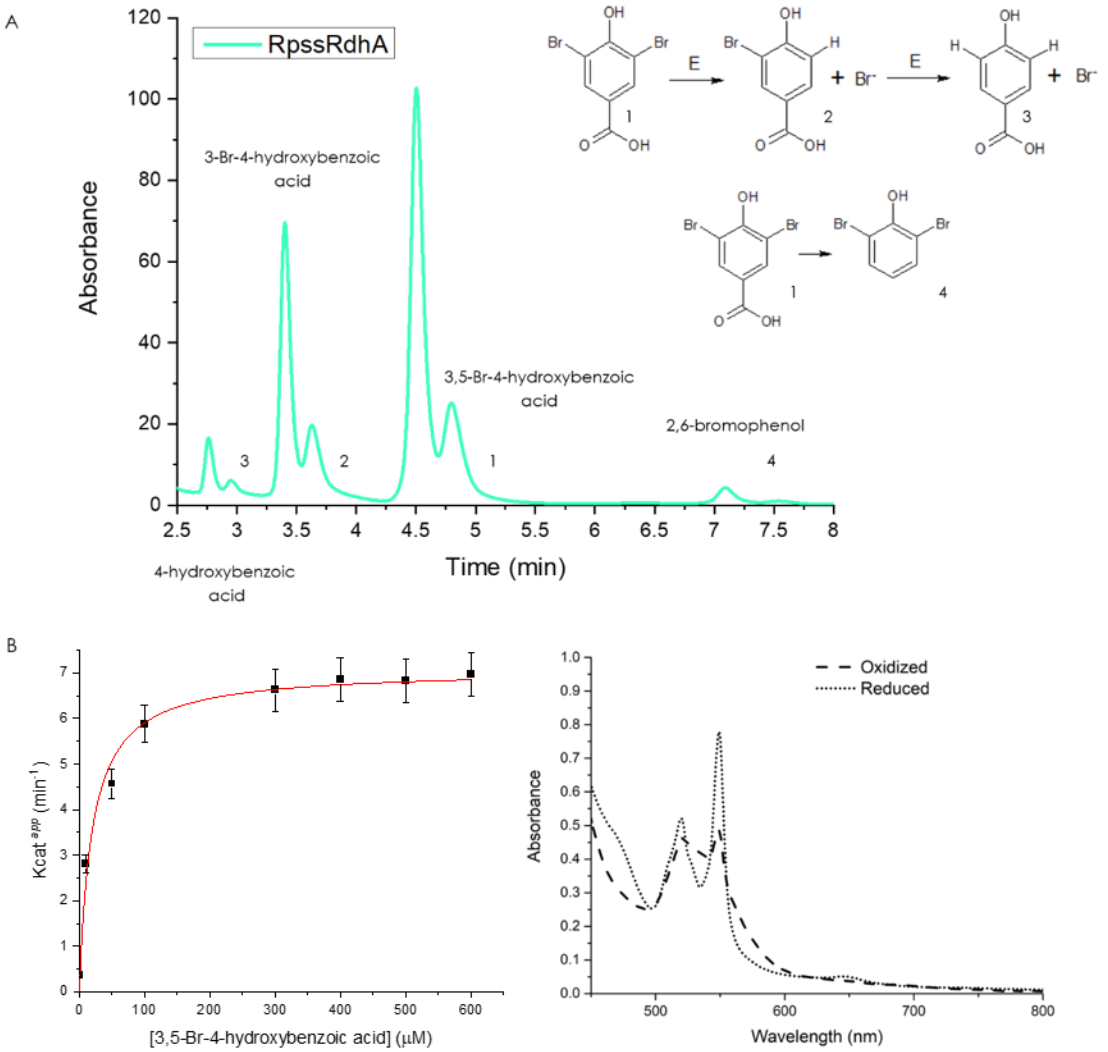
Opposite Page. Supporting material, Figure 1. NJ-phylogenetic dendrogram of the ssRdhAs. Only 7 % of the genomes analysed by Liu *et al*. [37] possess plasmids related to organohalide metabolism in α and β Proteobacteria. Although it is not clear if these operons contain only catabolic ssRdhA (or the naturally truncated counterparts such as NpRdhA), our PSI-BLAST results suggest that the self-sufficient RdhA architectures are limited only to these classes (and are not present in the γ, ε, δ Proteobacteria). The alignment of the ssRdhA sequences recovered after a PSI-BLAST search was generated using JalView (MAFFT with defaults) [34] and the neighbour-joining dendrogram was generated in MEGA® [38], the Bootstrap values are shown at the nodes. Although there is great diversity in terms eco-physiology of these host organisms, our results only show two distinct clades of mostly marine-dwelling Proteobacteria. We highlighted the ssRdhAs that were selected for heterologous production in this work. *Ottowia thiooxydans* (that encodes the OtssRdhA) is a Burkholderiales organism isolated from activated sludges in Germany and *Tropicibacter phthalicus* (TpssRdhA) is a Rhodobacteriales organism found in Japan's shallow-seas, able to degrade phthalate and dimethylphthalate.



Supporting Material, Figure 2. ssRdhA constructs in pPT7and pEt. Here we present the map of the ssRdhA constructs cloned in different expression vectors for heterologous expression in *B. megaterium* (pPT7) and *E. coli* (pEt systems). The maps were generated using SnapGene ® (SnapGene 5.1, Insightful Science).

Temperature (°C)			
10	○	✓	Low yields
12	◍	✓✓	Good yields/no degradation
15	◍	✓✓✓	Good yields/no degradation
17	◍	✓✓✓	Good yields/no degradation
≥20	●	✓	Aggregation (insoluble bodies)
Media			
LB	◍	✓✓✓	Good yields/no degradation
TB	○	✓	Aggregation (insoluble bodies)
M9	◍	✓✓✓	Good yields/no degradation
Autoinduction	●	✓	Degradation
Strain (E. coli)			
BL21	○	✓✓	Protein degradation
soluBL21	○	✓	Low yields
HMS	◍	✓✓✓	Good yields/no degradation
Artic Express	◍	✓✓✓	Good yields/no degradation
Time (h)			
6	◍	✓✓	Low yields
12	◍	✓✓✓	Good yields/no degradation
16	○	✓✓	Good yields/no degradation
≥24	●	✓	Degradation
Inductor (mM)			
0.25	◍	✓✓✓	Good yields/no degradation
0.5	○	✓✓	Good yields/degradation
1	●	✓	Aggregation (insoluble bodies)

Supporting Material, Figure 3. Table that summarises the expression conditions of the ssRdhA constructs cloned in pEt vectors. Here we present a rough system for the evaluation of the expression, accompanied by observations made during the experiments, describing the quality and the yield of the pure or semi-pure samples after purification. The variables during our experiments were temperature, time, and concentration of the inductor, but *E. coli* offered the additional possibility to test different expression strains and media. In the table we use a semaphore system, where the lightest circles (○) mean a good quality and the darkest ones (●) a lower quality of the samples, implying degradation or low cofactor incorporation. The total yields of soluble protein are represented by (✓). Our experiments suggest that lower temperatures are preferred to avoid aggregation and degradation of the ssRdhAs, although there is a compromise on the yields obtained. Long induction periods (16 h) lead to the aggregation of the protein into inclusion bodies. Lower concentrations of the inductor (IPTG) are preferred, as is the case of other complex and high-molecular-weight proteins. All *E. coli* strains produced the ssRdhAs, but some more specialised strains require media and additional antibiotics to grow adequately. The use of Artic Express® was viable, although it requires gentamycin, in addition to the antibiotic required by the pEt systems used, fact that compromises the obtention of viable colonies and expression. All the media utilised (Formedium) allowed the expression of the ssRdhAs, but when the autoinduction versions of each preparation were used, the protein was found aggregated into insoluble bodies and no cofactor loading was observed. M9 was used in some cases, depending on the strains and, interestingly, this media appears to be useful to produce soluble and fully complemented ssRdhAs, although the yields were variable in each growth. LB was a better media than TB for the expression of ssRdhAs.



Vmax	7.09534 ± 0.18763
Km	19.97117 ± 3.42418
Reduced Chi-Sqr	0.10046
R-Square (COD)	0.9852
Adj. R-Square	0.98274

Supporting Material, Figure 4. Whole-cell biotransformation in *B. megaterium*. A) HPLC chromatogram showing the *in vivo* or whole-cell formation of both products of dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid substrate after an overnight of reaction, at 25°C. Decarboxylation controls of the halogenated substrates and products performed with the heterologously expressed vanillic decarboxylase from *B. megaterium*, provided by Dr. Stephen Marshall, revealed that the di- and mono-brominated species, and specially the 4-hydroxybenzoic acid can be decarboxylated *in vitro*, suggesting that decarboxylation and other secondary reactions might take place *in vivo*, compromising the analysis of the dehalogenation reaction by affecting its detection. B) Steady-state kinetic profile for the dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid, using methyl viologen as a source of electrons and reduction of cytochrome *c* (40 mM).

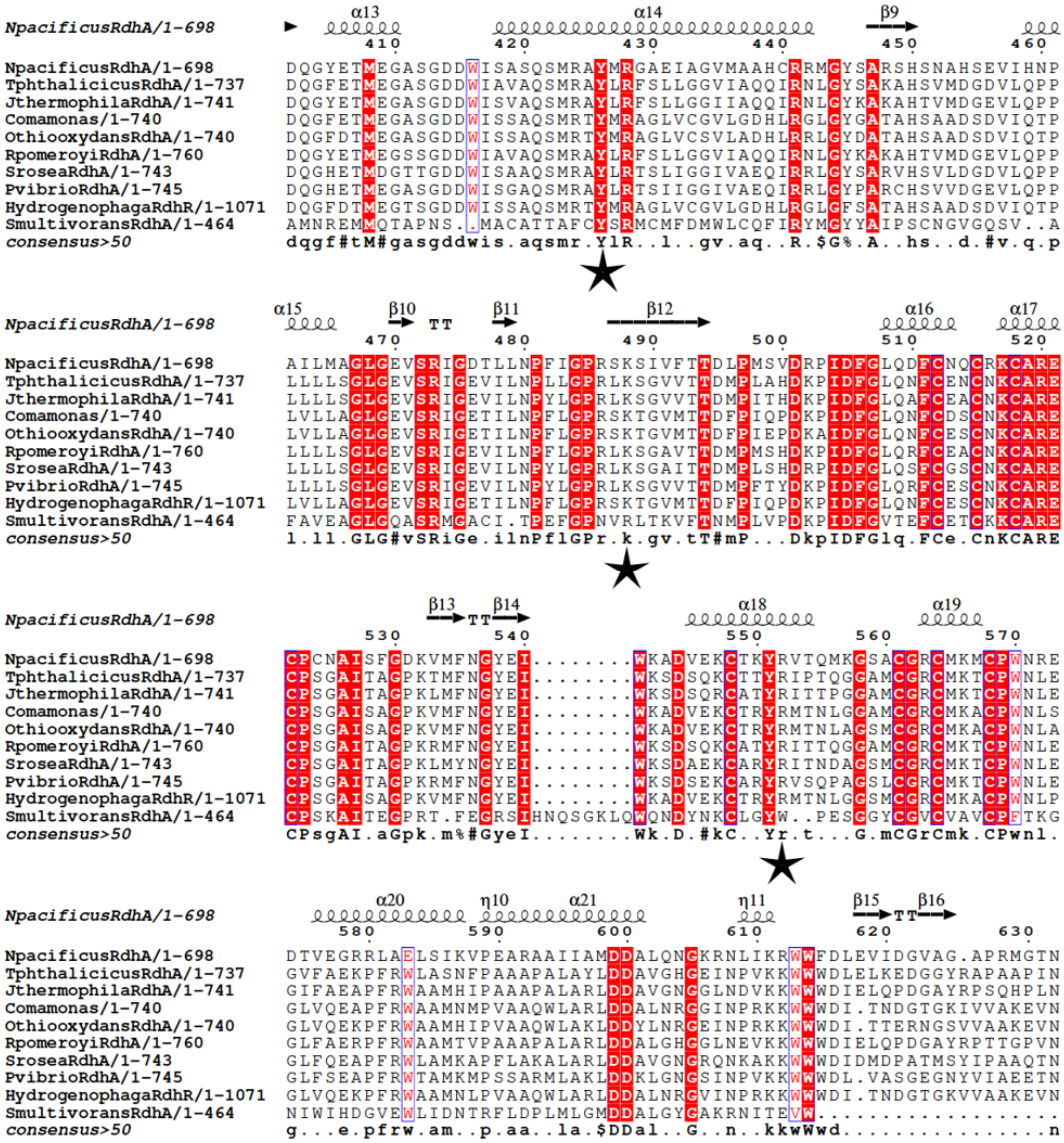
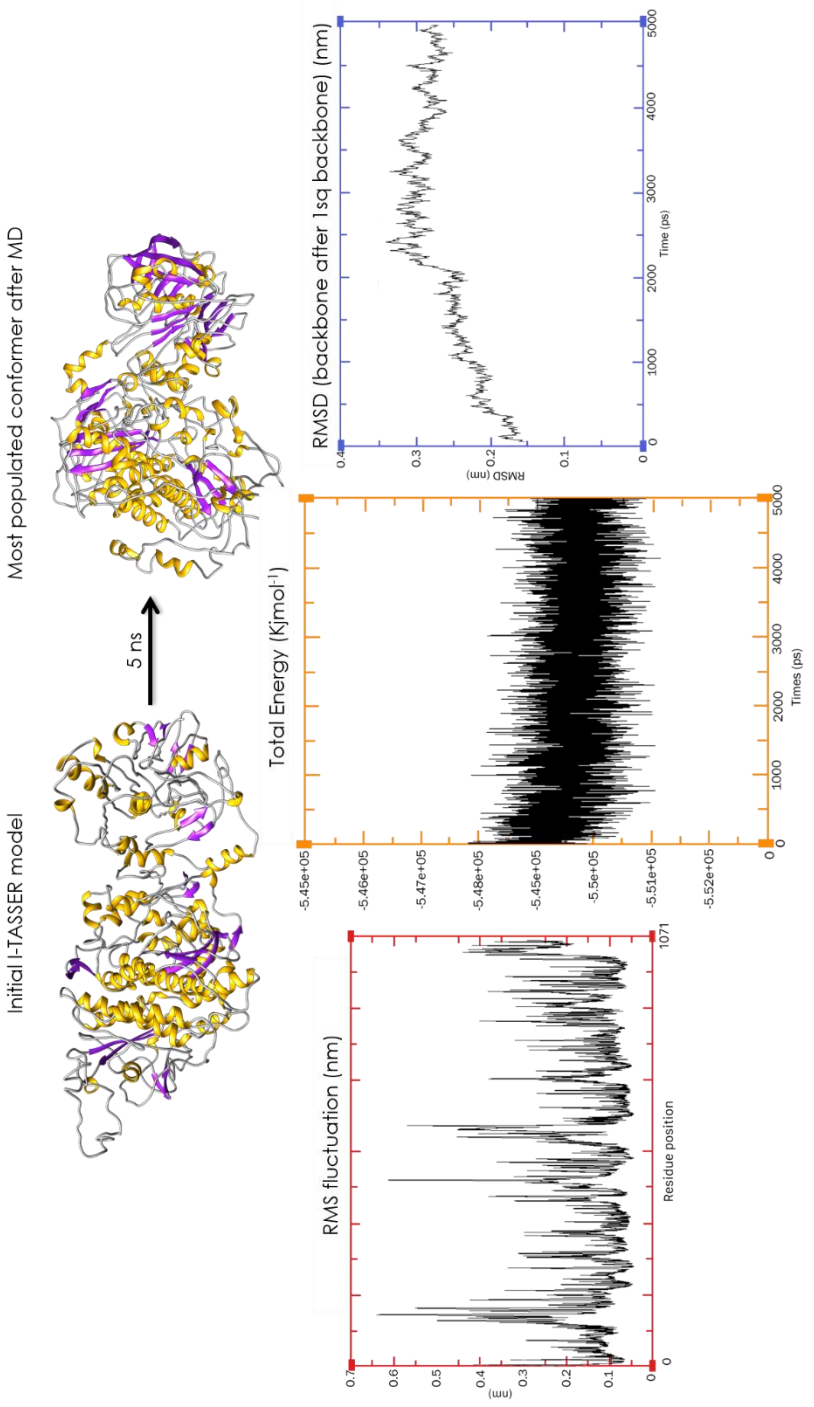


Table 2. EMBOSS Needle alignment OtssRdhR/PceA

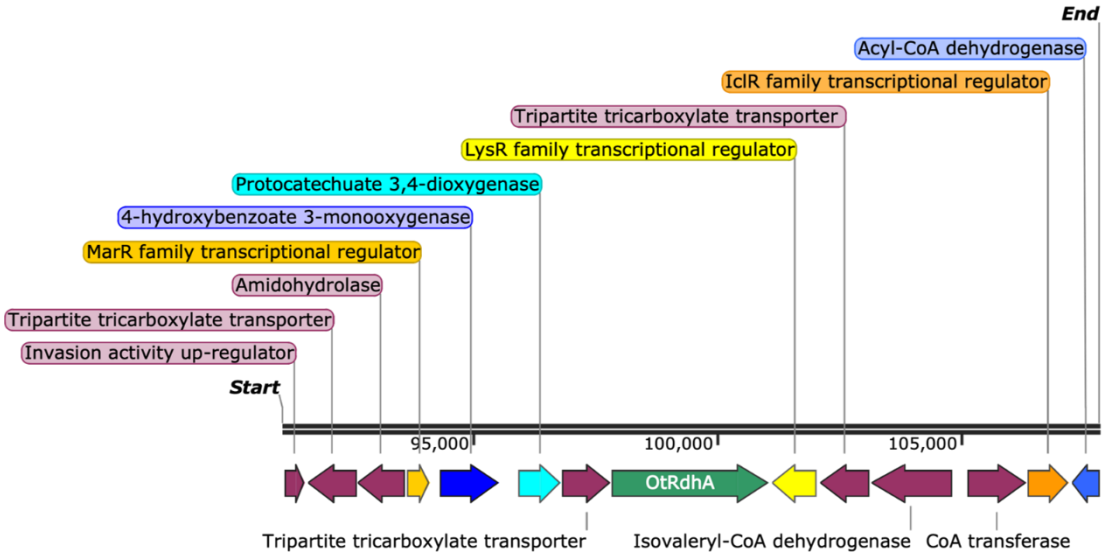
Identity	125/1148	10.9%
Similarity	194/1148	16.9%
Gaps	761/1148	66.3%

Length 1148, Score 271.5, Gap penalty 10, Extended penalty 0.5, Matrix BLOSUM 62

Supporting material, Figure 5. Alignment of a group of ssRdhA orthologous (including NpRdhA) and the respiratory RdhA PceA from *S. multivorans*. The conservation of residues in the central region of the B_{12}/Fe-S domains is highlighted in red, and the presence of the catalytic residues is shown by a star. Although the K has been substituted by R and R by a W residue in PceA, the catalytic Y residue was found in all RdhAs, hinting to its putative role in proton donation during the halogen substitution. The alignment was generated using ESPript 3.0 using the BLOSUM 62 matrix colouring scheme [39].



Supporting material, Figure 6. OtRdhA homology models from I-TASSER and MD simulation. The initial model was minimised using the HyperChem 8.0 software, which allows the correction of severely distorted bond-angles in the initial homology models. We performed a productive run of 5 ns of MD using Gromacs 2018.8 [40] using the Amber-99 force-field to further refine and relax the structure. In the figure, we present the initial I-TASSER model and the most populated conformer at the end of the MD simulation. We monitored the productive run following the total energy of the system, the RMSD and RMSF for the protein backbone, and the fluctuation of each residue, respectively). Our analysis shows that 5 ns of MD have been insufficient for the system (protein + solvent) to converge and more simulation time is required, given the size of the full-length ssRdhA (118 kDa).



Supporting Material, Figure 7. **A cartoon representation of the gene context where the *OtssrdhA* gene is present**. Multiple genes are found in the same *contig* as the ssRdhA. It is unknown if all these enzymes are involved in the same catabolism or organohalides, where the removal of the halogen substituents will allow the reduced products to be used as substrates in oxidative ring-opening reactions until the backbones can be reutilised as carbon sources or as synthetic intermediaries in secondary metabolism. Results generated by Chen *et al.* [14], suggest the putative role of the encoded monooxygenases and dioxygenases in the organohalide degradation pathways. Interestingly, a MarR-type transcription factor (also called a tripartite tricarboxylate transporter) can be seen in the same *contig* as the catabolic *ssrdhA* structural gene, according to the data deposited in the NCBI database for the GenBank entry WP_028602398. Figure generated by SnapGene ®.

4.7 Experimental section

Gene source information and construct design. The *btuB* gene was cloned from *Escherichia coli* K-12 genome and cloned into pLysS by Evelyne Deery (University of Kent) and kindly donated to us. After a BLAST search in the Ref-Seq database (NCBI), the following sequences were selected: *Ottowia thiooxydans* (WP_028602398), *Salinarimonas rosea* (WP_029029911), *Ruegeria pomeroyi* (WP_044027870), *Tropicibacter phthalicicus* (WP_099247216.1), *Pseudovibrio* sp. FO-BEG (WP_014287181) and *Comamonas* sp. 7D-2 (WP_015585978) for functional characterisation. The codon optimised *ssrdhA* genes were synthesised by GeneArt. Plasmid DNA extractions were carried out using the QIAprep Spin Miniprep Kit (QIAGEN), according to the manufacturer's instructions. Standard protocols were used for ligase independent cloning (In-Fusion Clonetech). Coding sequences were amplified by PCR using Q5 HF polymerase (NEB) and specific oligonucleotides designed to

have a Tm of 58 °C. DNA concentration was determined using the Nanodrop 2000c spectrophotometer (Thermo Scientific). After the In-fusion reaction, *E. coli* NEB5α or Stellar (Takara) cells were transformed as described in the standard protocol, and presence of *ssrdhA* insert was confirmed by restriction digest and DNA sequencing (MWG Eurofins). Visualisation was done by agarose gel electrophoresis using a 2 % gel in Tris-acetate-EDTA (TAE buffer), containing SafeView (NBS Biologicals). Constructs containing the *ssrdhA* insert were generated via In-Fusion into pEt28a(+)/pEt32a(+)/pETDuet-1 and pPT7. The ssRdhA constructs in pEt systems were co-transformed into *E. coli* (DE3) strains together with the BtuB-pLysS/pEt3a constructs, following the transformation protocol for DE3 or HMS competent cells. The ssRdhA-pPT7 construct with multiple affinity tags were transformed into *B. megaterium*.

Preparation of protoplasts and transformation of ssRdhA-pPT7 in B. megaterium. Approximately 5 µg of DNA from each ssRdhA-pPT7 construct cloned in *E. coli* Stellar cells (Takara) were mixed with a suspension of 500 µL of *Bacillus megaterium* strain MS941 protoplasts in sterile 15 mL Falcon tubes and mixed with 1.5 mL of an aqueous solution of PEG-6000 40 % were added and mixed carefully with the protoplasts, and then the contents of the tube were incubated for 2 min at 25 °C. 5 mL Hyp-medium [glucose 20 %, KH_2PO_4 0.5 M, yeast extract 10 %, $MnSO_4$ 2 $mgmL^{-1}$, Sucrose 50 % v/w, PEG-6000 40 %, and tetracycline 5 $mgmL^{-1}$] were added and mixed carefully. Cells were harvested by centrifugation at 6 000 rpm for 10 mins at room temperature, the supernatant was decanted immediately after centrifugation. 500 µL of Hyp-medium were added to the rest of the supernatant. Samples were incubated at 30 °C for 90 min with gentle mixing by rolling the tubes. 2.5 mL aliquots of top agar were prepared in sterile tubes. After growth, all cells were added to 2.5 mL top agar and mixed gently by rolling the tube and poured on a pre-warmed plate of LB containing the desired antibiotic. Samples were incubated overnight at 30 °C.

Heterologous expression of the ssRdhA-pPT7 in B. megaterium. The ssRdhA-pPT7 constructs (with N and/or C terminal tags) were transformed into *B. megaterium* MS941 containing the pT7-RNAP plasmid for xylose-inducible expression under the T7 promoter using the modified minimal medium for protoplast transformation protocol [41]. Successful transformation in *B. megaterium* cells was verified by colony PCR, using the CloneAmp HiFi PCR Premix (Clonetech) and the pPT7 Fw and Rv primers. Single colonies of *B. megaterium* DSM319 containing the ssRdhA-pPT7 constructs were inoculated in 250 mL of LB supplemented with tetracycline 10 μgmL^{-1} and chloramphenicol 4.5 μgmL^{-1} and grown overnight in a shaking incubator at 30 °C and 200 rpm. The next morning, 30 flasks

containing 1L of sterile LB supplemented with the same antibiotics were inoculated with 5 mL of the overnight culture. Cultures were grown at 30 °C until an OD $\lambda^{600\ nm} \leq 0.4$ was reached, the temperature was then lowered to 24 °C until an OD $\lambda^{600\ nm} \leq 0.6$-0.7 was reached. The cultures were induced with varying concentrations of xylose (0.25 or 0.50 % w/v). After the addition of filter sterilised solutions of hydroxocobalamin 2 µM and $(NH4)_2FeSO_46H_2O$ 50 µM, the flasks were moved to a shaker at 17 °C and 200 rpm. Cultures were induced overnight. Cells were harvested by centrifugation at 6 000 rpm for 15 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). Pellets were stored at -80 °C.

Heterologous expression in E. coli DE3 strains. *E. coli* BL21 (DE3) (NEB), Artic Express (DE3) (Agilent) and HMS174 (DE3) (Novagen) cells containing the constructs ssRdhA-pEt28(+), ssRdhA-pEt32(+), and ssRdhApETDuet-1 and the respective BtuB-pLysS/BtuB-pEt3a/ssRdhApETDuet-1 were inoculated in 50 mL of LB supplemented with kanamycin 50 μgmL^{-1}/ chloramphenicol 34 μgmL^{-1} (in case of the co-transformant) and kanamycin 50 μgmL^{-1} (for the single transformant) and grown in a 37 °C shaker (200 rpm) overnight. The next morning 24 flasks containing 1 L of sterile media supplemented with the appropriate antibiotics were inoculated with 2 mL of the overnight cultures. Cells were grown at 37 °C (180 rpm) until an OD $\lambda^{280\ nm} \leq 0.5$, then, the temperature was lowered to 24 °C until an OD $\lambda^{280\ nm} \leq 0.7$-0.8 was reached. Filter sterilised solutions of IPTG, hydroxocobalamin, and $(NH4)_2FeSO_46H_2O$ were added to the final concentrations of 0.25 mM, 2 µM, and 50 µM respectively. The temperature then was changed to 12 or 17 °C. Cultures were grown overnight. Cells were harvested by centrifugation at 6 000 rpm for 10 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). Pellets were stored at -80 °C.

Purification by IMAC with gravity flow Ni(II)-NTA and Ni(II)-TDE (Protino). Cell pellets were thawed and resuspended in the working buffer solution [HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM, trehalose 1%]. The buffer solution was filtered and degassed. EDTA-free protease inhibitor cocktail (Roche), lysozyme, DNase I, and RNase (0.01 $mgmL^{-1}$) (Sigma-Aldrich) were added to the suspension and stirred until completely homogenised. Cells were lysed using a cell disruptor at 20 kpsi (Constant Systems). The cell lysate was clarified by centrifugation at 40,000 rpm for 1 h with an Optima CE-80K ultracentrifuge (Beckman Coulter). The supernatant was filtered through a 0.45 µm filter and applied to a Ni (II)-NTA gravity column (Qiagen) or a Ni(II)-TDE Protino (Macherey-Nagel), previously equilibrated. All the purification steps were carried out at 4° C. The columns were washed in two steps with buffer supplemented with 10 mM imidazole and 40 mM imidazole. The protein was eluted with the working buffer supplemented with 250 mM imidazole. In case of Ni(II)-TED

Protino resin, just one wash step with 10 mM imidazole was made and the protein was eluted with the working buffer supplemented with 100 mM imidazole. Samples were subjected to SDS-PAGE analysis and fractions containing the purified protein were pooled. Imidazole was removed using a 10-DG desalting column (Bio-Rad) equilibrated with buffer solution [HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM]. Protein was concentrated as required using a Vivaspin concentrator (Sartorius). Alternatively, for anaerobic purifications, the cell pellets were resuspended in the working buffer solution after degassing and purging it with N_2 [HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM]. The cell extract was passed through a French press at 1-1.5 kPsi under N_2 atmosphere and then centrifuged at 40 000 rpm for 1 h at 4°C using an Optima CE- 80K ultracentrifuge and a Ti45 rotor (Beckman Coulter). All the purification steps were carried out in a Belle anaerobic glovebox, following the same protocol for both Ni (II) resins.

Purification using the ÄKTA Pure Protein Purification System (GE). IMAC and Hydroxyapatite. After clearing the lysates (from *E. coli* or *B. megaterium*) by centrifugation. The supernatant was applied to a Hi-Trap IMAC column (GE), loaded previously with $NiSO_46H_2O$ 100 mM, washed, and pre-equilibrated with the working buffer solution [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM]. After loading the lysate, a linear gradient was made until a final concentration of imidazole 200 mM was reached [Buffer B: HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM, Imidazole 250 mM]. Fractions that contained the protein of interest were collected and concentrated. Samples were subjected to SDS-PAGE analysis and fractions containing the purified protein were recovered. Imidazole was removed using a 10-DG desalting column (Bio-Rad) equilibrated with the working buffer solution. The protein samples were concentrated using Vivaspin concentrators with the appropriate cut-off (Sartorius). The CHT Ceramic Hydroxyapatite Class II (BioRad) was used according to the protocol for globular proteins (II) following the manufacturer's instructions. The column was packed manually using a column suitable for the ÄKTA (flow rate of 300 ch/h). The protein was buffer exchanged to working Buffer A [$NaPO_4$ 5 mM, NaCl 150 mM, pH 6.8]. And the column was then equilibrated with 10 CV of the Buffer A previous to the loading of the sample, a wash with 5 CV was performed before elution using a linear gradient of Buffer B [500 mM $NaPO_4$, pH 6.8]. The protein samples were concentrated again using Vivaspin concentrators with the appropriate cut-off (Sartorius).

Purification using Ion Exchange chromatography. Ion exchange was used as a polishing step. The fractions enriched in the protein of interest were concentrated and then loaded

into the anion exchange columns (Hi-Trap Q HP 5 mL (GE), Resource Q 6 mL (GE) and Mono Q 1 mL (GE)) using a Super-loop or a capillary loop depending on the volume. Buffers were filtered and degassed [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 50 mM] and [Buffer B: HEPES-NaOH 50 mM, pH 7.8, NaCl 1 M].

Purification by size-exclusion chromatography. The protein samples were concentrated using Vivaspin concentrators until reaching a volume of 500 µL, filtrated, and then injected using a capillary loop (2 mL). The buffer was filtered and degassed before use [Buffer: HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM]. The Superdex S-200 10/300 GL analytical column was mounted on the ÄKTA system and pre-equilibrated using 10 CV before the loading of the sample. The run was performed according to the manufacturer's recommendations regarding the flow-rate (0.5 $mLmin^{-1}$) and the system pressure. The samples were recovered in the first elution fractions and concentrated.

SDS-PAGE electrophoresis. Samples were made up of 20 µL with 10 µL SDS 2 x Sample Buffer and mixed before boiling at 100 °C for 5 mins. 10 µL of samples were loaded onto a Precast gel 10 % precast SDS-PAGE gel cassette (BioRad) in addition to 5 µL Page ruler pre-stained protein ladder (NEB). SDS 1 x running buffer and a voltage of 150 V was applied to the gel for a good separation of the protein samples. The gel was stained with Instant Blue (Expedeon).

Western Blot. After SDS-PAGE separation, the gel pas places in water or transfer buffer and then transferred to a nitrocellulose membrane (Trans-Blot Turbo Midi 0.2 µm) using a Trans-Blot Turbo Transfer System (BioRad). The membrane was stained using a Ponceau S solution to evaluate the quality of the transfer and then washed with TBST solution or a commercial blocking solution containing 3% BSA for 1 h. Alternatively, we used the WesternBreeze® Chemioluminscent Kit (Thermo-Fisher). Incubation with the antibody solution proceeded overnight at 4 °C. The antibody dilution was made according to the manufacturer's recommendation. In this case, we used the anti-His_6-peroxidase conjugated with horseradish peroxidase (HRP antibody). The blot was rinsed again doing 5 washes for 5 minutes each with TBST or the commercial washing solution supplemented with the kit. The chemiluminescent substrate (Western Sure from LI-COR) was added according to the manufacturer's recommendation and imaging was done using a C-DiGit-Blot scanner (LI-COR).

Raman scatter spectroscopy. Concentrated samples of the ssRdhA were dried and powdered and deposited on a CaF disc (10 µL spots). A 532 nm inVia Raman laser (Renishaw) at 10 % power and a 1 s pulse was used to run standard and static scans

centered at 1400 cm^{-1}. Baseline correction was performed using an asymmetric least-squares-algorithm developed by Dr. Bethan Mcavan.

UV-Visible spectroscopy and determination of the protein concentration. UV/Visible absorbance spectra were recorded with a Cary Win UV 60 spectrophotometer (Agilent). Protein concentration was determined at $\lambda^{280\ nm}$, using the value for the molar extinction coefficient generated from the primary sequences by the server ProtParam, included in ExPASy (http://www.expasy.org/). Samples were baseline corrected using the working buffer solution and by measuring transmittance 0 %. Scans were recorded at 0.5 ns between 200 – 800 nm. Alternatively, the concentration can be determined using the Protein Assay (Bio-Rad).

Metal estimations. The iron content of the protein samples was estimated with the bathophenanthroline colorimetric assay after acid denaturation. The protein sample was mixed with an equal volume of HCl 2M and heated at 80 ^{0}C for 10 min before the removal of the precipitate by centrifugation at 13 200 rpm. The supernatant was mixed with deionised water to a final volume of 750 µL. Samples were mixed after the addition of 200 µL of saturated $NH_4CH_3CO_2$. Solutions of sodium metabisulfite $Na_2S_2O_5$ (100 mM, 50 µL) and bathophenanthroline (100 mM, 10 µL) were added and the sample was incubated at room temperature for 10 min before measuring the absorbance at 535 nm in a Cary Win UV 60 spectrophotometer (Agilent). Concentrations were determined from a Fe^{2+} standard curve measured from 0 to 50 nmol.

Cobalamin extraction. Cobalamin was extracted from purified protein samples by incubating an aliquot at 80 °C, after the addition of KCN 500 µM. The sample was then centrifuged at 13,000 rpm for 15 mins and the supernatant was removed and scanned in a Cary Win UV 60 spectrophotometer (Agilent).

Methyl viologen spectrophotometric activity assay. The purified ssRdhA samples were buffer exchanged into degassed buffer A [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM] in a glove- box (Belle Technology, UK). A methyl viologen solution 30 mM reduced by titration with sodium dithionite was prepared for the assay. The substrate-dependent oxidation of dithionite reduced methyl viologen was measured at 578 nm (ε = 978 mM^{-1} cm^{-1}). Assays contain increasing concentrations of the substrate and sodium dithionite-reduced methyl viologen (300 µM). The oxidation reaction of methyl viologen proceeded for 5-10 min. Steady-state kinetics were measured using 3,5-dibromo-4-hydroxybenzoic acid as substrates and the data was fitted to a nonlinear fitting (Michaelis-Menten model) using Origin Pro.

NADPH-dependent spectrophotometric activity assay. The purified ssRdhA samples were buffer exchanged into degassed buffer A [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM] in a glove- box (Belle Technology, UK). The reaction mixtures were prepared by adding NADPH 40 mM, 1-10 μM ssRdhA, and varying concentrations of the organohalide substrate sing appropriate cuvettes to minimise the absorbance contribution of the reaction components. The reactions were started by the addition of the enzyme or the substrate, and the change in absorbance of the NADPH was monitored at 340 nm using the kinetic mode of a Cary Win UV 60 spectrophotometer (Agilent).

NADPH-dependent standard activity assay. Purified and concentrated ssRdhA samples were buffer exchanged into degassed buffer A [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 200 mM] under anaerobic conditions in a glove- box (Belle Technology, UK). The reaction mixtures were prepared to a final volume of 500 μL by adding an excess concentration of NADPH (2 mM), or NADH, and the halogenated substrate 3,5-dibromo-4-hydroxybenzoic acid (1 mM), and the reactions were started by the addition of the ssRdhA enzyme (1-10 μM). Both the NADPH and the halogenated substrate were solubilised in the working buffer solution, by addition of NaOH (if necessary, in the case of the 3,5-Br-4-hydroxybenzoic acid, monitoring the changes in the pH). The reactions were set up anaerobically using crimp seal vials that were then transferred to an incubator at 25° C (or orbital shaker at 250 rpm). The reactions were allowed to proceed for different times (15 min to 2 h) and the assays were terminated by the addition of TCA 5%. The samples were then spun down at 13 200 rpm using a Microfuge, to separate the precipitated components. The supernatant was recovered and put into HPLC vials for reverse-phase HPLC analysis.

Whole-cell biotransformations in B. megaterium. Approximately 500 mL of expression media were inoculated with 2.5 mL of an over-night culture of ssRdhA-pPT7 or ssRdhA-pEt transformed cells. The cultures were grown at 30 °C or 37 °C respectively at 200 rpm until an OD $\lambda^{600\ nm} \leq 0.4$ was reached, the temperature was then lowered to 24 °C until an OD $\lambda^{600\ nm} \leq 0.6$-0.7 was attained, and then the cultures were induced with xylose (0.50 % w/v) or IPTG (0.25 mM). After the addition of filter sterilised solutions of hydroxocobalamin 2 μM and $(NH4)_2FeSO_46H_2O$ 50 μM, the flasks were moved to a shaker at the appropriate induction temperature (12 - 17 °C) and agitation (200 rpm). The induction proceeded overnight. Cells were harvested by centrifugation at 6 000 rpm for 15 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). The cell pellets were then washed with the working buffer solution [HEPES-NaOH 50 mM, pH 7.8, without NaCl], spun down, and then resuspended in 5 mL of the same buffer to obtain an OD $\lambda^{600\ nm} \leq 300$. The

suspension was placed on ice and diluted ten times for the assays, using the buffer solution. Reaction mixtures (final volume 500 µL) were prepared by adding glucose 50 mM and the halogenated substrate at an excess concentration (5 to 10 mM). The reaction proceeded at 25 °C in an orbital shaker for fixed times (12 h). The reaction was stopped by the addition of acetonitrile (2:1) with TCA 5 %. The samples were mixed thoroughly and then centrifuged at 13 200 rpm; the supernatant was filtered transferred to HPLC vials for analysis.

Reverse-phase HPLC. Sample analysis for product formation *in vitro* and *in vivo* (whole-cell biotransformation) was performed using an Agilent System 1110 Series, equipped G1379A degasser, G1312A binary-pump, a G1367A well plate autosampler unit, a G1316A temperature-controlled column compartment, and a single-wavelength UV detector. The stationary phases were Kinetex C-18 (250 mm length, 4.6 mm diameter, 5mm particle size (Phenomenex)), and Syncronis C-18 (250 mm length, 4.6 mm diameter, 5mm particle size (Thermo Scientific)). The mobile phase was made with HPCL grade acetonitrile/water (50:50) with 0.1 % TFA as additive. Substrate and product standards were prepared in acetonitrile. The flow rate was maintained at 1 $mLmin^{-1}$. The detector was set at 245 nm.

EPR spectroscopy. Samples of the ssRdhA were prepared as isolated or reduced and transferred in volumes of 300 µL into 4 mm Suprasil quartz EPR tubes (Wilmad) under anaerobic conditions. The tubes were anaerobically sealed and frozen and stored in liquid nitrogen. Experimental parameters were as follows: microwave power 0.5 mW, modulation frequency 100 kHz, modulation amplitude 5 G, temperature 30 K. Spectra were obtained using a Bruker ELEXSYS E500 spectrometer, Super high Q resonator (ER4122SHQ), Oxford Instruments ESR900 Cryostat, and an ITC503 temperature controller.

In silico methodology. I-TASSER is a program used for protein structure prediction. The initial ssRdhA generated by I-TASSER models were improved by a cycle of energy minimisations using the Amber 99 force-field in Hyperchem 8.0 (Hypercube, Inc.), which employs the internal topologies of the model to assign the correct connectivity. The output models were then minimised using a slow procedure, to avoid distortion of the original fold. These new models were then scored using the Rosetta design-HMMer (Rd.HMM) [42]. After that, we performed a molecular dynamic simulation (MD) for 5 ns using GROMACS 2018.8 [40], in an octahedral box with 1.5 nm of distance to the wall, using water as the solvent and a fixed concentration of 100 mM of NaCl at 303 K (using a coupled Berendsen thermostat/barostat for constant P and T). We used the *leap-frog* algorithm for the integration of the classic movement equations during the MD simulation. Initially, we

performed a restricted simulation where the backbone and side chains of the protein were fixed to certain coordinates inside the box by the application of a harmonic force of 1000 N to favour the approximation of solvent molecules to the protein to achieve a solvation layer; this simulation continued for 200 ps and the neighbours list was updated every 20 fs. After that, we performed a productive simulation during 5 ns without restrictions. The most populated conformer at the end of the simulation was obtained by a clustering algorithm according to its occurrence and then minimised again for 5000 steps.

Bioinformatics. After a PSI-BLAST search (Position-Specific Iterated Basic Local Alignment Search Tool) against the NCBI (Ref-Seq) database (ncbi.nlm.nih.gov), using the translated sequence of the BhbA gene that encodes the ssRdhA from *Comamonas* sp. 7D-2 (WP_015585978) we obtained significative alignments of approximately 100 sequences after 10 iterations. The compositional matrix used was BLOSUM62, with the following algorithm parameters: Gap 10, Existence 11, Extension 1, and a PSI-BLAST threshold of 0.001. Sequence selection was done considering the quantitative values of similitude and identity generated during the alignment, that indicate the conservation or presence of identical sites on the aligned sequences, and the similarity of the sequences depending on the physicochemical properties of the aa residues, respectively. We recovered the sequences that presented an identity of more than 15% with respect to the query, and presented E values tending to 0 in FASTA format, eliminating the sequences from the same genera before aligning them using JalView (MAFFT with defaults). The NJ-phylogenetic dendrogram was generated in MEGA® [34, 38].

4.8 References

1. P. Adriaens and A. L. Barkovskii, in *Encyclopedia of Environmental Microbiology*, ed. G. Bitton, Wiley-Interscience, New York, 2003, p. 509.
2. S. Atashgahi, Y. Lu, Y. Zheng, E. Saccenti, M. Suarez-Diez, J. Ramiro-Garcia, H. Eisenmann, M. Elsner, A. J. M. Stams, D. Springael, W. Dejonghe and H. Smidt, *Environ. Microbiol.*, 2017, **19**, pp. 968–981.
3. F. Maphosa, S. H. Lieten, I. Dinkla, A. J. Stams, H. Smidt and D. E. Fennell, *Front. Microbiol.*, 2012, **3**. 351.
4. G. Öberg, *Applied Microbiology and Biotechnolgy*, 2002, **58**, pp. 565-581.
5. G. W. Gribble, *Chemosphere*, 2003, **52**, pp. 289-297.
6. M. Fincker and A. M. Spormann, *Annu. Rev. Biochem.*, 2017, **86**, pp. 357–386.
7. B.-E. Jugder, H. Ertan, M. Lee, M. Manefield and C. P. Marquis, *Trends Biotechnol.*, 2015, **33**, pp. 595-610.
8. K. A. P. Payne, C. P. Quezada, K. Fisher, M. S. Dunstan, F. A. Collins, H. Sjuts, C. Levy, S. Hay, S. E. J. Rigby and D. Leys, *Nature*, 2014, **517**, pp. 513-516.
9. T. Futagami, M. Goto and K. Furukawa, *Chem. Rec.*, 2008, **8**, pp. 1-12.

10. M. Bommer, C. Kunze, J. Fesseler, T. Schubert, G. Diekert and H. Dobbek, *Science*, 2014, **346**, pp. 455-458.
11. A. Kublik, D. Deobald, S. Hartwig, C. L. Schiffmann, A. Andrades, M. v. Bergen, R. G. Sawers and L. Adrian, *Environ. Microbiol.*, 2016, **18**, pp. 3044–3056.
12. A. Neumann, G. Wohlfarth and G. Diekert, *J. Biol. Chem.*, 1996, **271**, pp. 16515-16519.
13. J. Maillard, W. Schumacher, F. Vazquez, C. Regeard, W. R. Hagen and C. Holliger, *Appl Environ Microbiol.*, 2003, **69**, pp. 4628–4638.
14. K. Chen, L. Huang, C. Xu, X. Liu, J. He, S. H. Zinder, S. Li and J. Jiang, *Mol. Microbiol.*, 2013, **89**, pp. 1121–1139.
15. K. Chen, Y. Mu, S. Jian, X. Zang, Q. Chen, W. Jia, Z. Ke, Y. Gao and J. Jiang, *Appl Environ Microbiol.*, 2018, **84**. e02467-17.
16. A. Parthasarathy, T. A. Stich, S. T. Lohner, A. Lesnefsky, R. D. Britt and A. M. Spormann, *J. Am. Chem. Soc.*, 2015, **137**, p. 3525−3532.
17. B.-E. Jugder, K. A. P. Payne, K. Fisher, S. Bohl, H. Lebhar, M. Manefield, M. Lee, D. Leys and C. P. Marquis, *ACS Chem. Biol.*, 2018, **13**, pp. 548-552.
18. D. Leys, L. Adrian and H. Smidt, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120316.
19. T. Schubert, L. Adrian, R. G. Sawers and G. Diekert, *FEMS Microbiol. Ecol.*, 2018, **94**. fiy035.
20. F. A. Collins, K. Fisher, K. A. P. Payne, S. Gaytan, S. E. J. Rigby and D. Leys, *Biochemistry*, 2018, **57**, p. 3493−3502.
21. D. D. Shultis, M. D. Purdy, C. N. Banchs and M. C. Wiener, *Science*, 2006, **312**, pp. 1396-1399.
22. N. Noinaj, M. Guillier, T. J. Barnard and S. K. Buchanan, *Annu. Rev. Microbiol.*, 2010, **64**, pp. 43-60.
23. A. Suyama, M. Yamashita, S. Yoshino and K. Furukawa, *J. Bacteriol.*, 2002, **184**, pp. 3419-3425.
24. H. Sjuts, K. Fisher, M. S. Dunstan, S. E. Rigby and D. Leys, *Protein Expr. Purif.*, 2012, **85**, pp. 224-229.
25. F. A. Collins, PhD thesis, The University of Manchester, 2017.
26. S. Costa, A. Almeida, A. Castro and L. Domingues, *Front. Microbiol.*, 2014, **5**, pp. 1-20. 63.
27. K. A. P. Payne, K. Fisher, H. Sjuts, M. S. Dunstan, B. Bellina, L. Johannissen, P. Barran, S. Hay, S. E. J. Rigby and D. Leys, *J. Biol. Chem.*, 2015, **290**, pp. 27572-27581.
28. R. Banerjee and S. W. Ragsdale, *Annu. Rev. Biochem.*, 2003, **72**, pp. 209-247.
29. T. F. Bidleman, A. Andersson, L. M. Jantunen, J. R. Kucklick, H. Kylin, R. J. Letcher, M. Tysklind and F. Wong, *Emerg. Contam.*, 2109, **5**, pp. 80-115.
30. C. Kunze, G. Diekert and T. Schubert, *FEBS J*, 2017, **284**, pp. 3520-3535.
31. A. Roy, A. Kucukural and Y. Zhang, *Nature Protocols*, 2010, **5**, pp. 725-738.
32. A. Drozdetskiy, C. Cole, J. Procter and G. J. Barton, *Nucleic Acids Research*, 2015, **43**, pp. W389-W394.
33. H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, *Nucleic Acids Research*, 2016, **44**, pp. W344-W350.
34. A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton, *Bioinformatics*, 2009, **25**, pp. 1189-1191.
35. D. P. Dowling, Z. D. Miles, C. Köhrer, S. J. Maiocco, S. J. Elliott, V. Bandarian and C. L. Drennan, *Nucleic Acids Res.*, 2016, **44**, pp. 9965–9976.
36. H. Fang, D. Li, J. Kang, P. Jiang, J. Sun and D. Zhang, *Nat. Commun.*, 2018, **9**. 4917(2018).
37. J. Liu and M. M. Häggbloma, *Mbio*, 2018, **9**. e02471-18.
38. S. Kumar, G. Stecher, M. Li, C. Knyaz and K. Tamura, *Mol Biol Evol*, 2018, **35**, pp. 1547-1549.
39. X. Robert and P. Gouet, *Nucleic Acids Research*, 2014, **42**, pp. W320–W324.
40. D. V. d. Spoel, E. Lindhal, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *Journal of computational Chemistry*, 2005, **26**, pp. 1701-1718.

41. S. J. Moore, M. J. Mayer, R. Biedendieck, E. Deery and M. J. Warren, *N. Biotechno.l*, 2014, **31**, pp. 553-561.
42. L. P. Martínez-Castilla and R. Rodríguez-Sotres, *PLoS One*, 2010, **5**. e12483.

5.0 Chapter Five. Results Chapter III.

Designing an artificial self-sufficient catabolic reductive dehalogenase

Samantha Gaytan[1], Karl Payne[1], Godwin Aleku[1], and David Leys

[1] Manchester Institute of Biotechnology, School of Chemistry, Faculty of Science and Engineering, The University of Manchester, Manchester M1 7DN, United Kingdom

[2] Department of Biochemistry, Sanger Building, University of Cambridge, 80 Tennis Court Road, Cambridge, CB2 1GA, United Kingdom

* To whom correspondence may be addressed: Manchester Institute of Biotechnology, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK. Tel.: +44-161-306-5150; E-mail: david.leys@manchester.ac.uk

Keywords: reductive dehalogenation, cobalamin, BtuB transporter, heterologous expression, pyridine nucleotides, artificial self-sufficient dehalogenase

This manuscript has not been submitted to any scientific journal.

5.1 Abstract

A subfamily of B_{12}-dependent oxidoreductases called reductive dehalogenases (RdhAs) use organohalides as terminal electron acceptors during bacterial organohalide respiration. However, not all RdhAs are involved in energy conservation pathways and a subset participates instead in degradation pathways. These catabolic RdhAs are often found fused to a terminal reductase domain like the flavoprotein phthalate dioxygenase reductase (PDR). Unfortunately, efficient heterologous expression and functional

characterisation of such self-sufficient catabolic RdhAs has proven to be difficult. Here we present the design and heterologous expression of an artificial self-sufficient RdhA-PDR fusion, based on the well characterised NpRdhA from *Nitratireductor pacificus* pht-3B, and a PDR-like reductase module. However, our results show that although expressed as a soluble holo-enzyme, the artificial NpRdhA-PDR could not catalyse the dehalogenation of the 3,5-bromo-4-hydroxybenzoic acid. To test both the linkage and the activity of the independent domains, we generated an individual PDR-like module, demonstrating that the reductase mediates the electron transfer process from NADPH to NpRdhA, thus supporting the dehalogenation reaction, even when truncated from the original construct.

5.2 Introduction

While organohalides are produced as part of the natural biogeochemical halogen cycle, their numerous industrial, agricultural, and household applications, have led to the (inadvertent) release of anthropogenic compounds. These now constitute a large proportion of the most recalcitrant environmental pollutants, that accumulate in water systems and soils [1, 2]. Several anaerobic microorganisms known as organohalide respiring bacteria (OHRB) use organohalides as terminal electron acceptors [3]. OHRBs can be further subdivided into obligate and facultative organisms; the first category comprises only strict anaerobes (mainly from the Firmicutes and Chloroflexi phyla) that use organohalides as terminal electron acceptors during respiration [4], whereas the facultative and aerobic organisms, most of them from the phylum Proteobacteria, have a diverse respiratory metabolism and organohalides can be used as carbon sources rather than terminal electron acceptors [4, 5]. Research of the pathways that involve the dehalogenation of organohalides by OHRBs might help us guide bioremediation strategies. These will provide a biochemical solution, instead of physical and chemical processes that are expensive and inefficient, as contamination of soils and aquifers by persistent organic pollutants (POPs) is now one of the major challenges for sustainable development.

The OHRB dehalogenation metabolism depends on the enzymes called reductive dehalogenases (RdhAs), a subfamily of B_{12}-dependent enzymes classified as oxidoreductases (EC 1.97.1.8) [3, 6]. Although little is known about the mechanism of the dehalogenation, it is thought that the cobalamin chemistry in the RdhAs differs from that of other B_{12}-dependent enzymes [7-9]. Despite this, and the knowledge generated

by different research groups worldwide over almost 30 years, there are still key questions that remain unanswered in the field, and current advances are hindered by the lack of reliable RdhA sources and an ill-understood eco-physiological role of these enzymes and their host organisms.

Reductive dehalogenation was thought to be uncommon in aerobes and facultative anaerobic bacteria until Chen *et al.* [10] reported the characterisation of *Comamonas* sp. 7D-2, a β-Proteobacteria strain that was shown to degrade the herbicide bromoxynil under aerobic conditions. This organism relies on the activity of the reductive dehalogenase BhbA, that presents a C-terminal reductase similar to the phthalate dioxygenase reductase (PDR) in addition to the typical B_{12}-binding and Fe-S domains [11]. It is suspected that this module might be involved in the transfer of electrons intramolecularly from pyridine nucleotides like NADP(P)H. Hence, these enzymes act in a self-sufficient way, without the need for external reductase systems to provide electrons during the catalysis, and are thus referred to as self-sufficient architectures (ssRdhA). Gene fusion events generate chimeric proteins with new activities that with time can be selected, allowing the evolution of new phenotypes [12], as might have been the case of the ssRdhAs.

Our group has contributed to the structural and complete functional characterisation of the catabolic NpRdhA, which lacks the additional PDR-like reductase module [8]. Genomic analysis revealed that the *NprdhA* structural gene is clustered in an operon together with putative redox partners, such as NAD(P)H-dependent oxidoreductases plus some associated [4Fe-4S] and [2Fe-2S] ferredoxins, hinting to a specific pathway for channeling electrons to the active site's cobalamin, as opposed to the multiple and not well-defined electron sources for the respiratory dehalogenases [13]. Recently, Collins *et al.* [14] showed that NpRdhA requires the presence of redox partners to support the NADPH-dependent reduction of 3,5-dibromo-4-hydroxybenzoic acid **(Supplementary Figure 2)**. However, efficient organohalide reduction was only observed when using a non-physiological reducing system composed of the *Spinacia oleracea* ferredoxin (SpFd) and *Escherichia coli* flavodoxin reductase (EcFldr), that could be used in the future for bioremediation of heavily polluted sites.

Preliminary results obtained in our group with a set of self-sufficient RdhA orthologous to BhbA from *Comamonas* sp. 7D-2 show that these proteins depend on NAD(P)H to catalyse the reduction of organohalides, as shown previously by Chen *et al.* [10], linking the reduction of organohalides to the NAD(P)H/NAD(P)$^+$ pool in the cell

(Results Chapter II). Given the experimental difficulties found when heterologously expressing the self-sufficient RdhAs-PDR in both *B. megaterium* and *E. coli*, we decided to generate an artificial self-sufficient RdhA based on the already well characterised NpRdhA [8] by fusing it to the reductase module of a ssRdhA.

In principle, the artificial NpRdhA-PDR construct mimics the natural full-length enzymes but it could constitute a more tractable system than the natural self-sufficient architectures, if it retains the high solubility and cofactor incorporation of the non-self-sufficient NpRdhA. Additionally, to the role of the fusion linkage, we tested if the isolated PDR-like reductase domain could mediate the electron transfer from pyridine nucleotides to the B_{12}/Fe-S typical dehalogenase domain of the catabolic RdhAs. This will allow the development of an alternative assay to that established by Collins *et al.* [14] when using the non-physiological redox partners.

5.3 Experimental section

Gene source information and construct design. The *Nitratireductor pacificus* pht-3b genome contains the three putative reductive dehalogenases homologous to BhbA [15], including the *NprdhA* gene (WP_008597722.1), that lacks the PDR-like reductase domain. The gene was codon optimised for both *E. coli* and *B. megaterium* and synthesised by Genscript®. The *E. coli* K-12 (*btuB*) gene was donated as a *btuB*-pLysS construct by Evelyne Deery (University of Kent). Constructs containing the NpRdhA fused to the PDR (NpRdhA-PDR) were cloned via three-fragment In-Fusion into pEt30(+) and pPT7, and the PDR gene was generated by truncating the ssOtRdhA from *Ottowia thiooxydans* (WP_028602398) from the 740 aa and later was cloned into pEt28a. Standard protocols were used for ligase independent cloning (In-Fusion Clonetech ®). Insertion and the correct orientation and reading frame of the gene was verified also by DNA sequencing (MWG Eurofins). Plasmid DNA extractions were carried out using the QIAprep Spin Miniprep Kit (QIAGEN), following the manufacturer's instructions. The DNA concentration was determined using the Nanodrop 2000c spectrophotometer (Thermo-Scientific) and the insertion of the genes was confirmed by restriction digest and DNA sequencing (MWG Eurofins). Visualisation was done by agarose gel electrophoresis using a 2 % gel in Tris-acetate-EDTA (TAE buffer), containing SafeView (NBS Biologicals). The NpRdhA-PDR fusion constructs in pEt30a(+) were co-transformed into *E. coli* (DE3) strains together with the BtuB-pLysS construct,

following the NEB transformation protocol for DE3 competent cells. And the NpPvRdhA-pPT7 construct, together with NpRdhA-pPT7 were transformed into *B. megaterium*. The PDR-pEt28(+) construct was transformed into *E. coli* (DE3) strains.

Heterologous expression of the NpRdhA-PDR fusion and the PDR-like reductase module in E. coli DE3 strains. *E. coli* DE3 expression strains BL21 (NEB) and HMS174 (DE3) (Novagen) cells containing the constructs NpRdhA-PDR-pEt30(+)a/BtuB-pLysS and PDR-pEt28(+)a were inoculated in 50 mL of LB supplemented with kanamycin 50 μgmL^{-1} and chloramphenicol 34 μgmL^{-1} (in case of the co-transformant) and kanamycin 50 μgmL^{-1} (for the single transformant PDR), respectively and grown in a 37 °C shaker (200 rpm) overnight. The next morning 24 flasks containing 1 L of sterile LB supplemented with the appropriate antibiotics were inoculated with 2 mL of the overnight cultures. Cells were grown at 37 °C (180 rpm) until OD $\lambda^{600\ nm} \leq$ 0.7-0.8 was reached. Filter sterilised solutions of IPTG, hydroxocobalamin, and $(NH4)_2FeSO_46H_2O$ were added to the final concentrations of 0.25 mM, 2 µM, and 50 µM respectively. The temperature then was changed to 17 °C (or 20°C in case of the PDR truncation) and the cultures were grown overnight. Cells were harvested by centrifugation at 6 000 rpm for 10 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). The pellets were stored at -80 °C or used immediately.

Heterologous expression of the NpRdhA-pPT7 and NpPvRdhA-pPT7 fusion constructs in B. megaterium. The NpRdhA-pPT7 and NpRdhA-PDR-pPT7 fusion constructs were transformed into *B. megaterium* MS941 containing the pT7-RNAP plasmid for xylose-inducible expression under the T7 promoter using the modified minimal medium for protoplast transformation protocol **[16]**. Successful transformation in *B. megaterium* cells was verified by colony PCR, using the CloneAmp HiFi PCR Premix (Clonetech) and the pPT7 Fw and Rv primers. Single colonies of *B. megaterium* DSM319 containing the RdhA-pPT7 constructs were inoculated in 250 mL of LB supplemented with tetracycline 10 μgmL^{-1} and chloramphenicol 4.5 μgmL^{-1} and grown overnight in a shaking incubator at 30 °C and 200 rpm. The next morning, 30 flasks containing 1L of sterile LB supplemented with the same antibiotics were inoculated with 5 mL of the overnight culture. Cultures were grown at 30 °C until an OD $\lambda^{600\ nm} \leq$ 0.6-0.7 was reached. The cultures were induced with xylose (0.50 % w/v). After the addition of filter sterilised solutions of hydroxocobalamin 2 µM and $(NH4)_2FeSO_46H_2O$ 50 µM, the temperature was lowered to 17 °C and 180 rpm. Cultures were induced overnight. Cells

were harvested by centrifugation at 6 000 rpm for 15 minutes (Beckman Coulter Avanti J-26XP centrifuge with a JL8.1000 rotor). Pellets were stored at -80 °C.

Purification via IMAC by gravity flow using Ni(II)-NTA and Ni(II)-TDE (Protino). Cell pellets were thawed and resuspended in the working buffer solution [HEPES-NaOH 50 mM, pH 7.8), NaCl 250 mM]. The buffer solution was filtered and degassed. EDTA-free protease inhibitor cocktail (Roche), lysozyme, DNase I, and RNase (0.01 $mgmL^{-1}$) (Sigma-Aldrich) were added to the suspension and stirred until completely homogenised. Cells were lysed using a cell disruptor at 20 kpsi (Constant Systems). The cell lysate was clarified by centrifugation at 40,000 rpm for 1 h with an Optima CE-80K ultracentrifuge (Beckman Coulter). The supernatant was filtered through a 0.45 µm filter and applied to previously equilibrated gravity-flow columns, Ni (II)-NTA (Qiagen) was used in case of the artificial NpRdhA-PDR, but the Ni(II)-TDE Protino (Macherey-Nagel) resin was preferred for the truncated PDR. All the purification steps were carried out at 4° C. The Ni-NTA column was washed in two steps with buffer supplemented with 10 mM imidazole and 40 mM imidazole and the protein was eluted with the working buffer supplemented with 250 mM imidazole. In case of Ni(II)-TED Protino resin, just one wash step with 10 mM imidazole was made and the protein was eluted with the working buffer supplemented with 100 mM imidazole. Samples were subjected to SDS-PAGE analysis and fractions containing the purified protein were pooled. Imidazole was removed using a 10-DG desalting column (Bio-Rad) equilibrated with buffer solution [HEPES-NaOH 50 mM, pH 7.8, NaCl 250 mM]. Protein was concentrated as required using a Vivaspin concentrator (Sartorius).

Purification by ion-exchange chromatography using the ÄKTA Pure Protein Purification System (GE). The fractions enriched with the protein of interest were concentrated, buffer exchanged, and then loaded into the anion exchange columns (Hi-Trap Q HP 5 mL (GE) or Resource Q 6 mL (GE)) using a Super-loop or a capillary loop depending on the volume of the sample immediately after the first step of IMAC was performed. To adequately perform the purification the NaCl concentration of the protein samples was matched to that of the working buffer solution. Buffers were filtered and degassed before its use [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 50 mM] and [Buffer B: HEPES-NaOH 50 mM, pH 7.8, NaCl 1 M]. The runs were performed according to the column manufacturer's recommendations for the ÄKTA Purifier.

Purification size-exclusion chromatography using the ÄKTA Pure Protein Purification System (GE). The protein samples were concentrated using Vivaspin concentrators until 500 µL, filtrated, and then injected using a capillary loop (2 mL). The buffer was filtered and degassed before use [Buffer: HEPES-NaOH 50 mM, pH 7.8, NaCl 250 mM]. The gel-filtration using the Superdex S-200 10/300 GL was performed according to the manufacturer's recommendations.

SDS-PAGE electrophoresis. Samples were made up of 20 µL using the appropriate volume of Sample Buffer 2x and mixed before boiling at 100 °C for 5 mins. 10 µL of samples were loaded onto a Precast gel 10 % precast SDS-PAGE gel cassette (BioRad) in addition to 5 µL Page ruler prestained protein ladder (NEB). SDS 1 x running buffer and a voltage of 150 V was applied to the gel for a good separation of the protein samples. The gel was stained with Instant Blue (Expedeon).

UV-Visible spectroscopy and determination of the protein concentration. Protein concentration was determined at $\lambda^{280\ nm}$, using the value for the molar extinction coefficient generated from the primary sequences by the server ProtParam, included in ExPASy (http://www.expasy.org/). Samples were baseline corrected using the working buffer solution and by measuring transmittance 0 %. Scans were recorded at 0.5 ns between 200 – 800 nm using a Cary Win UV 60 spectrophotometer (Agilent).

Metal estimations and cobalamin extraction. The Fe content of the protein samples was estimated with the bathophenanthroline colourimetric assay after acid denaturation after mixing the sample with an equal volume of HCl 2M and heated at 80 ^{0}C for 10 min before the removal of the precipitate by centrifugation at 13 000 rpm. The supernatant was mixed with deionised water to a final volume of 750 µL. Samples were mixed after the addition of 200 µL of saturated $NH_4CH_3CO_2$. Solutions of sodium metabisulfite $Na_2S_2O_5$ (100 mM, 50 µL) and bathophenanthroline (100 mM, 10 µL) were added and the sample was incubated at room temperature for 10 min before measuring the absorbance at 535 nm in a Cary Win UV 60 spectrophotometer (Agilent). Concentrations were determined from a Fe^{2+} standard curve measured from 0 to 50 nmol. Cobalamin was extracted from purified protein samples by incubating an aliquot at 80 ° C, after the addition of KCN 500 µM. The sample was then centrifuged at 13,000 rpm for 15 mins and the supernatant was removed and scanned in a Cary Win UV 60 spectrophotometer (Agilent).

Methyl viologen spectrophotometric activity assay. Enzymatic assays were performed with a Cary UV-Visible spectrophotometer, under anaerobic conditions inside a glove-box (Belle Technology) purged with N_2(g) or Ar(g) atmosphere. The working buffer solutions were rendered anaerobic by flushing them with N_2 for at least 1h and additional reagents (weighed solids such as sodium dithionite or NAD(P)H were dissolved inside the box to guarantee the absence O_2 by diminishing its hygroscopic behaviour) and left to equilibrate over-night inside the box, if possible. The protein samples were buffer exchanged using desalting columns previously equilibrated with anaerobic buffer (10-DG from BioRad® or 0.3 to 1.5 mL from Zeba®). The samples were baseline corrected using the working buffer solution (and by measuring the 0% transmittance, if required). For methyl viologen assays reaction mixtures were prepared as follows:

Methyl viologen assays	
Methyl viologen (30 mM)	10 µL
Sodium dithionite (100 mM)	0.5 µL
Substate (20 mM)	20 µL
Enzyme RdhA (1-5 $mgmL^{-1}$)	x µL
Final volume (by addition of buffer)	1000 µL

The substrate-dependent oxidation of sodium dithionite-reduced methyl viologen was monitored at a single-wavelength (ε = 978 $mM^{-1}cm^{-1}$) for 5 to 15 mins. The data-analysis proceeded according to the Lambert-Beer's Law, solving for the delta or change of concentration ($\Delta C min^{-1}$), which gives the amount of product formed per minute ($Mmin^{-1}$). These units are then transformed into $\mu Mmin^{-1}$. This value is then divided by the total mg of RdhA enzymes used in the assay to determine specific activity ($\mu Mmin^{-1}mg^{-1}$).

$$\text{Lambert} - \text{Beer's Law:}$$

$$A = \varepsilon l C$$

A = Absorbance ($\Delta A min^{-1}$), ε = Molar extinction coefficient of methyl viologen (978 $mM^{-1}cm^{-1}$), l = Path-length of cuvette, c = Concentration of substrate ($\Delta C min^{-1}$).

NADPH-dependent activity assay. Purified and concentrated samples of both the NpRdhA-PDR protein and the individual domains NpRdhA and PDR reductase were buffer exchanged into degassed buffer A [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 250 mM] under anaerobic conditions in a glove-box (Belle Technology, UK). The reaction mixtures were prepared to a final volume of 500 µL by adding an excess

concentration of NADPH (2 mM), or NADH, and the halogenated substrate 3,5-dibromo-4-hydroxybenzoic acid (1 mM). The reactions were started by the addition of the enzyme (1-10 μM) and in case of the one-pot assay by the addition of both the NpRdhA and PDR. Both the NADPH and the halogenated substrate were solubilised in the working buffer solution. The reactions were set up anaerobically using crimp seal vials that were then transferred to an incubator at 25° C (or orbital shaker at 250 rpm). The reactions were allowed to proceed for different times (15 min to 2 h) and the assays were terminated by the addition of TCA 5%. The samples were then spun down at 13 200 rpm using a Microfuge, to separate the precipitated components. The supernatant was recovered and put into HPLC vials for reverse-phase HPLC analysis.

Reverse-phase HPLC. Sample analysis for product formation was performed using an Agilent System 1110 Series, equipped G1379A degasser, G1312A binary-pump, a G1367A well plate autosampler unit, a G1316A temperature-controlled column compartment, and a single-wavelength UV detector. The stationary phases were Kinetex C-18 (250 mm length, 4.6 mm diameter, 5mm particle size (Phenomenex)), and Syncronis C-18 (250 mm length, 4.6 mm diameter, 5mm particle size (Thermo Scientific)). The mobile phase was made with HPCL grade acetonitrile/water (50:50) with 0.1 % TFA as additive. Substrates standards and product markers were prepared in acetonitrile. The flow rate was maintained at 1 $mLmin^{-1}$. The detector was set at 245 nm.

Cytochrome c reduction assay. Purified samples of the OtPDR reductase were buffer exchanged into degassed buffer A [Buffer A: HEPES-NaOH 50 mM, pH 7.8, NaCl 250 mM] under anaerobic conditions in a glove-box (Belle Technology, UK). To determine the affinity for pyridine nucleotides, titrations with either NADH or NADPH were made in presence of cytochrome *c* reductase, at saturating concentrations (40 μM). Reactions were started by the addition of the nucleotide and monitored by recording the whole spectra with a Cary 60 UV/Visible spectrophotometer. The reduction rates were obtained by fitting the data to a non-linear model (Michaelis-Menten).

EPR spectroscopy. Samples were prepared as isolated in volumes of 300 μL into 4 mm Suprasil quartz EPR tubes (Wilmad) under anaerobic conditions. The tubes were anaerobically sealed and frozen and stored in liquid nitrogen. Experimental parameters during the run were established as follows: microwave power 0.5 mW, modulation frequency 100 kHz, modulation amplitude 5 G, temperature 30 K. Spectra were obtained

using a Bruker ELEXSYS E500 spectrometer, Super high Q resonator (ER4122SHQ), Oxford Instruments ESR900 cryostat and ITC503 temperature controller.

In silico methodology. The homology-models were generated using the I-TASSER server, designed for automated structure-function prediction [17]. The output file contains information about the coordinates and contacts in the modelled 3D structure. The original I-TASSER models relative to the PDR-like domain and its cognate RdhA were minimised using the Amber forcefield in HyperChem 8.0 (Hypercube) and then refined using the Rosetta-relax-fast protocol (rosettacommons.org). An evaluation of the quality of the models was performed using the Rosetta design-HMMer (Rd.HMM). In this protocol, a 3D homology model is considered to be close to an equilibrium structure and certain energy values are assigned to it; if it retrieves its original primary sequence from the databases the model is considered acceptable and no energy penalisation will occur. The score of the original amino acid sequence should be amongst the top scores (or ideally the first), and the sequences in this group should present high sequence similarity amongst them. Rd.HMMer was performed using 15 intermediates with randomised sequences and each was reconstructed 10 times. The searches were done against the Ref-Seq database included in the NCBI [18]. The server ClusPro [19], was used to determine the protein-protein interactions during a rigid-body-docking analysis that depends on energy parametrisations to generate low-energy clusters of the protein complexes.

5.4 Results and discussion

Cloning and expression of a soluble artificial ssRdhA. Homologous proteins to BhbA are distributed in the superphylum Proteobacteria, both as full-length enzymes and as versions that lack the PDR-like domain, such as the NpRdhA [8]. Recently, our group proved that NpRdhA and other B_{12}-dependent proteins, such as QueG [20], could be expressed as active holo-enzyme in *Escherichia coli* DE3 strains using the BtuB transporter. The yields were comparable to those obtained when expressing it in *B. megaterium* (2.0 to 2.5 mgL^{-1}), and there was no need for cofactor reconstitution protocols to incorporate both the B_{12} and Fe-S cofactors, as opposed to respiratory enzymes such as PceA and CprA [21, 22].

Preliminary data obtained with various catabolic self-sufficient RdhA-PDR enzymes homologous to BhbA [10] suggest that its expression in both heterologous

bacterial hosts is challenging due to its low expression yields and low cofactor incorporation, that in turn, affects the amount of soluble holo-enzyme. Our experience suggests that NpRdhA might be unique and a model RdhA, as its expression as a soluble and active holo-enzyme in high yields after a single purification step via IMAC could not be reproduced even with other orthologous catabolic RdhAs that lack the PDR-like reductase domain, such as the *Burkholderiales joshi* and *Entotheonella* (candidatus) RdhAs [23]. Interestingly, these truncated versions of catabolic RdhAs are not commonly found or distributed as their full-length counterparts in Proteobacteria. We speculate that the presence of a 44 aa length internal peptide located in the vestigial B_{12}-binding domain of the ssRdhAs (but not present in NpRdhA), might contribute to protein aggregation hence complicating its expression and purification (Results Chapter II). *Nitratireductor pacificus* pht-3b [15] encodes three isoforms of catabolic RdhAs, including a self-sufficient ssRdhA that has not been functionally characterised, mainly because small-scale expression trials showed that it was found mostly in the insoluble fraction. For this reason, we decided to produce an artificial NpRdhA-PDR fusion protein that potentially will constitute a more tractable option than the natural self-sufficient RdhA enzymes. We present an alignment of the NpRdhA-PDR, the non-self-sufficient NpRdhA, and the ssRdhA from *Pseudovibrio* sp. FO-BEG that donated the linker and terminal PDR-like domain used to build the artificial fusion **(Supplementary Figure 1)**.

The NpRdhA-PDR fusion (**Figure 1)** was generated using a three fragment In-Fusion (Clonetech®) reaction and the insertion of the two fragments into the pEt30a(+) vector was verified by sequencing (Eurofins). A similar method was used to generate a NpRdhA-PDR construct with the pPT7 vector, for transformation into *B. megaterium*.

10	20	30	40	50	60
MRLYSNRDRP	NHLGPLALER	LARVDDVVAQ	PARQPEDGFA	ASEDSLLGDV	EEYARLFTRF
70	80	90	100	110	120
LDGPVAPLGD	AIPDDPARRA	ENLKASAYFL	DASMVGICRL	DPDDRAGDCD	PSHTHALVFA
130	140	150	160	170	180
VQFGREPEAG	EAGAEWIRGT	NAARTDMRCA	EIAAILSGYV	RWMGFPARGH	FSGDAQVDLA
190	200	210	220	230	240
RLAVRAGLAR	VVDGVLVAPF	LRRGFRLGVV	TTGYALAADR	PLAPEGDLGE	TAPEVMLGID
250	260	270	280	290	300
GTRPGWEDAE	EEKRPLHMGR	YPMETIRRVD	EPTTLVVRQE	IQRVAKRGDF	FKRAEAGDLG
310	320	330	340	350	360
EKAKQEKKRF	PMKHPLALGM	QPLIQNMVPL	QGTREKLAPT	GKGGDLSDPG	RNAEAIKALG
370	380	390	400	410	420
YYLGADFVGI	CRAEPWMYYA	SDEVEGKPIE	AYHDYAVVML	IDQGYETMEG	ASGDDWISAS

```
        430         440         450         460         470         480
QSMRAYMRGA  EIAGVMAAHC  RRMGYSARSH  SNAHSEVIHN  PAILMAGLGE  VSRIGDTLLN

        490         500         510         520         530         540
PFIGPRSKSI  VFTTDLPMSV  DRPIDFGLQD  FCNQCRKCAR  ECPCNAISFG  DKVMFNGYEI

        550         560         570         580         590         600
WKADVEKCTK  YRVTQMKGSA  CGRCMKMCPW  NREDTVEGRR  LAELSIKVPE  ARAAIIAMDD

        610         620         630         640         650         660
ALQNGKRNLI  KRWWFDLEVI  DGVAGAPRMG  TNERDLSPDR  GDKIGANQKL  AMYPPRLQPP

        670         680         690         700         710         720
PGTTLDAVLP  VDRSGGLAEY  AAAETPAAAR  ARLASGSNDL  APVYEVSASS  SPVLAVRLSK

        730         740         750         760         770         780
VNAEAEGVSR  YEFTLDDGSP  LPVFEAGAHI  DVVVAPEFLR  QYSLAGDPAD  NSKYVIGVLE

        790         800         810         820         830         840
EKEGRGGSAL  MHRIFHEGRR  VFISKPINHF  PLHEDAKKTL  LFGGGIGVTP  MIAMAHRLHA

        850         860         870         880         890         900
LGKPFELHYC  FRHRHKAGFI  EEIQNTQWAD  KAFIHCSSEG  TRADLKSILA  GYEDGYHVYT

        910         920         930         940         950         960
CGPDVFMDGV  LDTAAANGWS  EESLHKEYFS  VPDQGDYVNT  SFFVKLASTG  ERIEIPEDKS

        970         980         990        1000        1010        1020
AADVLNEKGI  SVPTKCSDGI  CGVCTAKYKG  ADVEHRDFVL  SGKEKEDTLV  LCCSRAKDAG

AELLLEL
```

Figure 1. Sequence of the NpRdhA-PDR fusion (112 kDa (1027 aa)). The NpRdhA module is highlighted in pink and the PDR-like module in purple).

Although the transformation in *B. megaterium* MS941 protoplasts containing an internal plasmid pPT7-RNAP for xylose-inducible expression was successful, and the artificial fusion was expressed our results indicate that the purification via affinity chromatography (IMAC) could not be achieved to homogeneity, as was the case of the natural ssRdhAs (data not shown). Fortunately, the co-transformation of *E. coli* expression strains (BL21 and HMS174) with the NpRdhA-PDR fusion, together with the BtuB transporter (*btuB*-pLysS/pEt3a), allowed its expression using standard conditions of expression for this host (17 °C, 12 h, and IPTG 0.25 mM). The protein thus expressed was purified in high yields by a single step IMAC using the Ni-NTA. Further purification steps were done to increase the purity of the sample, such as gel-filtration chromatography. We summarise the results of the expression and purification of the NpRdhA-PDR in *E. coli* HMS174 in **Figure 2**. The UV/Visible spectroscopy of the purified sample revealed a similar spectrum to that obtained for other RdhAs, with a broad absorbance between 300 and 600 nm, corresponding to the Fe-S clusters and corrinoid

cofactors under aerobic conditions [8, 9, 22]. However, the thermal denaturation of the sample in presence of KCN revealed low incorporation of B_{12} (15 to 20 %); this observation could not be reinforced by EPR data.

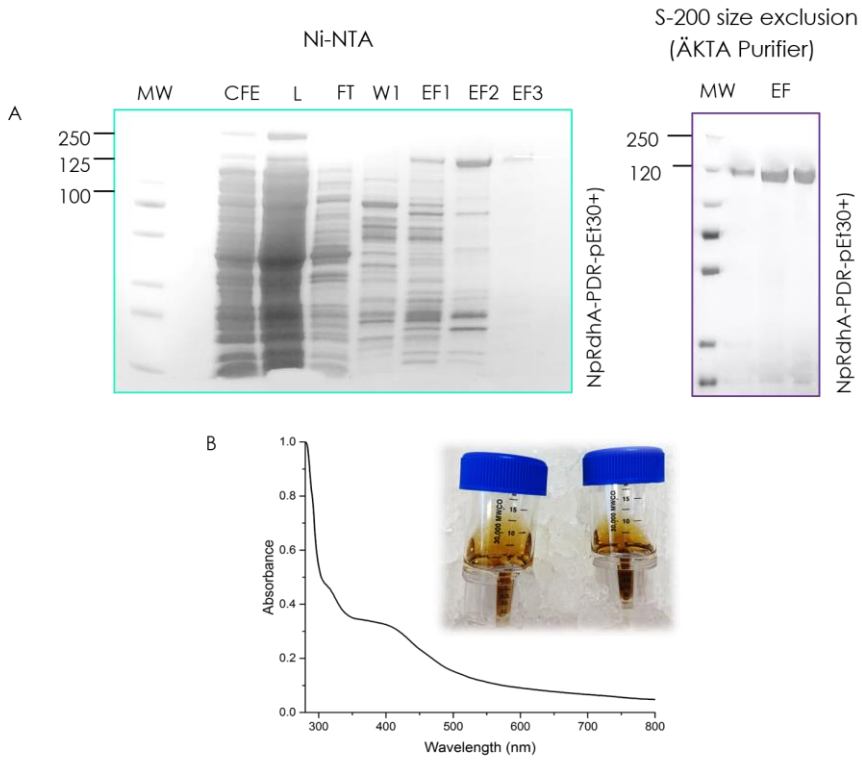


Figure 2. Heterologous expression of the NpRdhA-PDR in *E. coli*. A) SDS-PAGE of the IMAC (Ni-NTA) purification of the artificial protein NpRdhA-PDR, followed by size-exclusion using an analytical columns S-200 in the ÄKTA Purifier (GE). The protein of interest can be observed at approximately 120 kDa. B) UV/Visible spectrum of the purified samples, showing the usual spectral features for the B_{12} and Fe-S. Data was plotted using OriginPro ®.

The low cofactor complementation of the NpRdhA-PDR samples did not cause their aggregation into inclusion bodies, as was observed in case of the natural ssRdhA architectures (Results Chapter II). We speculate that the issues observed regarding the heterogeneous cofactor incorporation likely derived from using two vectors with the same origin of replication (OriC) when co-transforming *E. coli*, as we noticed that the yield and cofactor incorporation varied greatly in each co-transformation and large-scale growth.

Although our results indicate that the artificial NpRdhA-PDR fusion could be expressed as a soluble protein in a comparatively straightforward manner as NpRdhA, the artificial fusion protein could not support NAD(P)H-dependent dehalogenation of

the 3,5-dibromo-4-hydroxybenzoic acid substrate *in vitro* consistently, even when the B_{12} and Fe-S cofactors were present, as is shown in **Figure 3**.

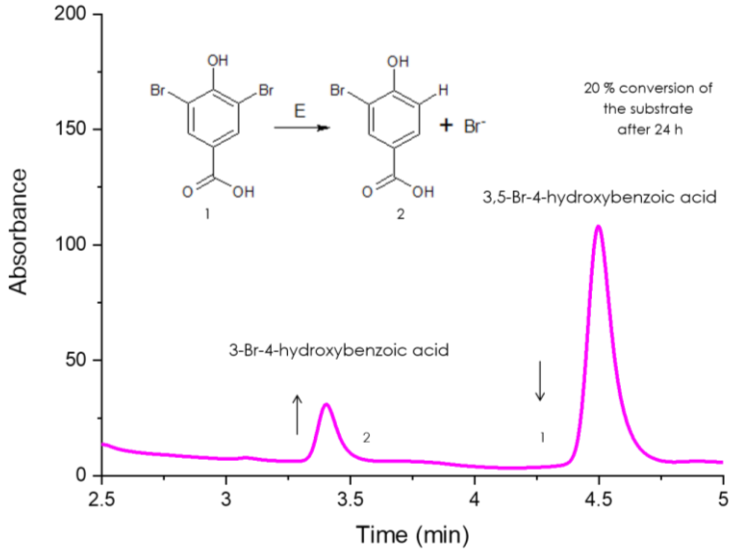


Figure 3. The NpRdhA-PDR fusion enzyme does not support NADPH-dependent dehalogenation. HPLC chromatogram showing that the dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid substrate did not proceed quantitatively *in vitro* in presence of 5 mM NADPH, at 25°C, after 24 h of reaction. The formation of 3-bromo-4-hydroxybenzoic acid can be attributed to the dehalogenation of the substrate catalysed by light. Product formation was analysed by reverse phase HPLC. Data plotted using OriginPro ®.

Whole-cell biotransformation assays in both *E. coli* and *B. megaterium* further reinforced this observation (data not shown), and we concluded that the low cofactor inclusion was not directly responsible for the lack of activity. Instead, we suspected that either the individual domains were not active or that the artificial fusion might acquiere an erroneous folding due to the inadequate truncation of the long linker region during cloning. This led us to consider the characterisation of the independent domains to test the hypothesis.

Cloning and expression of the individual PDR-like reductase module. We generated the individual PDR-like domain from the ssOtRdhA from *Ottowia thiooxydans.* The sequence was truncated at the beginning of the linker region between the B_{12}/Fe-S domains and the reductase module (740 aa). The truncated gene was then cloned into pEt28a(+), in the hope that its independent characterisation might shed some light as to why no activity was observed in case of the artificial fusion. We expressed the PDR module in *E. coli* BL21 and purified it in high yields via IMAC using the Ni-TDE resin and a polishing step of anion-exchange with the Res-Q column (ÄKTA Purifier). The

spectrum after purification revealed the typical features observed for FMN containing proteins, overlapping the broad absorbance of the rhombic 2Fe-2S cluster, as is shown in **Figure 4.** A bathophenanthroline colourimetric determination revealed that two equivalents of Fe are present per mole of protein, as expected with the inclusion of the 2Fe-2S cofactor, that couald be analysed by EPR.

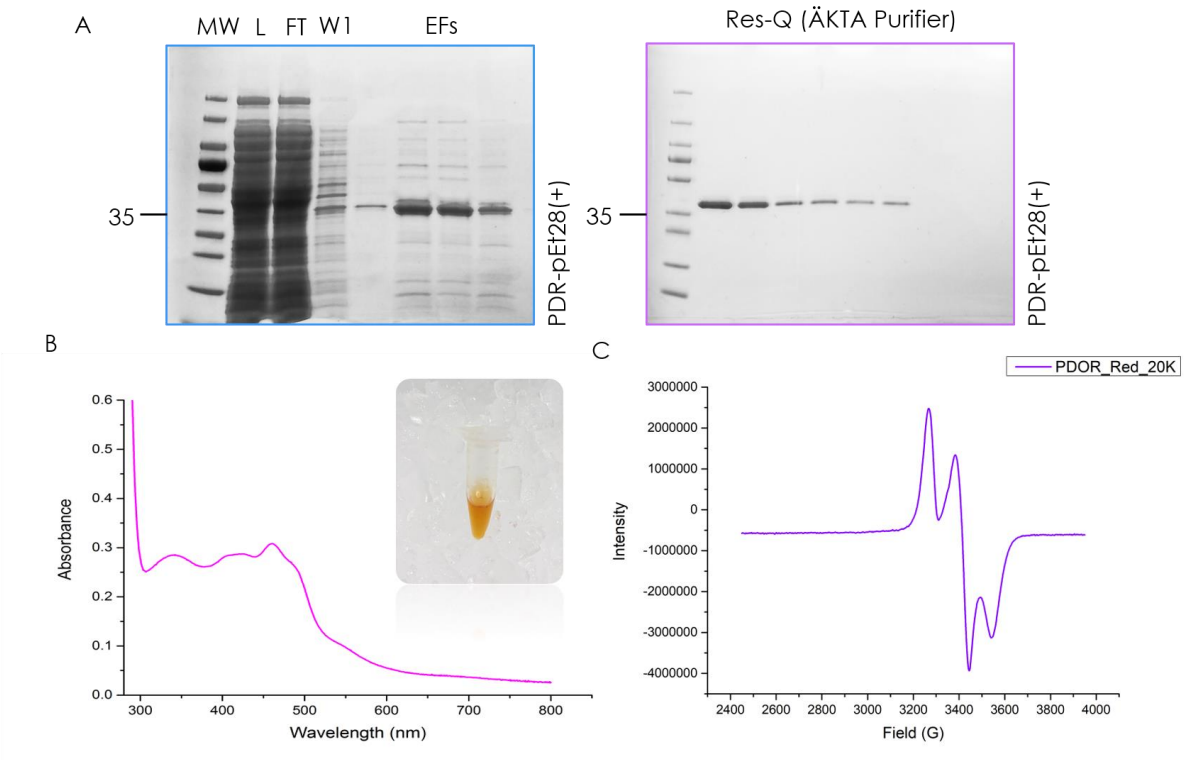
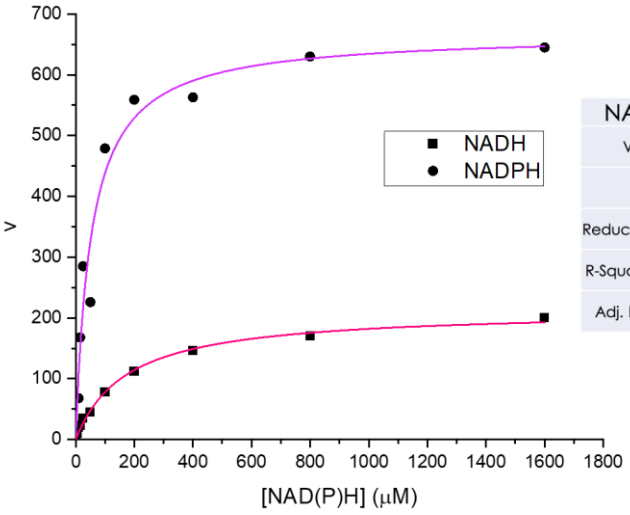


Figure 4. Purification of the PDR-like domain via IMAC. A) IMAC purification of the OtPDR-like reductase using a Ni-TDE resin domain, followed by a polishing step of Res-Q (ÄKTA Purifier (GE)). SDS-PAGE revealed a band at approximately 36 kDa, relative to the purified reductase module. B) Additionally, we present the UV/Visible spectrum of the doubly purified samples, showing the typical FMN absorbance features at 300-350 nm, being overlapped by the broad Fe-S absorbance. C) EPR spectra of the sodium dithionite reduced OtPDR sample (100 µM), showing the features relative to the rhombic 2Fe-2S cluster. The reduced FMN is not observed, as semiquinones (SQs) are not paramagnetic. Data plotted using OriginPro ®.

The FMN quantification by UV/Vis absorbance suggests that one equivalent of FMN is present per mol of protein after the two sequential purification steps, and that at least 75 % of the sample possessed the FMN when freshly purified. Although the dissociation constant of the FMN was not determined, we noticed that this cofactor could be readily reconstituted when added to the working buffer solution.

We performed cytochrome *c* reduction assays in presence of growing concentrations of both NADH and NADPH to determine if the PDR was active. Our results indicate that this particular PDR module is able to catalyse the reduction reaction with both pyridine nucleotides, but that it binds the NADPH substrate with higher affinity, as suggested by our saturation curves. Our results are sumarised in **Figure 5**.



NADPH		NADH	
Vmax	667.83271 ± 34.60853	Vmax	213.79167 ± 6.25753
Km	52.50243 ± 10.96116	Km	173.99489 ± 16.31793
Reduced Chi-Sqr	2504.76145	Reduced Chi-Sqr	32.63731
R-Square (COD)	0.96179	R-Square (COD)	0.99402
Adj. R-Square	0.95701	Adj. R-Square	0.99336

Figure 5. NADPH/NADH-dependent cytochrome *c* reduction by PDR-like reductase. Titrations with increasing concentrations of pyridine nucleotides performed at a fixed cytochrome *c* concentration (40 μM) under anaerobic conditions. Data was fitted to a non-linear model, revealing that the OtPDR preferentially binds NADPH over NADH, given the sharpness of the inflection in each curve and the approximate values for the calculated kinetic parameters. Data was plotted using OriginPro ®.

Once that the preference for the phosphorylated nucleotide (NADPH) was established and to test if these reductase modules could support dehalogenation by transferring electrons to the central dehalogenase domain we developed a multienzyme or "one-pot" assay, by adding either NpRdhA or the OtssRdhA dehalogenase module. Unfortunately, the B_{12}/Fe-S cognate domain cloned into pPT7 for expression in *B. megaterium*, could not be expressed and purified in high yields, so we opted for NpRdhA to represent the RdhA domain instead **(Supplementary Figure 3)**. The NADPH-dependent dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid performed in the one-pot assay suggests that the individual reductase module is active and able to transfer electrons to the dehalogenase active site, thus supporting the dehalogenation of the substrate when an excess concentration of NADPH was added to the reaction mixture. As has been reported for other prototypical phthalate dioxygenase reductases, such as those from *B. cepacia* [11] and *Pseudomonas putrida* [24], apparently our

truncated PDR can mediate both intramolecular/intermolecular electron transfer reactions. The electrons are chanelled from the NAD(P)H subatrate, through an array of Fe-S clusters, and finally to the cognate active sites, depending on the structure of specific multi-domain oxidoreductases or the individual components of an electron transfer chain. Our results indicate that PDR can support dehalogenation in a 1:1 ratio with percentages of conversion to the first product of dehalogenation of approximately 64% after 1.5 h of reaction, at 25 °C, although the conversion rates are still not comparable to those obtained by Collins *et al.* [14], where a vast excess of the non-physiological redox partners was required for turnover **(Figure 6)**.

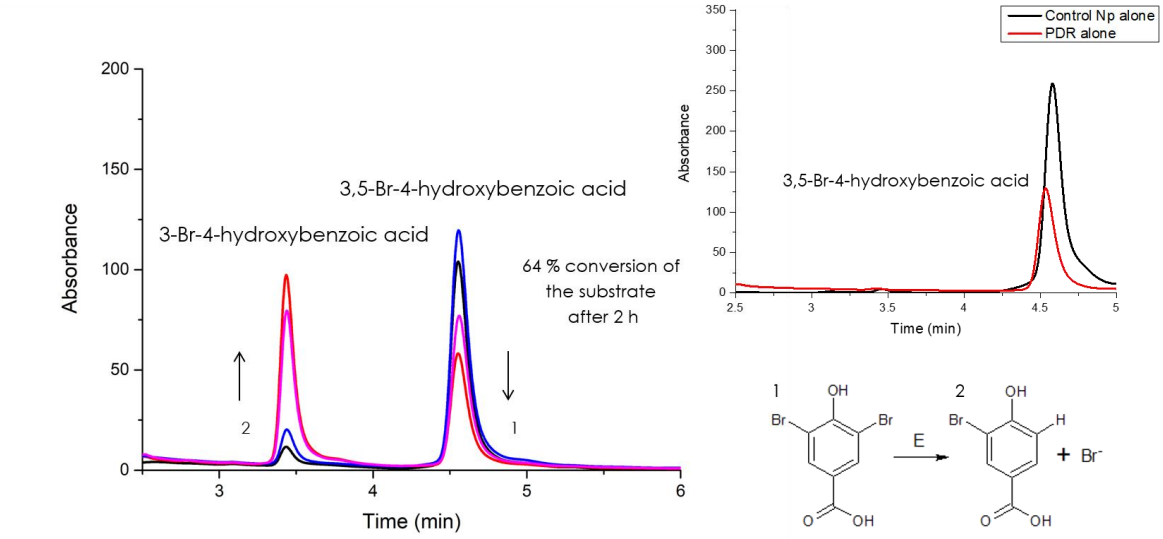


Figure 6. NADPH-dependent dehalogenation in "one-pot" assay, that combines the NpRdhA and PDR. HPLC chromatograms showing the *in vitro* formation of 3-bromo-4-hydroxybenzoic acid (the first product of dehalogenation of the 3,5-bromo-4-hydroxybenzoic acid) after 15 min, 30 min, 1 h, and 2 h of reaction in presence of NADPH (5 mM), at 25 °C. Our results suggest that the reduction occurs quantitatively at NpRdhA/PDR rations of 1:1 (10 µM each). In addition, we present the controls relative to the individual proteins, that were assayed at the same reaction conditions and analysed after 2 h of incubation. Data was plotted using OriginPro ®.

Albeit this, proper optimisation of the reaction conditions (pH, temperature, ionic strength, and protein ratios) is needed to evaluate if the assay will be rubust enough to be used routinely. Despite the high yields and purity of the preparations, we could not perform a complete characterisation of this module because of the poor stability of the protein at room temperature, fact that complicates its handling during assays, even when stabilisers such as glycerol were added to the buffer. For the same reason, we could not obtain crystals suitable for diffraction, albeit extensive screening of

conditions using the ThermoFluor assay as a guide. We generated different truncated versions of the PDR with varying lengths of the N-terminal linker region, but time was insufficient to test them. In addition to this, other PDR modules from different ssRdhA orthologues were cloned and preliminary results suggest that these might be better targets for biophysical and kinetic characterisation, so that a complete description of the PDR reductase can be achieved before the standarisation of the ''one-pot'' assay is attemped. A redox analysis performed by EPR, combined with stopped-flow experiments might shed light into the oxidation state of each cofactor and the thermodynamics of the electron transfer during catalysis.

The analysis of the protein-protein interactions between both modules, in turn, will help to understand the domain motion of the ssRdhA during catalysis. We generated homology models of the individual domains using I-TASSER [17] to compare them with NpRdhA and the crystals structure of the PDR reductase domain from *Burkholderia cepacia,* in order to establish if the key amino acid residues that determine its specificity towards NADPH or NADH are present. An inspection of the PDR-like model, refined with Rosetta-relax-fast, and later evaluated with the Rosetta HMMer [18] protocol, revealed that it contains a Ser (177) that appears to be conserved in all ssRdhA analysed. This residue allows the preferential binding of NADPH but is substituted by Asp (173) in *B. cepacia* and other PDR-like reductases characterised to date [11, 25] **(Figure 7)**.

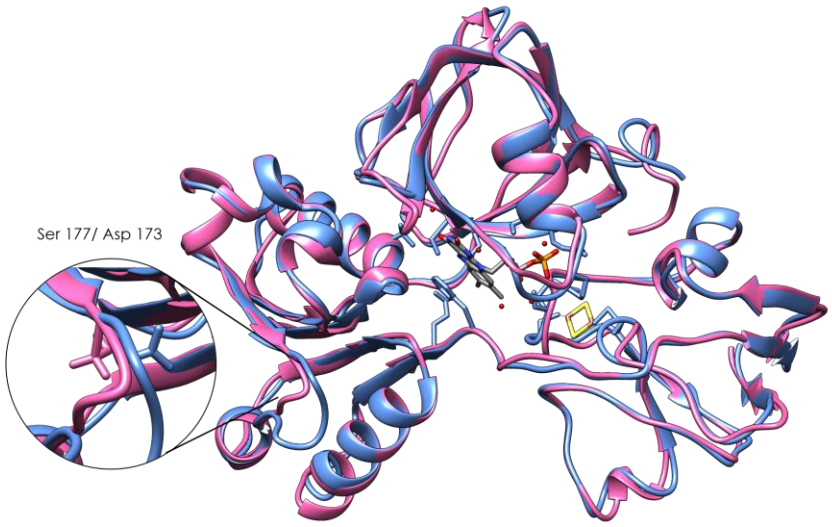


Figure 7. Structural alignment. The homology model of the PDR-like from *Ottowia thiooxydans* (OtPDR) generated by I-TASSER and refined with the guide of Rosetta HMMer protocol was aligned with the crystal structure PDR from *B. cepacia* (2PIA) solved at 2 Å. The structural alignment between OtPDR (pink) and BcPDR (ice blue) was performed with Chimera revealing an RMSD of 0.345 Å (for 311 atom pairs). In the

inset, we present the residues that determine the NAD binding properties; the presence of a Ser (instead of Asp) indicates the preference for NADPH.

A rigid body-docking [19] performed with our homology models showed that the PDR domain can dock to the RdhA domain in multiple orientations. However, not all of them will allow the electron transfer within physiological limits, that establish that distances longer than 14 Å will result in inefficient transfer and uncoupling of the reaction **(Figure 8)**. We hypothesise that the interaction of both domains defines the catalysis *in vitro* and *in vivo*, depending on the conformational freedom that the linker region allows, considering its dynamic behaviour in the solvent.

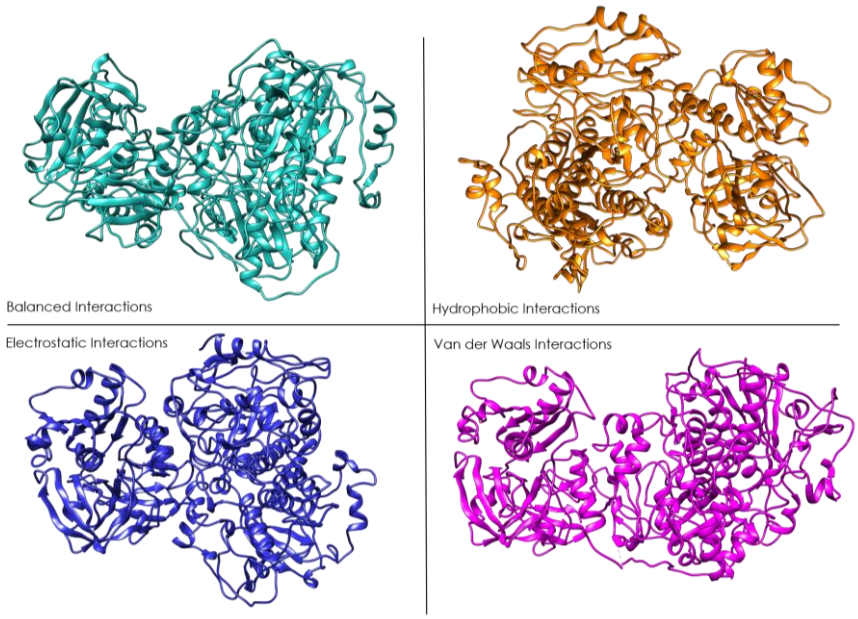


Figure 8. Rigid-body-docking analysis performed by ClusPro. Models of the individual RdhA (B_{12}/Fe-S) and PDR-like modules generated by homology modelling with I-TASSER and refined with the guide of the Rosetta HMMer protocol were utilised for rigid-body-docking analysis. Multiple docked orientations were ordered in terms of the proximity of the cofactors and depending on the type of interactions (and the energy) that governed the clustering. The clusters with the lowest energies were preferred (-845 to -958.6 $KJmol^{-1}$).

Our current results, however, do not explain why the artificial NpRdhA-PDR protein is not active when its independent modules are. We think that erroneous folding of the fusion prevents the adoption of conformations that allow productive electron transfer, mainly due to issues with the linker region between both modules. Linkers might be determinant for the activity of the protein by allowing the right conformational freedom and orientation of the modules during catalysis. It has been reported that

linkers allow the independent folding of the functional domains [26], and it is hypothesised that when the linker regions cannot retain specific secondary structure to restrict their movement in solution (such as coils), this affects the folding process of each domain thus leading to misfolded proteins. Redesigning the linker region might help to avoid these issues, but more work is needed to determine why this specific NpRdhA-PDR fusion did not catalyse the reduction of the 3,5-dibromo-4-hydroxybenzoic acid, even when the independent B_{12}/Fe-S and PDR-like modules were active, as revealed by the spectrophotometric methyl viologen (data not shown) and cytochrome *c* reduction assays. In addition to this, the low cofactor complementation, which is comparable in magnitude to that reported for other B_{12}-dependent enzymes [8] and therefore, cannot be responsible for the lack of activity solely.

Our results suggest that this specific PDR-like reductase mediates the electron transfer from the two-electron donor NADPH to the one-electron acceptor Fe-S in the form of a hydride ion (H^-), in a similar manner to other prototypical iron-sulfur dependent flavoproteins (FNR) [27]. FNR-type reductases participate in the electron transfer from pyridine nucleotides to diverse oxidoreductases by forming complexes [25], but these can also be found fused to other domains, as is the case of some cytochrome P450 fusions [28]. Consequently, we claim that the PDR-like module also participates as a redox capacity by channeling electrons to the B_{12} cofactor during organohalide reduction **(Supplementary Figure 4)**.

It is fundamental that a proper functional and structural characterisation of the individual domains of the ssRdhAs is performed (including the cognate RdhA), as perhaps these catabolic self-sufficient architectures will offer a more direct possibility to study the mechanistic details of reductive dehalogenation, since there is no requirement for exogenous redox partners, as is the case of NpRdhA. However, given the fact that the artificial protein NpRdhA-PDR did not support the dehalogenation of the 3,5-dibromo-4-hydroxybenzoic acid, this means that instead, the optimisation of the "one-pot" assay has to be attempted as an alternative, until a standard protocol for the heterologous expression and purification of the natural full-length ssRdhA is developed.

5.5 Conclusions

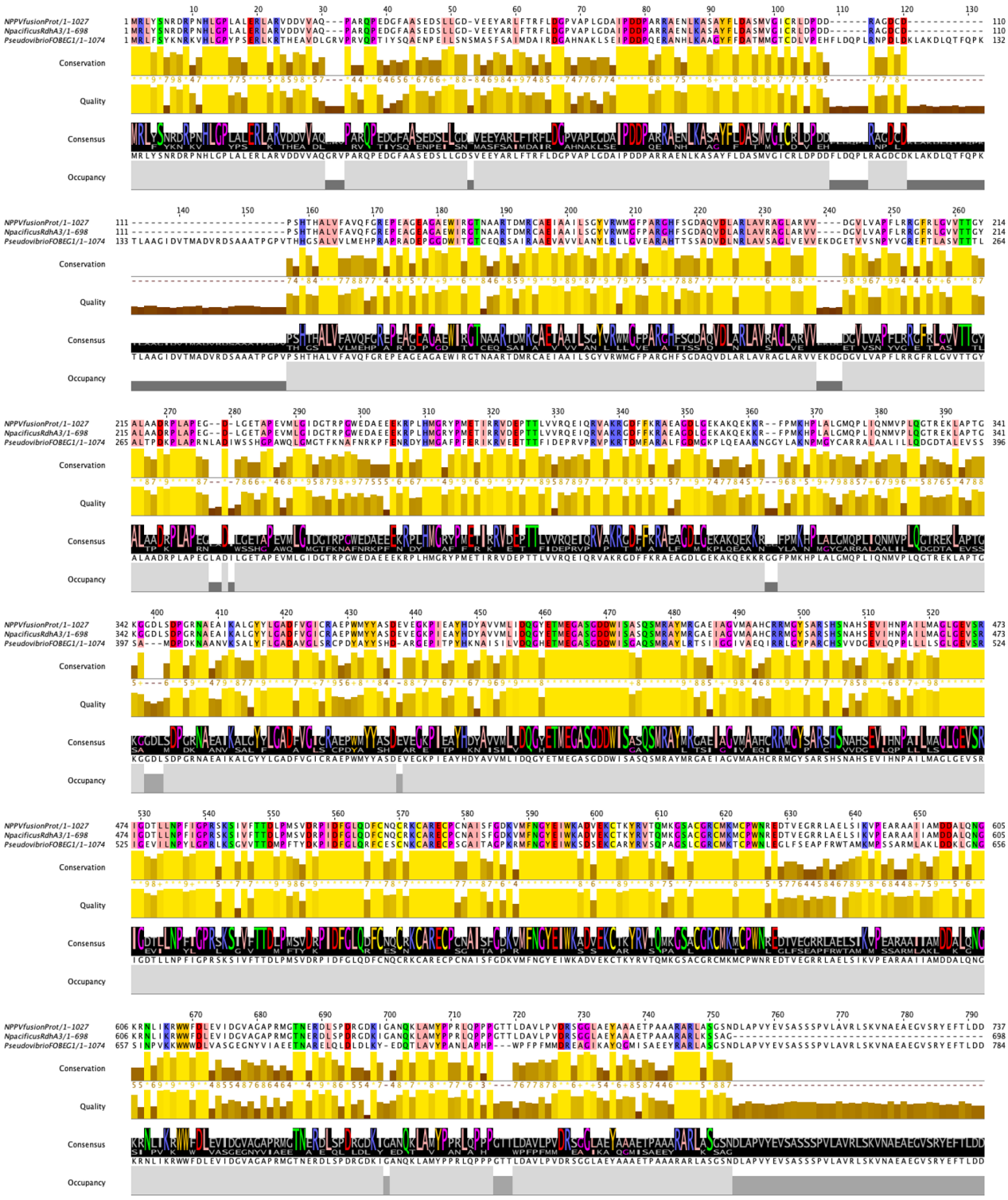
Traditional strategies of bioremediation of highly contaminated sites are expensive and inefficient, in contrast to new techniques that exploit the potential of

microbiota, specifically the strategies evolved by some consortia of both Gram-positive and Gram-negative microorganisms that can survive under environmental stresses derived from the presence of naturally occurring organohalide compounds. However, OHRBs have been reported to grow at slow rates and are highly sensitive to temperature, pH, redox conditions, light, and symbiotic interactions, hence the need to improve the current methodologies for heterologous expression of soluble and active enzymes in high yields. The development of artificial fusions that support the dehalogenation of different organohalides and the optimisation of multienzyme assays, such as the one developed by Collins *et al.* **[14]** or the one-pot assay presented in this work, can provide us with valuable strategies for the implementation of new long-term applications in cell-free bioremediation of heavily polluted sites.

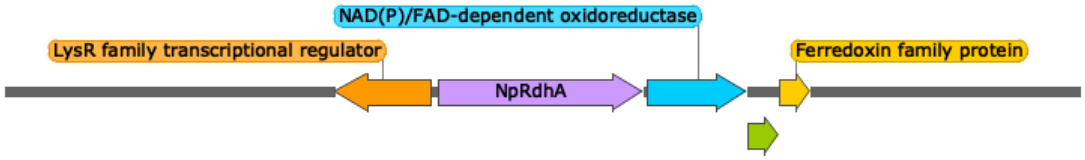
Although the NpRdhA-PDR artificial protein could be expressed in high yields (1.5 mgL^{-1}) in both heterologous bacterial hosts, it did not support the dehalogenation of the 3,5-bromo-4-hydroxybenzoic acid substrate. We hypothesise that the fusion protein is not active due to the lack of structure of the linker region between the B_{12}/Fe-S and the PDR-like modules. Despite this, our results serve as a proof of principle by showing that other dehalogenase-based chimeric architectures can be made using this methodology to generate other fusions with the advantage of using a well characterised enzyme such as NpRdhA and other oxidoreductase modules from the FNR family with relative ease, as possibly these will offer numerous alternative possibilities to study the electron transfer process.

The authors declared no conflict of interest. Karl Payne cloned the artificial ssRdhA construct and Samantha Gaytán (S.G) the individual domains. Karl Payne transformed *B. megaterium*. S.G expressed and purified the artificial RdhAs and individual domains, performed metal-cofactor determination, UV/Visible spectroscopy, and prepared the samples for EPR. S.G measured the activity of the enzyme. S.G. performed the *in silico* analysis presented in this work. S.G. wrote the manuscript. David Leys designed and directed the research. **Acknowledgments.** Many thanks to Professor Martin Warren (University of Kent) for sharing the BtuB expression vector with the Leys group. Samantha Gaytán received a scholarship from CONACyT (Mexico). D.L. is a Royal Society Wolfson Research Merit Award holder.

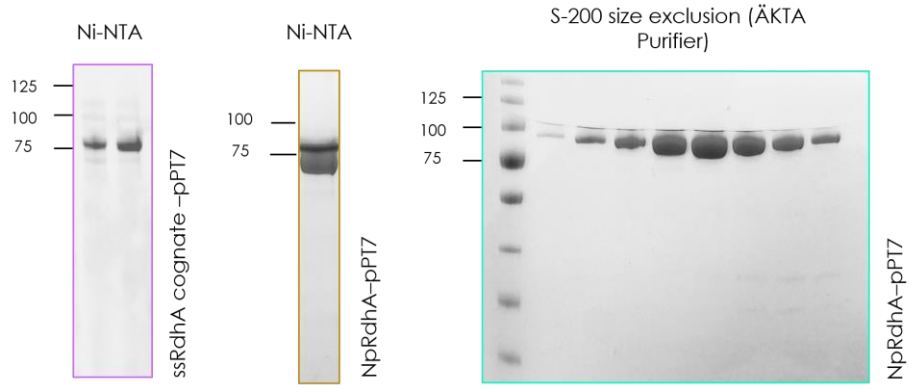
5.6 Supplementary Figures



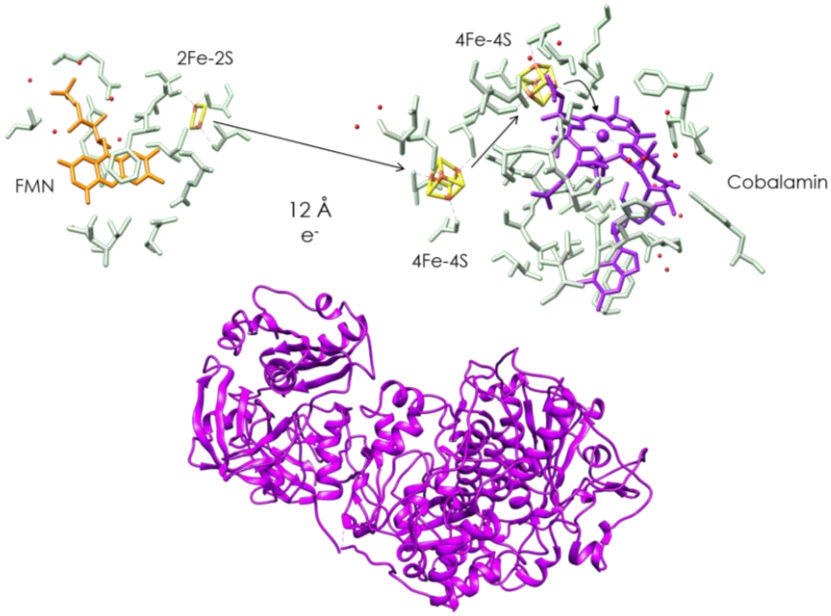
Supplementary Figure 1. Section of the MAFFT sequence alignment of the artificial fusion NpRdhA-PDR, the non-self-sufficient NpRdhA from in *N. pacificus* pht-3b, and the self-sufficient RdhA from *Pseudovibrio* sp. FO-BEG, that "donated" the linker region and the PDR-like domain used to generate the artificial fusion. The alignment was generated by JalView and only the B_{12}/Fe-S and the linker region are shown.



Supplementary Figure 2. Cartoon representation of the genomic context of the *NprdhA* cluster. In the same contig, adjacent to the reductive dehalogenase structural gene. The RdhR (LysR) transcriptional regulator, the FNR – NAD(P)H-dependent flavodoxin reductase, FD2 (2Fe-2S), and FD4 (4Fe-4S) or Fe-S clusters binding ferredoxin. Interestingly PDR-like reductases are not found in the same gene cluster (or within a *contig*) as the *NprdhA* structural gene. Figure generated with SnapGene®.



Supplementary Figure 3. Purification of NpRdhA (75 kDa) and cognate OtssRdhA (78 kDa) cloned into pPT7 for heterologous expression in *B. megaterium.* The samples were purified via IMAC (Ni-NTA) and size exclusion chromatography (S-200 analytical column).



Supplementary Figure 4. Docked models of the individual RdhA (B_{12}/Fe-S) and PDR-like reductase. The cluster presented was selected depending on the proximity of the cofactors in both modules. In the figure, we represent one of the multiple orientations for productive electron transfer, according to physiological distances (maximum 14 Å).

5.7 References

1. G. W. Gribble, *Am. Sci.*, 2004, **92**, pp. 342-349.
2. P. Adriaens and A. L. Barkovskii, in *Encyclopedia of Environmental Microbiology*, ed. G. Bitton, Wiley-Interscience, New York, 2003, p. 509.
3. B.-E. Jugder, H. Ertan, M. Lee, M. Manefield and C. P. Marquis, *Trends Biotechnol.*, 2015, **33**, pp. 595-610.
4. M. Fincker and A. M. Spormann, *Annu. Rev. Biochem.*, 2017, **86**, pp. 357–386.
5. F. Maphosa, S. H. Lieten, I. Dinkla, A. J. Stams, H. Smidt and D. E. Fennell, *Front. Microbiol.*, 2012, **3**. 351.
6. T. Schubert, L. Adrian, R. G. Sawers and G. Diekert, *FEMS Microbiol. Ecol.*, 2018, **94**. fiy035.
7. M. Bommer, C. Kunze, J. Fesseler, T. Schubert, G. Diekert and H. Dobbek, *Science*, 2014, **346**, pp. 455-458.
8. K. A. P. Payne, C. P. Quezada, K. Fisher, M. S. Dunstan, F. A. Collins, H. Sjuts, C. Levy, S. Hay, S. E. J. Rigby and D. Leys, *Nature*, 2014, **517**, pp. 513-516.
9. D. P. Dowling, Z. D. Miles, C. Köhrer, S. J. Maiocco, S. J. Elliott, V. Bandarian and C. L. Drennan, *Nucleic Acids Res.*, 2016, **44**, pp. 9965–9976.
10. K. Chen, L. Huang, C. Xu, X. Liu, J. He, S. H. Zinder, S. Li and J. Jiang, *Mol. Microbiol.*, 2013, **89**, pp. 1121–1139.
11. C. C. Correll, C. J. Batie, D. P. Ballou and M. L. Ludwig, *Science*, 1992, **258**, pp. 1604-1610.
12. J. S. Bonde and L. Bülow, in *Brenner's Encyclopedia of Genetics*, eds. S. Maloy and K. Hughes, Elsevier, Massachusetts, 2 edn., 2013, pp. 519-520.
13. A. Kublik, D. Deobald, S. Hartwig, C. L. Schiffmann, A. Andrades, M. v. Bergen, R. G. Sawers and L. Adrian, *Environ. Microbiol.*, 2016, **18**, pp. 3044–3056.
14. F. A. Collins, K. Fisher, K. A. P. Payne, S. Gaytan, S. E. J. Rigby and D. Leys, *Biochemistry*, 2018, **57**, p. 3493–3502.
15. Q. Lai, G. Li and Z. Shao, *J. Bacteriol.*, 2012, **194**, p. 6958.
16. S. J. Moore, M. J. Mayer, R. Biedendieck, E. Deery and M. J. Warren, *N. Biotechno.l*, 2014, **31**, pp. 553-561.
17. A. Roy, A. Kucukural and Y. Zhang, *Nat. Protoc.*, 2010, **5**, pp. 725-738.
18. L. P. Martínez-Castilla and R. Rodríguez-Sotres, *PLoS One*, 2010, **5**. e12483.
19. D. Kozakov, D. R. Hall, B. Xia, K. A. Porter, D. Padhorny, C. Yueh, D. Beglov and S. Vajda, *Nat. Protoc.*, 2017, **12**, pp. 255-278.
20. K. A. P. Payne, K. Fisher, H. Sjuts, M. S. Dunstan, B. Bellina, L. Johannissen, P. Barran, S. Hay, S. E. J. Rigby and D. Leys, *J. Biol. Chem.*, 2015, **290**, pp. 27572-27581.
21. A. Parthasarathy, T. A. Stich, S. T. Lohner, A. Lesnefsky, R. D. Britt and A. M. Spormann, *J. Am. Chem. Soc.*, 2015, **137**, p. 3525–3532.
22. B.-E. Jugder, K. A. P. Payne, K. Fisher, S. Bohl, H. Lebhar, M. Manefield, M. Lee, D. Leys and C. P. Marquis, *ACS Chem. Biol.*, 2018, **13**, pp. 548-552.
23. F. A. Collins, PhD thesis, The University of Manchester, 2017.
24. J. Peter W. Roome, J. C. Philley and J. A. Peterson, *J. Biol. Chem.*, 1983, **258**, pp. 2593-2598.
25. G. T. Gassner, M. L. Ludwig, D. L. Gatti, C. C. Correll and D. P. Ballou, *FASEB J.*, 1995, **9**, pp. 1411-1418.
26. W. Strohl, in *Reference Module in Life Sciences*, ed. B. Roitberg, Elsevier, 2017.
27. P. Ferreira, M. Martínez-Júlvez and M. Medina, in *Flavins and Flavoproteins*, eds. S. Weber and E. Schleicher, Humana Press, London, 2014, ch. 5, pp. 79-94.

28. A. W. Munro, D. G. Leys, K. J. McLean, K. R. Marshall, T. W. B. Ost, S. Daff, C. S. Miles, S. K. Chapman, D. A. Lysek, C. C. Moser, C. C. Page and P. L. Dutton, *Trends Biochem. Sci.*, 2002, **27**, pp. 250-257.

6.0 Chapter Six. Closing remarks

The work that we present here contributes to the understanding of the RdhAs regulation and function. In Chapter I we studied a MarR-type transcriptional regulator that controls the *rdhA* expression found in the genome of *Dehalococcoides mccartyi* CBDB1, while the heterologous expression and initial biochemical characterisation of a catabolic self-sufficient RdhAs (ssRdhAs) from marine aerobic Proteobacteria (Chapter II), and the characterisation of its independent PDR-like reductase module (Chapter III) benefitted our incipient understanding of these enzymes. Although the transcriptional regulator that we studied belongs to a strict anaerobic OHRB, the ssRdhA enzymes studies are somewhat distinct from the OHRB respiratory RdhA. Regardless, understanding of reductive dehalogenation in the wider sense (*i.e*. whether as part of OHRB respiration or not) in terms of regulation and mechanism is required to support future applications in bioremediation. The fact the catabolic ssRdhA are O_2-tolerant and self-sufficient makes them attractive targets. The fact that phylogenetically diverse OHRBs present similar OHR genes suggests that reductive dehalogenation metabolism can be the result of horizontal transfer, via transposon-mediated events, as discussed by Atashgahi *et al*. **[1]**

6.1 Transcriptional regulation in OHRBs

In this work, we provided a rationale for the binding specificity of the $RdhR_{CbdbA1625}$ for certain aromatic halogenated compounds, thus contributing to the understanding of the transcriptional regulation of reductive dehalogenases genes in one of the best-studied OHRBs: *Dehalococcoides mccartyi* CBDB1. This bacterium encodes 32 *rdhAB* structural gene pairs in its genome and most *rdhA* genes are associated with either the two-component signal transduction systems or the MarR transcriptional regulators **[2-4]**. The *Dehalococcoides* genera is known for its ability to dechlorinate some of the most toxic environmental pollutants, such as polychlorinated dibenzo-p-dioxins (PCDD) and pentachlorophenol (PCP) **[5]**.

The $rdhR_{CbdbA1625}$ is associated with the $rdhA_{CbdbA1624}$ structural gene, which encodes a putative respiratory reductive dehalogenase whose expression is upregulated by the 1,2,4-trichlorobenzene (1,2,4-TCB) *in vivo*, according to results reported by Wagner *et al.* **[5]**. Based on this, we studied the ligand binding properties of the $RdhR_{cbdb1625}$ to determine

its affinity for different halogenated ligands. Although the chlorobenzene derivatives were considered initially the best effectors for our experiments with the $RdhR_{CbdbA1625}$ [6], their low solubility in H_2O/MeOH mixtures and the negative effects of adding organic solvents to the $RdhR_{CbdbA1625}$ samples lead us to test instead a range of more soluble chlorophenol ligands, as putative analogues of 1,2,4-TCB. Our results revealed that the 1,2,3-TCP isomers bind with almost nanomolar affinity to $RdhR_{CbdbA1625}$, with the 2,3,4-TCP being the tightest binding ligand. The binding data relative to the 3,4,5-TCP (K_D of 1.23 ± 0.24 μM), interestingly shows that the relative position of the hydroxyl group towards the halogen substituents does not affect the recognition of the ligand by the $RdhR_{CbdbA1625}$. This was previously suggested by Dr. Carolina P. Quezada, who obtained and solved the crystal structures of the RdhR $_{CbdbA1625}$ protein bound with different dichlorophenols. The $RdhR_{CbdbA1625}$ crystal structures were solved in its ligand-free and ligand-bound conformations revealing a typical folding found in other MarR transcription factors [7-9]. Building upon this groundwork, we determined an additional crystal structure of the $RdhR_{CbdbA1625}$ with a TCP regioisomer, using theirs as a model for molecular replacement to develop more specific and informative binding experiments. The ligand-bound structures with the dichlorophenols and trichlorophenols revealed a single hydrophobic pocket per monomer and that the stabilisation of the organohalides depends on hydrophobic interactions and putative halogen-hydrogen bonds.

Additionally, we showed that $RdhR_{cbdb1625}$ binds the 42 bp palindromic intergenic region between the *rdhR*$_{CbdbA1624}$-*rdhR*$_{CbdbA1625}$ with nM affinity, according to both isothermal titration calorimetry (ITC) and intrinsic fluorescence-quenching experiments. Originally, it was hypothesised that the MarR-type transcription factor $RdhR_{CbdbA1625}$ from *D. mccartyi* CBDB1 was a negative response regulator [10], and our analytical ultracentrifugation interference experiments (AUC) and electrophoretic mobility shift assays (EMSAs) indeed suggest that the $RdhR_{CbdbA1625}$ may act as a repressor of the transcription in absence of the 2,3,4-TCP, at least *in vitro*. Although the precise repression mechanism remains unknown, we speculate that the $RdhR_{CbdbA1625}$ may prevent the transcription of the *rdhA*$_{CbdbA1626}$ gene by sterically hindering the RNA polymerase (RNAP) [11], after binding between the core promoter elements. DNA microarrays or experiments designed to obtain proof of the transcription repression/de-repression behaviour *in vivo* are required to establish if the steric blocking of the RNAP binding rather than a roadblocking mechanism (that prevents the transcription elongation when the repression factor binds to the start of the coding region) is responsible for the observed behaviour of the $RdhR_{CbdbA1625}$ [12] **(Figure 1)**.

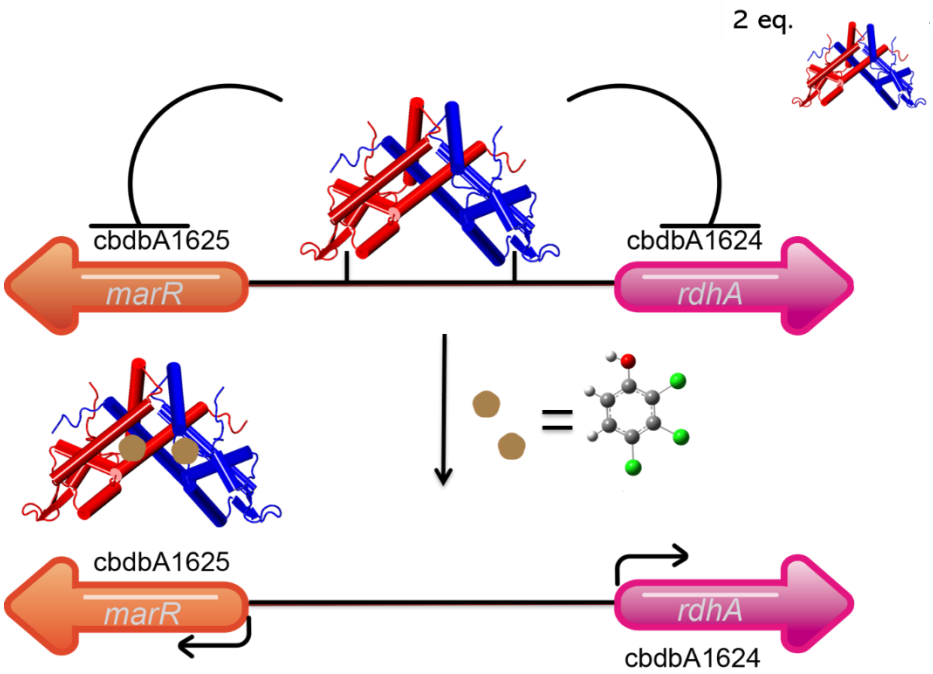


Figure 1. The characterisation of the RdhR$_{CbdbA1625}$ transcription factor helped us to establish important structure/function relationships between the binding site features on each protein monomer (shown in blue or red cartoon) and specific chlorinated ligands that mimic its putative physiological effector, such as the 2,3,4-TCP (shown in the image, with the chlorine substituents shown in green, but also represented as brown pentagons). In the picture, we illustrate the potential role of the RdhR $_{CbdbA1625}$ as a repressor of the transcription of the structural *rdhA* gene (cbdbA1624, illustrated in pink) and its own (cbdbA1625, shown in orange).

Our AUC, ITC, and gel-filtration experiments showed that the RdhR$_{CbdbA1625}$ exhibits a dynamic behaviour in solution, and that at certain concentrations it is found mostly as a dimer, largely independent of NaCl concentration, pH, or temperature. Upon recognition of the 42 bp palindromic dsDNA sequence, two dimeric RdhR$_{CbdbA1625}$ bind to the repeat motifs of the *rdhA* promoter. These results suggest positive cooperativity after the binding of the first dimer, as has been reported in the case of other MarR regulators, where the first binding event facilitates the second, although more experiments need to be conducted **[13]**. Some of the residues that participate in the interaction with the DNA were identified by comparing our crystal structure with the SlyA-dsDNA complex **[14]**. Despite many experimental trials, co-crystallisation of the RdhR$_{cbdb1625}$-dsDNA complex could not be obtained yet. Therefore, the specific interactions between the RdhR$_{CbdbA1625}$ and the palindrome target remain unknown, and hence, a specific repression mechanism is still elusive.

*

It is worth pointing out that most MarR homologs in Bacteria and Archaea negatively regulate operons that result in critical adaptative strategies for different environmental pressures, although MarR activators have also been reported [15]. In the absence of physiological ligands, MarR homologs can bind to the palindromic DNA, achieving repression in most cases, and upon binding of the ligand, de-repression of the transcription occurs [16]. Although the interaction with most ligands can cause attenuation of the binding of DNA, the fact that some of the physiological ligands have not been determined limits our understanding of the underlying molecular mechanism of regulation and the number of possible applications of the MarR homologous proteins.

The signals to which the MarR homologues from different metabolic pathways respond have been evolutionarily selected as adaptative strategies to specific environmental changes [15] and this is no exception in the context of reductive dehalogenation. Originally, when this research was planned the main objective was to establish a link between the primary structure of the MarR-type transcription factors and their physiological ligands, so that the substrate specificity of the RdhAs will be known without the need to express and purify the RdhAs, given the many challenges that this entails, particularly in case of the respiratory RdhAs [17]. It was also thought that the *in silico* determination of the physiological effectors of each of the MarR-type regulators found in the gene clusters of phylogenetically diverse OHRBs could be used as an additional tool for the classification of RdhAs, to select new research targets. Recently there has been an increased interest from the pharmaceutical and agricultural industries to study the allosteric mechanism of the MarR-type transcriptional regulators involved in the regulation of genes from distinct metabolic pathways, such as the lignin degradation. MarR transcriptional regulators have attracted attention for its potential as biosensors, and their possible applications in environmental sciences, as is the case of the whole-cell biosensors that aid in the detection of specific compounds by coupling the transcription factors (TFs) to reporter genes or fluorescent dyes, that give way to optical signals, even when the analyte concentration in solution is very low [18].

6.2 Characterisation of a self-sufficient RdhA and its PDR-like domain

In this thesis, we determined that it is possible to heterologously produce the catabolic self-sufficient RdhA enzymes with more success than previously reported by our group with the *Comamonas* sp. 7D-2 BhbA and other catabolic ssRdhA (as is the case of the *Pseudovibrio* sp. FO-BEG1). Earlier attempts at expressing the membrane-bound

respiratory RdhA enzymes and the full-length catabolic RdhAs in T7 based vectors both in *E. coli* and *B. megaterium* failed given the fact that either the protein was not expressed, expressed below detectable levels or was encountered mostly in the precipitate, or insoluble fraction, even when the TAT peptide (if present) and *rdhB* sequences were removed and only the *rdhA* genes were included in the construct design **[19, 20]**. Our modest results represent an important step forward, as both the expression and isolation process are methodologically so complicated that for many years have prevented the complete kinetic and structural characterisation of this family of B_{12}-dependent enzymes **[17, 21]**.

Prior research in our group showed that the catabolic reductive dehalogenase NpRdhA could be expressed as a soluble and active protein using *B. megaterium* as heterologous host **[22]**. However, both the lack of diversity in T7-based expression vectors and the difficulties found during the preparation of protoplasts, including a long 3 - 4 day transformation protocol, were impractical and led the group to refocus on *E. coli* as the main host for heterologous expression by co-transforming the expression strains with the BtuB system, that actively transports cyanocobalamin across the cell membrane **[23]**. This mechanism bypasses the transcriptional mechanism that prevents the intracellular accumulation of B_{12} when supplemented in the media during growth. We used this methodology as a guide for our experiments with the self-sufficient catabolic RdhAs, but the previous success was apparently limited to isolated RdhA dehalogenase domains (such as the RdhAs from *Entotheonella* (candidatus) and *Burkholderiales joshii*), and the high expression yields of the NpRdhA holo-enzyme could not be reproduced.

In this context, the validation of both heterologous hosts for the expression of different B_{12}-dependent enzymes was still necessary, and the research interest of the group was focused on the self-sufficient RdhAs homologous to BhbA, as this system presented the most promising characteristics to allow the study of the electron transfer from NAD(P)H, via three Fe-S, to the corrinoid cofactor. This would eventually enable a more detailed kinetic study of these enzymes, not hindered by the constraints of utilising NpRdhA, in presence of excess exogenous redox partners like flavodoxin reductase and ferredoxins **[24]** or the methyl-viologen **[25]**.

Major adjustments in the construct design of a number of ssRdhA orthologous genes (such as the addition of solubility and affinity tags) had to be done to achieve our objective of optimising the production of the fully complemented holo-enzymes *in vivo*. Other

reductive dehalogenases that were expressed in *E. coli* frequently required (largely unsuccessful) reconstitution protocols, such as PceA **[26-28]** and CprA **[20]**.

When *B. megaterium* was used as the host for heterologous expression, most of our ssRdhAs constructs produced soluble and active protein. However, expression yields were low, and the samples could not be purified to homogeneity since the activity was lost after various purification steps. Working under anaerobic conditions or the addition of detergents/sugars/salts to the working buffer solution did not improve the yields, and it became obvious that it was an unsustainable strategy. Despite this, the dehalogenation activity could be measured in the crude extract and the semi-purified samples after two steps of ion-exchange and hydroxyapatite chromatography. The isolation of at least four orthologous ssRdhAs that had already been cloned and transformed into this organism was attempted in parallel through different purification techniques but achieving similar results **(Figure 2)**.

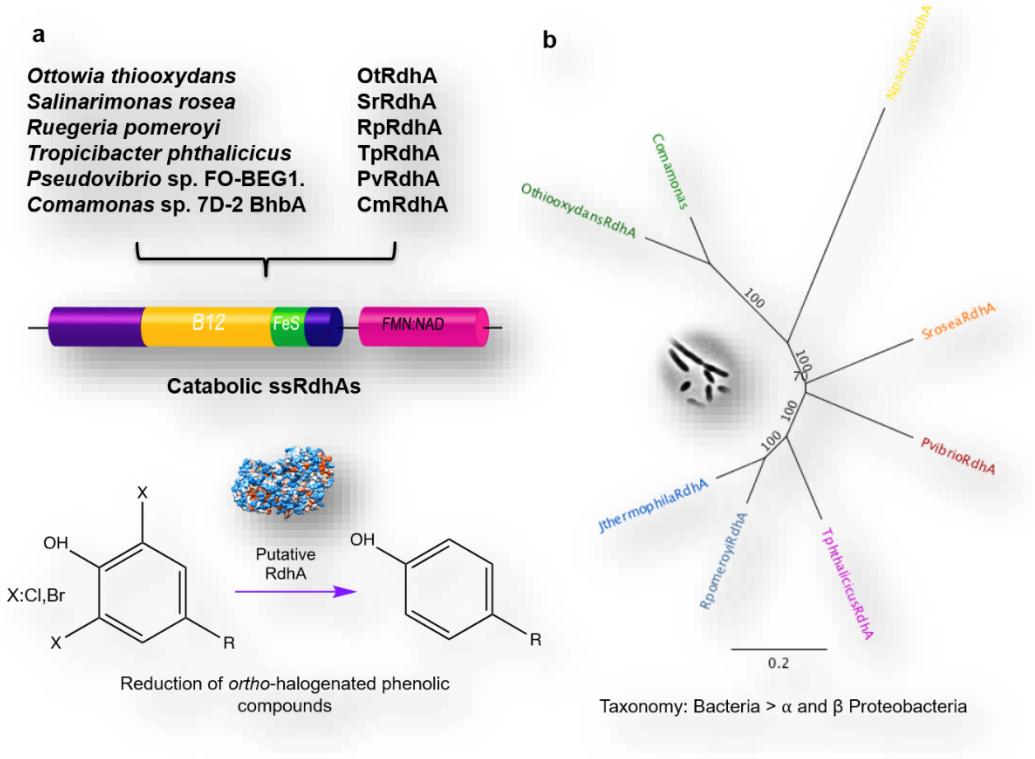


Figure 2. Orthologous self-sufficient RdhAs from the phylum Proteobacteria, that were selected for heterologous expression in *B. megaterium* and *E. coli*. All these homologous ssRdhAs can be grouped *ortho*-bromophenol RdhAs, similar to BhbA from *Comamonas* sp. 7D-2 and all present the same domain structure, showing the additional PDR-like reductase domain (a). Their phylogenetic relationship as orthologues is highlighted by an NJ-dendrogram, with the Bootstrap values showed at the nodes and expressed according to the number of replicates (b) (MEGAX®).

If *B. megaterium* is selected as prime heterologous host, more work needs to be done to identify a suitable purification strategy. Our work demonstrates that regardless of the affinity tag used, the purification via immobilised metal affinity chromatography (IMAC) or StrepTactin XT gravity-flow columns will not yield pure samples[1], even if the recombinant ssRdhAs can be confirmed to be in the crude extract both by Western Blot and matrix-assisted laser desorption ionisation time-of-flight mass spectrometry (MALDI-TOF). These results, together with our experience with the expression in *Vibrio natrigens* (BioCat) were very disappointing, although interestingly, very similar to those reported by Kunze *et al.* **[30]** when using *S. blattae* as heterologous host.

Better results were obtained when using *E. coli,* although the heterologous expression in this host has its challenges. Nonetheless, these strategies have started to demonstrate that it is possible to express complex metalloenzymes (as QueG and the naturally truncated NpRdhA) when co-expression of the BtuB transporter is ensured by co-transformation. even when the formation of apo-enzyme aggregates was observed previously when expressing recombinant fusion proteins in *E. coli* (such as the PceA enzyme linked to the trigger-factor) **[26]**. In this case, the changes made in the construct design, particularly the multiple affinity and/or solubility tags added to the full-length *rdhA* structural genes, allowed us to partially overcome both the solubility and low expression issues.

Even though a variety of expression vectors, strains, and induction strategies were undertaken, much more must be done to achieve a high yielding heterologous production system for the self-sufficient RdhAs found in Proteobacteria. After our experiences with numerous BhbA orthologous proteins, from marine α and β-Proteobacteria, we hypothesise that part of the solubility problems might be due either to misfolding caused by incomplete cofactor binding or because of the presence of highly disordered regions present both in the middle of the typical B_{12}/Fe-S dehalogenase domain and in the linker region between the N-terminal and the PDR-like reductase domains. These were found by a close inspection of the primary sequences and our homology 3-D models, generated by I-TASSER and evaluated by ROSETTA-design-HMMer (Rd.HMMer). These apparently unstructured regions could

[1] Although it is known that the excess of cobalamin disrupts Fe metabolism in *B. megaterium*, the addition of free cobalt and B_{12} in the media (at nM - μM concentrations) helps to increase the synthesis of corrinoid cofactors, and to repress the expression of the cobaltochelatase ($CbiX^L$), that often complicates the purification using Ni(II) based affinity chromatography, due to the fact that it contains regions with a high content of His residues **[29]**.

indicate the presence of yet unidentified protein-protein interactions in the physiological setting.

In further attempt to characterise the ssRdhAs system, we decided to truncate the *ssrdhA* genes to produce the isolated modules consisting either of the typical B_{12}/Fe-S dehalogenase domain or the PDR-like reductase domain. This allowed me to perform a partial characterisation of the reductase module. Our results suggest that these orthologous truncated PDR-like reductases rely on the electron transfer from pyridine nucleotides (NAD(P)H) in the form of a hydride ion (H^-) to allow the reduction of organohalides, similarly to other prototypical Fe-S flavoproteins that participate in the metabolism of inactivated aromatics, in respiration and photosynthesis **[31]**, commonly found fused to other protein architectures, probably as a result of gene fusion events driven by adaptative evolution **[32]**. Although the results presented in this work are limited to the OtPDR reductase (generated from the *ssOtrdhA*) and the non-self-sufficient NpRdhA we hypothesise that the electron transfer proceeds in a similar manner with the corresponding RdhA module and that this system can be illustrative of the catalytic behaviour of the full-length enzymes. We chose to work with NpRdhA due to its availability and the difficulties encountered when purifying the truncated dehalogenase domains. We demonstrated that the independent modules supported the dehalogenation of 3,5-dibromo-4-hydroxybenzoic at ratios of NpRdhA/PDR (1:1), as opposed to the assay developed by Collins *et al*. **[24]**, where substantial excess of the plant ferredoxin and flavodoxin reductase were required to drive the reduction reaction in presence of NADPH. We also produced an artificial fusion of the NpRdhA and one of the available PDRs, but sadly, although this construct was fully complemented and proved to be soluble, it was not active. We can speculate that this may be either a consequence of incorrectly fusing the PDR to NpRdhA in a way that there is no conformational freedom for the electron transfer to occur productively or, perhaps more likely because the particular reductase chosen during the construct design is not active on its own. Unfortunately, time was insufficient to perform more in-depth steady-state experiments. Our work shows that the truncated modules can represent more tractable systems than the full-length ssRdhAs, but more experiments are necessary to draw definitive conclusions about this fused Fe-S flavoprotein.

It is not surprising that the production of ssRdhA holo-enzymes is still the most significant challenge to overcome. Our initial results are encouraging and we trust that the heterologous expression systems will improve in the future and will enable us to meet the yield requirements necessary to perform the biophysical experiments. Despite the group

focus on structure/function relationships, a multidisciplinary approach will help link any current or future knowledge to the metabolic significance of reductive dehalogenation, and how this determines the ecophysiology of the OHRBs in diverse aquatic ecosystems (either marine, lentic or lotic), considering that most of them possibly present some degree of anthropogenic pollution with organohalides.

6 .3 Perspectives

The determination of the self-sufficient RdhA structures either by X-ray crystallography or cryo-electron microscopy (Cryo-EM) might play a crucial role in the future and must be considered a priority in the field of reductive dehalogenation. The structural information at atomic resolution will not only allow the comparison of these new architectures to the respiratory RdhAs, like PceA, but also shed light about the type of protein-protein interactions that exist between the dehalogenase domain and the PDR-like reductase domains of the ssRdhAs and maybe, their dynamic behaviour during the catalysis and the resting state. A proper kinetic characterisation of the ssRdhAs enzymes, combined with spectroscopic techniques such as EPR and standard and ultrafast UV/Visible spectroscopy would help us to identify the rate-limiting step during the dehalogenation reaction providing altogether a better understanding of the catalytic mechanism. This will ultimately allow unravelling of its individual steps and the efficiency of the electron-transfer processes from the simple Fe-S in the PDR-like domain to the cubane Fe-S and the cobalamin. However, any future work on the ssRdhAs requires further optimisation of the heterologous expression and purification protocols. The utilisation of a fermenter or bioreactor where the temperature, time, and aeration of the media can be controlled with precision will aid enormously to increase the reproducibility and help to scale-up the bacterial growths, guaranteeing better yields of the recombinant proteins.

Given the many challenges encountered when using the natural bacterial hosts, but particularly because of their slow-growing rates, complex nutrient requirements and their need for symbiotic relationships, the utilisation of most OHRBs for large-scale enzyme production is impossible at the moment. In addition to this, the lack of technologies to isolate and properly identify numerous OHRBs from the groundwater ecosystems and marine sediments and of high-throughput tools to culture these organisms in the laboratory implies that most of its gene ontology, physiology, and ecology remain unknown.

However, even if the research on reductive dehalogenation is still at a basic stage, the diversity of organisms from the phylum Proteobacteria means that there might also be a great variety of *rdhA* gene structures with different domain composition and metabolical or physiological roles. For this reason, determining the specific subcellular localisation of the ssRdhAs *in vivo* must also form part of the future scope, as this could clarify their metabolic role as cytosolic proteins or as part of a molecular assembly of multiple proteins, similar to the membrane-attached respiratory RdhA-complex **[17, 33]**.

In silico models could also be used as a guide for site-directed mutagenesis to refine or tune the activity of the RdhAs towards specific substrate targets or to develop new soluble and stable artificial architectures, emulating the diversity found in Nature **[17, 34, 35]**, where different combinations of the central B_{12}-binding domain of the dehalogenases plus different types of reductases and membrane anchors could be of help to explore the dynamics between the dehalogenase/reductase pairs, as we think that this will constitute a promising research target in the future.

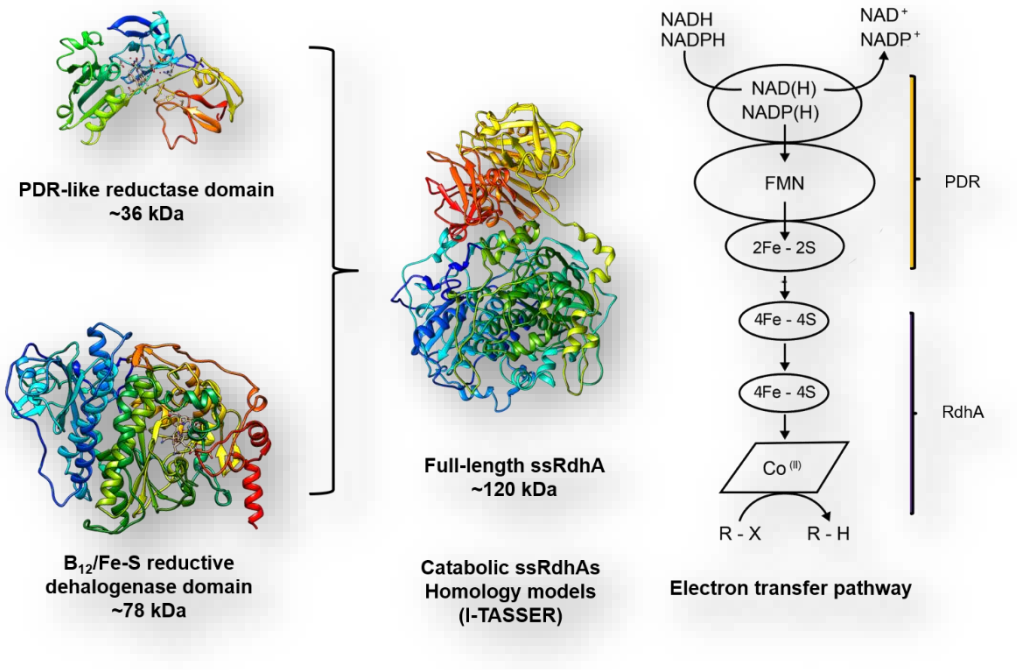


Figure 3. The self-sufficient reductive dehalogenases (ssRdhAs). 3-D homology models of the individual domains and the full-length RdhA of *Ottowia thiooxydans* generated by I-TASSER, evaluated, and further refined with the guide of the protocol ROSSETA-design-HMMer. In the diagram, we represent the electron transfer between the multiple cofactors, starting from the NAD(P)H substrate, through FMN, three different Fe-S and finally, the cobalamin and the aromatic halogenated substrate. **[32]**

A complete understanding of the biochemistry behind the reductive dehalogenation reactions will also allow the development of new strategies for the cell-free (or free-enzyme) bioremediation, as discussed by Scott *et al.* [36], by protein engineering approaches. Mutagenesis of key active site residues should be attempted if higher catalytic rates and a more specific/broader substrate scope are desired, however, the design of new and artificial enzymes for organohalide degradation with improved temperature, pH, and O_2 stability by the modification of motifs and even domains will also be an interesting challenge.

6.4 Closing remarks

What is the role of the catabolic RdhAs? Are they involved in cytosolic degradation pathways, or anchored to the membrane as part of a respiratory complex, as was suggested by Chen *et al.*? [37] The requirement for exogenous electron donors (such as H_2 or formate) and therefore the presence of the respiratory complex associates, such as HupL and the Mo-dependent enzymes will become redundant in case of the self-sufficient RdhA systems, as the presence of the PDR-like reductase module guarantees a direct link to NAD(P)H oxidation. Does this mean that there is no real need for interaction with other enzymes or redox partners? Interestingly, most of the dehalogenase operons in Proteobacteria lack the *rdhB* gene, as shown by Liu *et al.* [34]. Are these proteins located in the cytosol but interacting with the membrane, even when the RdhB is not present? Now it is clear that reductive dehalogenases are not just involved in organohalide respiration process and that many new architectures exist, apart from the self-sufficient RdhAs studied here, as is the case of the fusion protein between the RdhA and RdhB found also in some Proteobacteria as reported by Atashgahi *et al.* [38], that might also give testament to the diversity of organisms that not only populate the shallow seas but also freshwater ecosystems as rivers and lagoons. What does this tell us about the evolution of reductive dehalogenation metabolism?

One of the reasons why I decided to embark on this project, despite its challenging nature was because of its potential to develop bioremediation strategies for highly polluted sites. It is calculated that nearly 50 % of the rivers worldwide have been polluted by POPs, as a result of anthropogenic activity [39]. Ecosystem restoration projects are nowadays one of the top priorities in environmental sciences and a huge interest from local communities and researchers has arisen consequently, because of the ecological degradation of most surface-water systems and aquifers. Although the application of RdhAs-based catalysts in

bioremediation will not be achieved immediately, this remains our ultimate goal. It is possible that in the future the field may be revolutionised by the discovery of new enzymes that might be much more suitable for a complete functional/structural characterisation than our current systems. These might even be able to catalyse or be modified to catalyse the reverse reaction: oxidative halogenation, given its biocatalytic potential to produce valuable synthetic intermediates of industrial interest (due to the polarity of the C-X bonds) and in biomimetic chemistry.

Finally, even when bioremediation has driven research on reductive dehalogenation, its scientific value is not limited to its potential applications; the understanding of the catalytic mechanism and biochemical role of the RdhAs, as well as the determination of its structure, have a high intrinsic scientific value. If complemented with the identification of genomic sequences of OHR related genes of different phylogenetic origins, the research will enable the discovery of new enzymes with the potential to catalyse the dehalogenation of a new and/or a broader spectrum of organohalides and as a whole, this knowledge will provide us with a clearer picture of the evolution of the reductive dehalogenation metabolism and the ecophysiology of OHRBs. In turn, all this potentially will lead us to the optimisation of new expression methodologies in both the native organisms and other bacterial hosts.

6.5 References

1. S. Atashgahi, Y. Lu and H. Smidt, in *Organohalide-Respiring Bacteria*, eds. L. Adrian and F. E. Löffler, Springer, Berlin, 2016, ch. 5, pp. 63-105.
2. J. Maillard, W. Schumacher, F. Vazquez, C. Regeard, W. R. Hagen and C. Holliger, *Appl Environ Microbiol.*, 2003, **69**, pp. 4628–4638.
3. D. Leys, L. Adrian and H. Smidt, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120316.
4. C. Levy, K. Pike, D. J. Heyes, M. G. Joyce, K. Gabor, H. Smidt, J. v. d. Oost and D. Leys, *Mol. Microbiol.*, 2008, **70**, pp. 151-167.
5. A. Wagner, L. Segler, S. Kleinsteuber, G. Sawers, H. Smidt and U. Lechner, *Philos. Trans. R. Soc. Lond., B,*, 2013, **368**. 20120317.
6. S. P. Cohen, S. B. Levy, J. Foulds and J. L. Rosner, *J. Bacteriol.*, 1993, **175**, pp. 7856-7862.
7. R. Zhu, Z. Hao, H. Lou, Y. Song, J. Zhao, Y. Chen, J. Zhu and P. R. Chen, *J. Biol. Inorg. Chem.*, 2017, **22**, pp. 685-693.
8. T. Kumarevel, in *Antibiotic Resistant Bacteria - A Continuous Challenge in the New Millennium*, ed. M. Pana, Intech, London, 2012, ch. 16, pp. 403-418.
9. R. E. Richardson, *Curr. Opin. Microbiol.*, 2013, **24**, pp. 498-505.
10. A. Wagner, L. Adrian, S. Kleinsteuber, J. R. Andreesen and U. Lechner, *Appl Environ Microbiol.*, 2009, **75**, pp. 1876-1884.
11. L. Krasper, H. Lilie, A. Kublik, L. Adrian, R. Golbik and U. Lechner, *J. Bacteriol.*, **198**, pp. 3130-3141.

12. S. A. F. T. v. Hijum, M. H.Medema and O. P. Cuipers, *Microbiol. Mol. Biol. R.*, 2009, **73**, pp. 481-509.
13. M. N. Alekshun, S. B. Levy, T. R. Mealy, B. A. Seaton and J. F. Head, *Nature*, 2001, **8**, pp. 710-714.
14. R.-y. Wu, R.-g. Zhang, O. Zagnitko, I. Dementieva, N. Maltzev, J. D. Watson, R. Laskowski, P. Gornicki and A. Joachimiak, *J. Biol. Chem.*, 2003, **278**, pp. 20240-20244.
15. A. Grove, *Comput. Struct. Biotechnol. J.*, 2017, **15**, pp. 366-371.
16. S. P. Wilkinson and A. Grove, *Curr. Issues Mol. Biol.*, 2006, **8**, pp. 51-62.
17. T. Schubert, L. Adrian, R. G. Sawers and G. Diekert, *FEMS Microbiol. Ecol.*, 2018, **94**. fiy035.
18. R. Fernandez-López, R. Ruiz, F. d. l. Cruz and G. Moncalián, *Front. Microbiol.*, 2015, **6**. 648.
19. B.-E. Jugder, K. A. P. Payne, K. Fisher, S. Bohl, H. Lebhar, M. Manefield, M. Lee, D. Leys and C. P. Marquis, *ACS Chem. Biol.*, 2018, **13**, pp. 548-552.
20. A. Parthasarathy, T. A. Stich, S. T. Lohner, A. Lesnefsky, R. D. Britt and A. M. Spormann, *J. Am. Chem. Soc.*, 2015, **137**, p. 3525–3532.
21. B.-E. Jugder, H. Ertan, M. Lee, M. Manefield and C. P. Marquis, *Trends Biotechnol.*, 2015, **33**, pp. 595-610.
22. K. A. P. Payne, C. P. Quezada, K. Fisher, M. S. Dunstan, F. A. Collins, H. Sjuts, C. Levy, S. Hay, S. E. J. Rigby and D. Leys, *Nature*, 2014, **517**, pp. 513-516.
23. D. P. Chimento, A. K. Mohanty, R. J. Kadner and M. C. Wiener, *Nature*, 2003, **10**, pp. 394-401.
24. F. A. Collins, K. Fisher, K. A. P. Payne, S. Gaytan, S. E. J. Rigby and D. Leys, *Biochemistry*, 2018, **57**, p. 3493–3502.
25. P. Adriaens and A. L. Barkovskii, in *Encyclopedia of Environmental Microbiology*, ed. G. Bitton, Wiley-Interscience, New York, 2003, p. 509.
26. H. Sjuts, K. Fisher, M. S. Dunstan, S. E. Rigby and D. Leys, *Protein Expr. Purif.*, 2012, **85**, pp. 224-229.
27. A. Neumann, G. Wohlfarth and G. Diekert, *J. Biol. Chem.*, 1996, **271**, pp. 16515-16519.
28. A. Suyama, M. Yamashita, S. Yoshino and K. Furukawa, *J. Bacteriol.*, 2002, **184**, pp. 3419-3425.
29. S. J. Moore, M. J. Mayer, R. Biedendieck, E. Deery and M. J. Warren, *N. Biotechno.l*, 2014, **31**, pp. 553-561.
30. C. Kunze, M. Bommer, W. R. Hagen, M. Uksa, H. Dobbek, T. Schubert and G. Diekert, *Nat. Commun.*, 2017, **8**. 15858.
31. C. C. Correll, C. J. Batie, D. P. Ballou and M. L. Ludwig, *Science*, 1992, **258**, pp. 1604-1610.
32. G. T. Gassner, M. L. Ludwig, D. L. Gatti, C. C. Correll and D. P. Ballou, *FASEB J.*, 1995, **9**, pp. 1411-1418.
33. M. Bommer, C. Kunze, J. Fesseler, T. Schubert, G. Diekert and H. Dobbek, *Science*, 2014, **346**, pp. 455-458.
34. J. Liu and M. M. Häggbloma, *Mbio*, 2018, **9**. e02471-18.
35. M. Fincker and A. M. Spormann, *Annu. Rev. Biochem.*, 2017, **86**, pp. 357–386.
36. C. Scott, C. Begley, M. J. Taylor, G. Pandey, V. Momiroski, N. French, C. Brearley, S. E. Kostsonis, M. J. Seleck, F. A. Carino, C. M. Bajet, C. Clarke, J. G. Oakshot and R. J. Russell, in *Pesticide mitigation strategies for surface water quality*, eds. K. S. Goh, B. L. Bret, T. L. Potter and J. Gan, ACS Symposium Series, Washington, D.C., 2011, ch. 11, pp. 155-174.
37. K. Chen, L. Huang, C. Xu, X. Liu, J. He, S. H. Zinder, S. Li and J. Jiang, *Mol. Microbiol.*, 2013, **89**, pp. 1121–1139.
38. S. Atashgahi, Y. Lu, Y. Zheng, E. Saccenti, M. Suarez-Diez, J. Ramiro-Garcia, H. Eisenmann, M. Elsner, A. J. M. Stams, D. Springael, W. Dejonghe and H. Smidt, *Environ. Microbiol.*, 2017, **19**, pp. 968–981.
39. L. Z. González, personal communication.

Fin