Structural and Photocatalytic Studies with a High Activity Formate Dehydrogenase

Diogo Jorge Graça de Lemos Grilo

Thesis to obtain the Master of Science Degree in

Biotechnology

Supervisors: Dr. Inês Antunes Cardoso Pereira

Dr. Arsénio do Carmo Sales Mendes Fialho

Examination Committee

Chairperson: Dr. Leonilde de Fátima Morais Moreira

Supervisor: Dr. Inês Antunes Cardoso Pereira

Member of the Committee: Dr. Cristiano de Sousa Mota

September 2021

Preface

The work presented in this thesis was performed at Instituto de Tecnologia Química Biológica António Xavier, ITQB-NOVA (Lisbon, Portugal), during the period of September 2020 to July 2021, under the supervision of Dr. Inês Cardoso Pereira, and within the frame of the grant PTDC/BII-BBF/2050/2020. The thesis was co-supervised at Instituto Superior Técnico by Dr. Arsénio Fialho.

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Abstract

CO_2 emissions are an urgent problem facing humanity, with an unpredictable impact on world climate. Formate dehydrogenase FdhAB from *Desulfovibrio vulgaris* is highly active for CO_2 reduction. In this work, a simplified version of FdhAB comprising only the α-subunit (FdhA) was produced. However, it showed a significant decrease in activity for formate oxidation and CO_2 reduction (5.28 and 1.34 s^{-1}, respectively for FdhA compared to 961 and 234 s^{-1}, respectively for FdhAB). Circular dichroism analysis showed a different secondary structure profile for FdhA versus FdhAB, which combined with the decrease in melting temperature (78.9 to 66.3 ºC) lead us to believe that the enzyme is unstable, highlighting the structural importance of the β-subunit. Since photocatalysis is a promising application for FdhAB, a photocatalytic setup, using the complete light spectrum, with TEOA, TiO_2 and FdhAB was successfully developed, with high formate production and turnover (5.4 mM and 42 s^{-1} after 3 h). Furthermore, TiO_2 sensitized with Eosin Y enabling visible-light absorption had also a good performance (0.6 mM of formate and 5.1 s^{-1} after 3 h). A completely metal-free setup directly coupling Eosin Y to FdhAB was also shown to work, even though with lower performance. Computational design of improved forms of FdhA and FdhAB was attempted using PROSS, with several designs obtained that need to be experimentally validated, which could help solve the stability problems of FdhA and improve even further FdhAB, that would be of high biotechnological relevance.

Keywords: CO_2 Reduction; Formate dehydrogenase; Simplified Enzyme; Photocatalysis; Computational Design; PROSS

Resumo

Emissões de CO_2 são um problema urgente que a humanidade enfrenta, com um impacto imprevisível no clima. A formato desidrogenase FdhAB de *Desulfovibrio vulgaris* é altamente ativa para redução de CO_2. Neste trabalho, uma versão simplificada da FdhAB apenas com a subunidade-α (FdhA) foi produzida. Porém, apresentou uma diminuição significativa de atividade para oxidação de formato e redução de CO_2 (5,28 e 1,34 s^{-1}, respetivamente para FdhA comparativamente a 961 e 234 s^{-1}, respetivamente para FdhAB). Análise de dicroísmo circular mostrou um perfil de estrutura secundária distinto para a FdhA comparativamente à FdhAB que, combinado com a diminuição da temperatura de *melting* (78,9 para 66,3 ºC), leva-nos a acreditar que a enzima é instável, destacando a importância estrutural da subunidade-β. Como a fotocatálise é uma aplicação promissora para a FdhAB, um sistema fotocatalítico que usa o espectro completo da luz com TEOA, TiO2 e FdhAB foi desenvolvido, apresentando elevada produção de formato e *turnover* (5,4 mM e 42 s^{-1} após 3 h). Ademais, TiO_2 sensibilizado com Eosina Y, permitindo absorção de luz visível, também apresentou bom desempenho (0,6 mM de formato e 5,1 s^{-1} após 3 h). Uma variante completamente sem metais acoplando diretamente a Eosina Y à FdhAB também funcionou, embora com menor desempenho. *Design* computacional de formas melhoradas da FdhA e FdhAB foi tentado com o PROSS, obtendo-se vários modelos que precisam de validação experimental, podendo ajudar a resolver os problemas de estabilidade da FdhA e melhorar ainda mais a FdhAB, algo que seria de alta relevância biotecnológica.

Palavras-Chave: redução de CO_2; Formato Desidrogenase; Enzima Simplificada; Fotocatálise; *Design* Computacional; PROSS

Table of Contents

List of Figures

List of Tables

List of Abbreviations

IEA – International Energy Agency

IPCC – Intergovernmental Panel on Climate Change

UNFCCC – United Nations Framework Convention on Climate Change

CCS – Carbon capture and storage

NDCs – Nationally determined contributions

CCUS – Carbon Capture, utilization, and storage

NRC – National Research Council

FDH – Formate dehydrogenases

Mo/W-bisPGD – Molybdenum/tungsten-bis pyranopterin guanosine dinucleotide-containing enzymes

D. vulgaris – *Desulfovibrio vulgaris*

OAT – Oxygen atom transfers

HAT – Hydrogen atom transfers

LED – Light-emitting diode

SED – Sacrificial electron donor

TEOA – Triethanolamine

PROSS – Protein Repair One Stop Shop

PSSM – Position-Specific Scoring Matrix

1. Thesis Outline

This work was focused on studying the formate dehydrogenase (FdhAB) from *Desulfovibrio vulgaris* Hildenborough, which is capable of catalysing the two-electron reduction of CO_2 to formate at high rates. This thesis will be divided into two distinct parts: the first based on an experimental approach and the second on computational analysis.

The experimental part consists in:

(i) the generation of a simplified version of the enzyme to understand its reaction mechanism and to further elucidate structural and functional features;

(ii) comparing the functional and structural characteristics of the simplified version to the FdhAB;

(iii) testing different photocatalytic setups with the FdhAB.

The computational part consists in:

(iv) *in silico* stabilization of FdhAB and of the simplified version obtained.

Each of these topics will be further developed in the second section (introduction), with a more detailed explanation of the work and its objectives given in the third section (motivation and objectives). These two sections will be followed by three sections: results (fourth section), discussion (fifth section), and concluding remarks and future perspectives (sixth section).

2. Introduction

2.1. Context

In the next 50 years, two major energy-related problems confront humanity. On one hand, we have the high global demand for energy and the depleting fossil fuels reserves, which will eventually lead to higher costs[1,2]. On the other hand, the levels of carbon dioxide in the atmosphere are at a record high and due to its greenhouse effect, it is responsible for large and unpredictable consequences on the world climate[1,2]. The International Energy Agency (IEA) data on this matter exhibit the problem in a simple and clear manner. According to their website, the CO_2 emissions are steadily increasing and in 2018 have reached a value of 33 513 million metric tons[3]. As a result, the atmospheric concentration of CO_2 has risen from the preindustrial value of 280 ppm to 415 ppm at the end of the year 2020[4].

The Intergovernmental Panel on Climate Change (IPCC) publishes several reports to sensitize the public and the governments to the problem of CO_2 emissions and possible solutions. In 2018, a special report on the impacts of global warming of 1.5 °C above pre-industrial levels stated: "it was estimated with high confidence that anthropogenic global warming is currently increasing at 0.2 °C (likely between 0.1 °C and 0.3 °C) per decade due to past and ongoing emissions."[5].

For these reasons, global awareness regarding CO_2 emissions and its effects has risen during recent years, both within the governments and within the general public[5,6], which lead to efforts to reduce its environmental impact by utilizing both preventive and remediation techniques[6]. The first step was done in 1994 when the United Nations Framework Convention on Climate Change (UNFCCC) entered into force and bound members, as stated in Article 3, "to act to protect the climate system of the present and future generations of humankind, on the basis of equity and in accordance with their common but differentiated responsibilities and respective capabilities"[7]. Later on, in 1997 under the UNFCCC, 37 industrialized countries made a commitment to reduce greenhouse gas with agreed targets for each country that would totalize an average reduction of 5% emissions (Kyoto Protocol)[6,8], with the premise that emerging carbon capture and storage (CSS) techniques would be effective on minimizing CO_2 discharges[6]. However, the establishment of flexible market mechanisms, such as the international emissions trading (emissions could be traded/sold between countries), clean development mechanism, and joint implementation mechanism (allowed a country to, in essence, trade part of its emissions by the implementation of emission-reduction projects and emission-removal projects in other countries, respectively)[8], resulting in a flawed protocol that benefits countries that can "pay for their emissions".

In 2015, the Paris Agreement was signed under the UNFCCC to strengthen the global response to climate change. In summary, it was agreed to limit the temperature increase to 1.5 °C above pre-industrial levels and to undertake rapid reduction in greenhouse gas emissions, again implementing the principle of "common but differentiated responsibilities"[9]. Unlike the Kyoto Protocol, that only established emissions reduction targets for developed nations, the Paris Agreement unites all countries and each country can voluntarily set their emissions target (NDCs: nationally determined contributions)[8,9].

The latest emissions gap report written by the UN Environment Programme in 2019 estimates that considering current policies estimations, a global mean temperature rise of 3.5 °C will occur by 2100, considering implementation of the conditional NDC scenario a rise of 3.2 °C and considering implementation of both a conditional and unconditional NDC scenario a rise of 3.0 °C[10]. As stated by

the same report: "it is clear that neither current policies nor NDCs are adequate to limit warming to the temperature limits including in the Paris Agreement"[10].

Most recently, the European Union (EU) proposed the European Green Deal as an action plan to meet the goals set on Paris and making the EU carbon-neutral by 2050, decoupling economic growth from resource use[11]. For this, a European Climate Law was proposed and on 23rd October 2020, the European Council has reached an agreement on a partial general approach[12].

Despite the efforts to minimize CO_2 emissions, this is clearly not enough as tens of billions of tons are still released yearly, demanding a more aggressive approach to target this problem[3,10,13]. Thus, a report from the National Research Council (NRC) proposed CO_2 capture as a crucial technique to hinder emissions and urged for the need to rapidly develop negative-carbon technologies (that can remove CO_2 from the atmosphere)[6]. Therefore, in terms of policymaking, several governments are increasing their support and investing in carbon-neutral technologies and negative-carbon technologies, namely carbon capture and storage (CCS) and carbon capture/utilization and storage (CCUS)[14]. More recently, direct air capture (DAC) has gained a lot of attention in regards to public policy and among investors, being advertised as a big potential negative-carbon technology with the advantage of allowing utilization-post capture for the generation of added value products, such as fuel[6,14]. Thus, to achieve a carbon-neutral economy, capture and utilization of CO_2 is highly attractive[1,2,6].

Nevertheless, the thermodynamic and kinetic stability of CO_2 is high, making its activation difficult[2]. As stated in the 2008 report from the US Department of Energy on "Basic Research Needs: Catalysis for Energy": "The major obstacle preventing efficient conversion of carbon dioxide into energy-bearing products is the lack of catalysts"[15]. IEA states in their Energy Technology perspectives report (2015) that using hydrogen-based on low-carbon sources as energy carriers is an important option to reduce CO_2 emissions[16]. Furthermore, it states that the gap between the current CCS/CCUS technology and the willingness of governments to adopt and invest in it for hydrogen production is small compared to other sectors[16]. Thus, as stated by Voldsund *et al.* (2016): "hydrogen production with CO_2 capture could be a key transition technology for moving into a sustainable hydrogen-using society"[16,17].

2.2. Overview of the Chemistry and Biology of CO_2 Utilization

Two concepts can be defined when mentioning CO_2 emissions: point-source (large stationary sources, such as oil refineries, natural gas, and fossil-fuel power plants or small stationary sources such as small heating boilers used in the residential sector) and distributed sources (or mobile sources, associated to the transport sector, which account for 50% of all greenhouse gas emissions)[6,13,18].

The CCS approach essentially comprehends three distinct processes: CO_2 separation from industrial and energy-related sources, its transport to be stored, and its long-term isolation from the atmosphere[13]. When the CO_2 is utilized, the process is known as CCUS[14].

Several articles have been published in the last two decades that describe the CO_2 capturing technologies available and discussing is viability (for detailed information regarding this topic, consult [6,13,14,19–21]). The advantage of using these techniques is that the CO_2 captured can be used to generate added-value products, such as fuel[6,16]. CO_2 utilization is mostly divided into two general groups: direct use of CO_2 or conversion of CO_2 into chemical products. Direct CO_2 use has been widely applied in several industries such as soft drinks, food, agro-chemistry, welding, etc. However, most attention has

been directed towards the conversion of CO_2 into chemical and energy products. On this matter, the production of fuels by CO_2 hydrogenation has been gaining a lot of attention[22].

2.2.1. Chemistry of CO_2 Utilization

CO_2 is a highly stable molecule, being the end-product of any biological or chemical combustion reaction involving carbon-based fuels. It can be utilized in low-energy processes or high-energy processes. The first are mostly the ones where CO_2 incorporation into a compound occurs (carboxylation reactions), where energetically speaking, the process is favourable. The latter occurs in processes where CO_2 reduction is required, where energy supply is needed, and can be done in the form of electricity (electrochemical reduction), hydrogen (hydrogenation reactions), metals, or radiation (CO_2 conversion into CO and ½ O_2)[23]. This happens because CO_2 needs to be activated, mostly by bending the molecule in order to decrease the C-O bond orders and prepare the molecule for a two-electron reduction[1].

As seen in the previous section, anthropogenic CO_2 emissions are one of humanity's biggest challenges. Therefore, ways of tackling both the need for energy, in a sustainable manner as the world gets more overpopulated every day and reducing CO_2 emissions are extremely appealing. When referring to CO_2 hydrogenation, the intention is to recycle the emitted CO_2 to an energy-bearing/storing product, such as formic acid or methanol[17,22]. When comparing both products, while to produce formic acid there is a need for two hydrogen equivalents (eq. 1), with methanol three hydrogen equivalents are required, being water formed, resulting in a loss of one hydrogen equivalent (eq. 2).

$$CO_2 + H_2 \leftrightarrow HCOOH \qquad (eq.\,1)$$
$$CO_2 + 3H_2 \leftrightarrow CH_3OH + H_2O \qquad (eq.\,2)$$
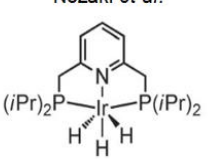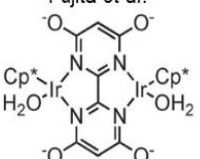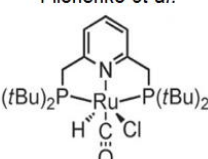
Even though methanol can store two equivalents of hydrogen, while formic acid only stores one, the latter has a 100% transfer rate, with no equivalents lost. Moreover, formic acid has several advantages, starting by being able to address the safety concerns that are typically associated with H_2 usage, being non-flammable and non-toxic[17,24]. Furthermore, it is liquid at room temperature, facilitating transportation and storage[25]. Nowadays, fuel cells capable of utilizing formic acid are already being developed, which means the molecule can be used directly to produce fuel. Another hypothesis is the dehydrogenation of formic acid on-site, which leads to the formation of hydrogen that can be used in hydrogen fuel cells. For these reasons and more, formic acid is recognized as one of the most promising hydrogen storing materials[17,24,25].

When reacting CO_2 and hydrogen (hydrogenation reaction) at high temperatures/pressures, the water gas-shift reaction occurs, resulting in the formation of carbon monoxide and water[17]. Therefore, in order to selectively generate formic acid a catalyst is required, which will improve the efficiency of the reaction and will lower the energy needs by reducing the energy of activation[17]. Afterwards, decomposition of formic acid can proceed by two pathways: generation of H_2 and CO_2 by decarboxylation or generation of CO and H_2O by decarbonylation (that should be inhibited)[17,25]. Therefore, for both the hydrogenation of CO_2 to generate formic acid and for the decarboxylation of formic acid to restore the H_2, suitable catalysis and control of the reaction conditions are demanded in order to increase the efficiency of the reactions by limiting unwanted pathways[17,25].

The hydrogenation of CO_2 with the use of homogenous catalysis has been the focus of research for several decades, dating back to the late 1970s with the use of triphenylphosphine (PPh_3) ligands with transition metals such as Ru, Rh, and Ir[17,24,25]. On the contrary, formic acid dehydrogenation is not as well studied, mostly due to the higher attractiveness of converting CO_2 into useful chemicals to be used in applications other than the energy sector[25]. Nevertheless, a great deal of literature focuses on this matter (for detailed information, consult [17,24]).

The problem is that even the systems that are considered state-of-the-art for reversible CO_2 hydrogenation reported until now (Figure 1) requires the use of organic solvents or additives for the catalysis to be efficient, which are not desirable when we seek to use the H_2 produced[19,26,27]. Besides that, even though the conditions used are not as extreme as the first catalytic systems described, it is still not possible to have an efficient catalyst under ambient conditions[24,28]. Furthermore, noble metals that are widely used for this application have a high cost. The application of non-precious metals (such as Ni, Fe, Co, and Mo) pioneered by Inoue with Ni complexes[29] has been investigated and while these systems are active, they have low performances[24,27]. Thus, all of these factors emphasize the need for the discovery of better and more efficient catalysts[24,30].

	Nozaki et al.	Fujita et al.	Filonenko et al.
Hydrogenation:			
T / P	200 ºC / 80 bar	80 ºC / 50 bar	120 ºC / 40 bar
solvent / base	H_2O, THF / KOH	H_2O / $KHCO_3$	DMF / DBU
TOF, h^{-1}	**150 000**	**54 000**	**1 100 000**
Dehydrogenation:			
T	80 ºC	90 ºC	90 ºC
solvent / base	$tBuOH$ / NEt_3	H_2O / HCO_2Na	DMF / NEt_3
TOF, h^{-1}	**120 000**	**228 000**	**257 000**

Figure 1: Most active homogeneous catalytic systems for reversible CO_2 hydrogenation. Taken from Filonenko et al. (2013)[27].

2.2.2. Biology of CO₂ Utilization

In biological systems, the fixation/conversion of CO_2 to organic carbon is crucial for life itself. There are in total six different metabolic pathways known to carry this process: the Calvin Cycle, the reductive citric acid cycle, the reductive acetyl-CoA or Wood-Ljungdahl pathway, the dicarboxylate/4-hydroxybutyrate cycle, and, lastly, the 3-hydroxyproprinate cycle. A detailed explanation of each pathway can be consulted in [31,32].

A significant research effort has been directed to the adoption of a single enzyme strategy, meaning the utilization of specific enzymes that are part of these pathways to capture/utilize CO_2[31]. A crucial class of enzymes are oxidoreductases (that catalyse the transfer of electrons from one molecule to another) and a lot of research has been directed to studying the ones able to transform CO_2 into

useful compounds, namely CO_2 reductases such as formate dehydrogenases (FDH), which catalyse the two-electron reduction of CO_2 to formate[31]:

$$CO_2(g) + H^+ + 2e^- \rightarrow HCO_2^-(aq) \qquad E° = -0.43\ V \quad (eq.3)$$
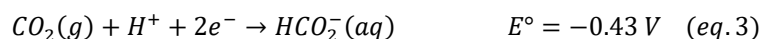
2.2.2.1.   Formate Dehydrogenases (FDHs)

Formate dehydrogenases can be divided into two major classes: metal-independent FDHs and metal-dependent FDHs[1,2]. The first group (metal-independent FDHs) can be found mostly in aerobic bacteria, yeasts, fungi, and plants[33]. The reaction catalysed by these enzymes is the oxidation of formic acid to CO_2 accompanied by the reduction of NAD^+ to NADH, therefore being of great utility for cofactor regeneration in processes of chiral synthesis[34]. Multiple alignment methods indicate that this group of proteins belong to the superfamily of D-isomer specific 2-hydroxyacid dehydrogenases, having relevant sequence similarities with D-lactate dehydrogenase, even though not being a hydroxy acid dehydrogenase[35,36]. The second group is exclusive of prokaryotes and the enzymes from this group belong to the superfamily of molybdenum/tungsten-bis pyranopterin guanosine dinucleotide-containing enzymes (Mo/W-bisPGD – Figure 2) and are of great biotechnological relevance due to their general ability to catalyse CO_2 reduction[30,37–41].
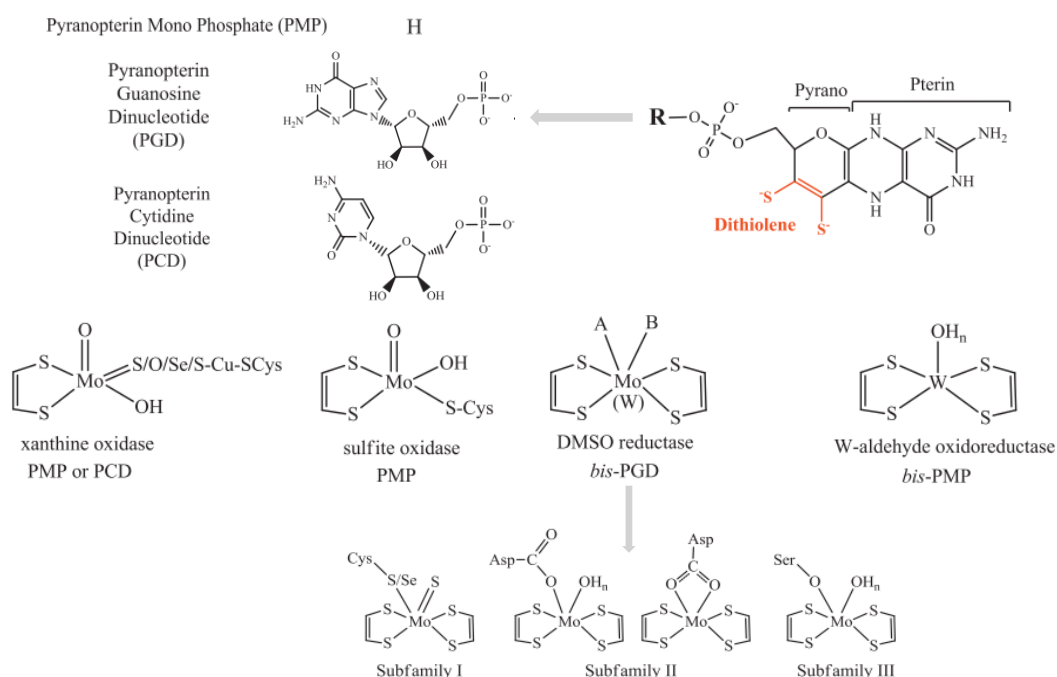
Figure 2: Representation of the families belonging to the Mo/W-bisPGD superfamily and their active site geometry, along with the chemical structure of the cofactors that they possess. The DMSO reductase family is further divided into three subfamilies depending on how the metal is coordinated in the two variable positions of the coordination sphere. Adapted from P.J. Gonzalez *et al.* (2013)[42].

This superfamily is divided into four large families: xanthine oxidases (XO), sulphite oxidases (SO), dimethyl-sulfoxide reductases (DMSO reductases), and W-aldehyde oxidoreductases[42]. These families are characterized by the presence of a mononuclear metal center (that can either be Mo or W) conjugated with a pyranopterin cofactor that can either be a pyranopterin mono phosphate (PMP), a pyranopterin guanosine dinucleotide (PGD), or a pyranopterin cytidine dinucleotide (PCD)[42]. The

remaining ligands do present some variability, but in all cases, the coordination sphere is completed with ligands with sulphur, selenium, and oxygen atoms. With the exception of the XO family, in the remaining proteins, the metal center is commonly coordinated by an amino acid of the polypeptide chain[42]. Due to the high variability among DMSO reductases, they are further divided into three subfamilies. FDHs belong to the subfamily I, being composed by Mo/W center coordinated to a cysteine (Cys) or selenocysteine (SeCys), a bis-PGD (typically referred in the literature as MGD – metal guanosine dinucleotide), and, lastly, to a sulphide group (=S/-SH).

Although the metal center of metal-dependent FDHs is conserved, its quaternary structure is quite diverse, varying highly in redox center composition such as iron-sulphur centers (Fe/S), which are extremely common, haems and flavins[43]. This is added to the variable metal center that can either be Mo or W, depending on the enzyme, some of which can even work with both. For instance, the FdhAB from *Desulfovibrio vulgaris* is able to incorporate either Mo or W, whilst $FdhABC_3$ from the same microorganism is only able to incorporate Mo[44].

Nonetheless, even though FDH belongs to the Mo/W-bisPGD superfamily, they are one of the only enzymes in this group that catalyse hydrogen atom transfers (HAT), while the remaining catalyse hydroxylation reactions, oxygen atom transfers (OAT), or hydration/dehydration reactions[42].

The subcellular location and physiological function of FDHs in prokaryotic organisms is vast. Generally, FDHs are used for two main goals: reducing CO_2 into formate (used in the C1 metabolism) or by coupling formate oxidation to the reduction of several terminal acceptors to obtain energy[33,43]. Metal-dependent FDHs are known for using several redox partners, such as cytochromes, ferredoxins, NAD^+, coenzyme $F_{420,}$ or membrane quinols, that work like mobile electronic wires that lead the electrons to accomplish their respective functions[40].

For instance, *E. coli* possesses three different FDHs: FDH-H (which is a component of the large formate hydrogen lyase complex – FHL – that works in anaerobic conditions), FDH-N (a membrane-bound complex that is induced by nitrate), and FDH-O (expressed in aerobic conditions). In the complex FHL, the cytoplasmic FDH-H is responsible for formate oxidation, with the electrons being then directed to a membrane-bound hydrogenase to generate hydrogen[33]. In the membrane-bound nitrate-formate respiratory system, the periplasmatic FDH-N oxidizes formate with the electrons further reducing the membrane quinone pool that will consequently reduce nitrate with the generation of proton motive force[45]. FDH-O is proposed to be part of another nitrate-formate respiratory complex that works similarly to the one expressed with FDH-N but is also expressed under aerobic conditions[45].

Sulphate-reducing organisms (SRO) such as the ones from the *Desulfovibrio* species can use formate as a reducing equivalent source for several pathways using different periplasmatic FDHs[45]. In this case, periplasmic *c*-type cytochromes are crucial in "transporting" the electrons generated by formate oxidation, with some FDHs having a dedicated cytochrome subunit such as the $FdhABC_3$ from *D. desulfuricans*[46] and *D. vulgaris*[44]. These electrons can be used by several membrane-bound respiratory complexes connected to nitrate reduction and sulphate reduction, highlighting the important role of formate as an energy conservation mechanism, especially in anaerobic organisms[46]. In microorganisms such as *Rhodobacter capsulatus* and *Cupriavidus necator* cytoplasmic metal-independent FDHs use the electrons from formate oxidation to regenerate NADH, contributing to the

important $NAD^+/NADH$ balance[45]. In addition, some FDHs physiologically perform CO_2 reduction, such as the ones from acetogens and syntrophs. The former are anaerobic bacteria that use the Wood-Ljungdahl pathway to convert CO_2 into acetic acid[47], while the latter produce formate and H_2 in a phenomenon referred to as "interspecies H_2-transfer"[37]. Interestingly, SRO are known to adopt a syntrophic lifestyle in the absence of sulphate, explaining its CO_2-reducing capabilities[39]. These groups of microorganisms are known to be superior CO_2-reductors, with hydrogen-dependent CO_2 reductases complexes (HDCR) being common in acetogens such as *Acetobacterium woodii* and *Thermoanaerobacter kivui*, where CO_2 reduction by a formate dehydrogenase is coupled to hydrogen production by a hydrogenase, and syntrophic microorganisms dynamically exchanging hydrogen equivalents by formate production through CO_2 reduction[30,37,41,48].

Structurally speaking, metal-independent FDHs are less diverse, forming mostly homodimers that include a NAD(H) binding pocket and a formate binding pocket, with both having several important conserved residues[36,43]. The metal-dependent FDHs are not structurally simple, and even though the metal coordination site is highly conserved, the same cannot be said for the composition of the protein, with different subunit compositions, resulting in heterogenous quaternary structures[43].

Initially, it was thought that only metal-dependent FDHs, which had tungsten in their active site were capable of CO_2 reduction. Right now, this is not the paradigm, since it has been shown that both Mo- and W-dependent FDHs and metal-independent FDHs can achieve CO_2 reduction, even though at very different rates. For this reason, metal-dependent FDHs are considered a far superior choice when compared to metal-independent FDHs, even with all the drawbacks associated with low-O_2 tolerance (such as protein purification and handling). Therefore, moving forward my focus will be on their structure and mechanism, since understanding the structural characteristics that can make an enzyme catalyse CO_2 reduction at the highest rates and understanding how the reaction is performed can help us design better catalysts for CO_2 reduction.

2.3. Metal-Dependent Formate Dehydrogenases

2.3.1. Structural Characterization

Metal-dependent FDHs range from the monomeric cytoplasmic FDH-H (α) possessing a molybdenum center and only one Fe/S cluster to the heteromeric FDH-N (($αβγ)_3$) with a molybdenum center, five Fe/S clusters, and two hemes, both in *E. coli*[45,49]. A degree of complexity above are the multimeric cytoplasmic NAD^+-dependent FdsABG from *C. necator* and RcFDH from *R. capsulatus*, both consisting in ($αβγ)_2$ heterotrimers with the α-subunit possessing the molybdenum center and five different Fe/S clusters, the β-subunit possessing another Fe/S cluster, the FMN cofactor and the binding site for NAD^+, and with the γ-subunit possessing an additional Fe/S cluster[50,51].

Interestingly, even though each protein has peculiar characteristics, metal-dependent FDHs are one example of the "puzzle-like systems" that redox proteins are, acquiring several different tools (subunits) that make them capable of performing their purposed function. Using this metaphor as a basis for the structural understanding of metal-dependent FDHs, the core piece of the puzzle, the one that cannot be missed in any structure is the α-subunit. This subunit possesses in all structures the conserved active site, with the dichotomies W/Mo and SeCys/Cys. From the structures of metal-dependent enzymes available in PDB, FDH-H and FDH-N from *E. coli* are Mo-SeCys enzymes[45,52],

FdsABG from *C. necator* and RcFDH from *R. capsulatus* are Mo-Cys enzymes[50,51], while both FDHs from *D. gigas* and *D. vulgaris* are W-SeCys FDHs[39,53]. It is worth noting, that in the case of FdsABG, the structure available is from the FdsBG subcomplex, without the catalytic subunit[51]. Even though no W-Cys FDH was structurally resolved until now, it is known that they exist with an example being the cytoplasmatic NAD$^+$-dependent FDH from *Methylobacterium extorquens*[54]. Furthermore, in the first characterized structure, FDH-H by Boyington *et al.* (1997), an oxo ligand was thought to complete the coordination sphere, with its presence mostly not considered when mentioning metal geometry because it was considered as highly labile[55]. In 2002, Jormakka *et al.* identified that the ligand was not an oxo group (=O) but a hydroxide group[45]. In the same year, the FDH from *D. gigas* was crystallized by Raaijmakers *et al.* and this study was the first suggesting that a sulphur ligand completed the coordination sphere of the metal center, instead of a hydroxide group[53]. This latter suggestion was supported by the identification of an *E. coli* and a *R. capsulatus* sulfotransferase that is responsible for the insertion of this ligand, which is essential for activity[56–58]. Nowadays, it is widely accepted that this ligand completes the coordination sphere in the active enzyme[2,39]. In 2006, Raaijmakers *et al.* re-analysed the crystallographic data from FDH-H and described a loop region (residues 138 to 146) that was mistraced on the original reduced state of the enzyme, suggesting that the selenium atom of SeCys$_{140}$ (part of the loop) was actually shifted away 12 Å from the molybdenum atom, with no connection at all with it[49]. This raised a question that is controversial still today: is there a displacement of the selenocysteine residue from the metal in the reduced state to allow catalysis, or does the metal maintain the coordination sphere intact between oxidized and reduced states? Nowadays, even though this question is still under debate, all the interpretations of available structures show the selenocysteine bound to the metal center[39,45,53,55]. It is worth noting that in the FdhAB from *D. vulgaris*, the loop (Ile$_{191}$-His$_{193}$) does suffer a small distortion, which results in the SeCys$_{192}$ being slightly shifted as an adjacent α-helix (Ser$_{194}$-Pro$_{198}$), but it remains bound to the metal[39]. Furthermore, previous experiments with iodoacetamide (capable of alkylating SeCys if the selenol group is free) that suggest that this dissociation occurs[59], was used by Raaijmakers *et al.* (2006) to support the hypothesis. However, recent studies with FdhAB from *D. vulgaris* show that iodoacetamide does not react with the selenocysteine, therefore no displacement of the selenocysteine occurs[39]. The original FDH-H structure and reinterpretation can be found in Figure 3A and the structures from FDH-N, FDH from *D. gigas* and FdhAB from *D. vulgaris* can be found in Figure 3B, Figure 3C and Figure 4, respectively.

According to a large phylogenetic study by Nielsen *et al.* (2019), the α-subunits of FDHs are structurally similar[60]. Moreover, the vast majority of the enzymes from the superfamily Mo/W-bisPGD, even though in variable size, possess the same base scaffold with four αβ conserved domains[61], with all the metal-dependent FDHs characterized until now possessing these four domains[39,53,55], with the exception of FDH-N which possesses an additional domain[45]. It is worth mentioning that this fifth domain is present in some Mo/W-bisPGD enzymes, however, it is not as conserved as the remaining domains, with its proposed role being to define the substrate-binding cavity[61]. In FDH-H, the first domain (I) coordinates the only Fe/S cluster close to the protein surface. The second and third domains (II and III, respectively) are responsible for binding the MGD cofactors, therefore having a classical dinucleotide binding fold. The selenocysteine that is coordinated to the metal center is part of a short loop at the NH$_2$-

terminus of domain III. The IV domain sits on top of the pterin segment of the cofactors, closing the structure. Therefore, the metal and its coordination sphere are in the middle where the four domains intersect, all of them interacting with the metal and the cofactors by a highly structured mesh of hydrogen bonds, salt bridges, and van der Waals interactions[55]. Regarding the FDH-N, the extra domain (V) makes the α-subunit of this protein larger than the one from FDH-H, with this domain being proposed to act as a binding site for an unknown regulatory molecule since it contained a hole connected to the active site cleft by a channel with positively charged residues, nonetheless, the core structure is similar to the one from FDH-H[45]. Even further evidencing the similarity between the α-subunits, when the α-subunit from *D. vulgaris* FdhAB is superimposed to the one from *D. gigas*, and FDH-H and FDH-N from *E. coli*, it showed a RMSD of 1.90 Å in all atoms[39]. All of these enzymes, showed a conserved [4Fe-4S] cluster binding domain ($CxxCxxCx_nCP$), with the presence of this cluster not only being characteristic of the α-subunit of FDHs but also for the majority of the enzymes from the Mo/W-bisPDG superfamily[61].

In this subunit, it is worth noting several conserved residues, such as a lysine that bridges the pterin of one of the MGD cofactors and the iron-sulphur cluster present in this subunit[39,45,53,55]. This suggests that the first MGD (bridging to the lysine) is associated with the electron transfer pathway, with some authors proposing that the second might modulate the redox potential of the metal[39]. The two well-conserved residues Arg_{333} and His_{141} (numeration from FDH-H), even though their exact role on catalysis differs from one mechanistic proposal to another, are known to be crucial and can be found on all characterized structures[39,45,53,55]. In the FDHs from *D. gigas* and *D. vulgaris* it is possible to find a disulphide bridge (Cys_{845}-Cys_{872}, numeration from *D. vulgaris*) located in a hydrophobic site close to the guanidine of the MGD that is not involved in electron transfer[53]. The opening of this bridge possibly allows the opening of the entry of the formate channel, which might explain the need for activation with a reducing agent in an anaerobic environment for them to gain full activity[39,53]. The authors proposed this as a mechanism to protect the enzyme to mild oxygen exposure, in contrast to FDH-H that is permanently inactivated[53].

The only structures crystallized in the presence of formate were FDH-H from *E. coli* and FdhAB from *D. vulgaris.* In the case of FDH-H, besides the analysis of the SeCys displacement by Raaijmakers *et al.* (2006), the analysis by Boyington *et al.* (1997) identified a slight rotation of one of the pterin portion of one of the MGD cofactors, and with a slight overall increase in bond lengths between the metal and the ligands in the oxidized state. The increase in bond length between the Mo-$SeCys_{140}$, results in the movement of the ligand allowing His_{141} and Arg_{333} to interact indirectly by one water molecule. In the reduced state, formate was proposed to bind the metal as nitrate, causing the displacement of this water molecule. The alpha proton of formate would be located near the selenium of $SeCys_{140}$, which was proposed to be oriented by a positive charge and a hydrogen bond from Arg_{333}, therefore being the sidechain of His_{141} "moved" to the other side of the selenium atom[55]. In the case of FdhAB from *D. vulgaris*, the shift in the $SeCys_{192}$ mentioned above results in the shifting of the imidazole ring of the conserved His_{193} away from $SeCys_{192}$ and the establishment of a hydrogen bond with Gly_{442}, resulting in the conserved residue (Arg_{441}) suffering a small rearrangement. On the other hand, the MGDs cofactors suffer conformational changes as well, with the formation of new hydrogen bonds between

neighbouring amino acids (Gln$_{890}$ bind to Glu$_{433}$, with the former stabilized by a water molecule that is not present in the oxidized state, and that is also H-bonded to the second MGD cofactor)[39].
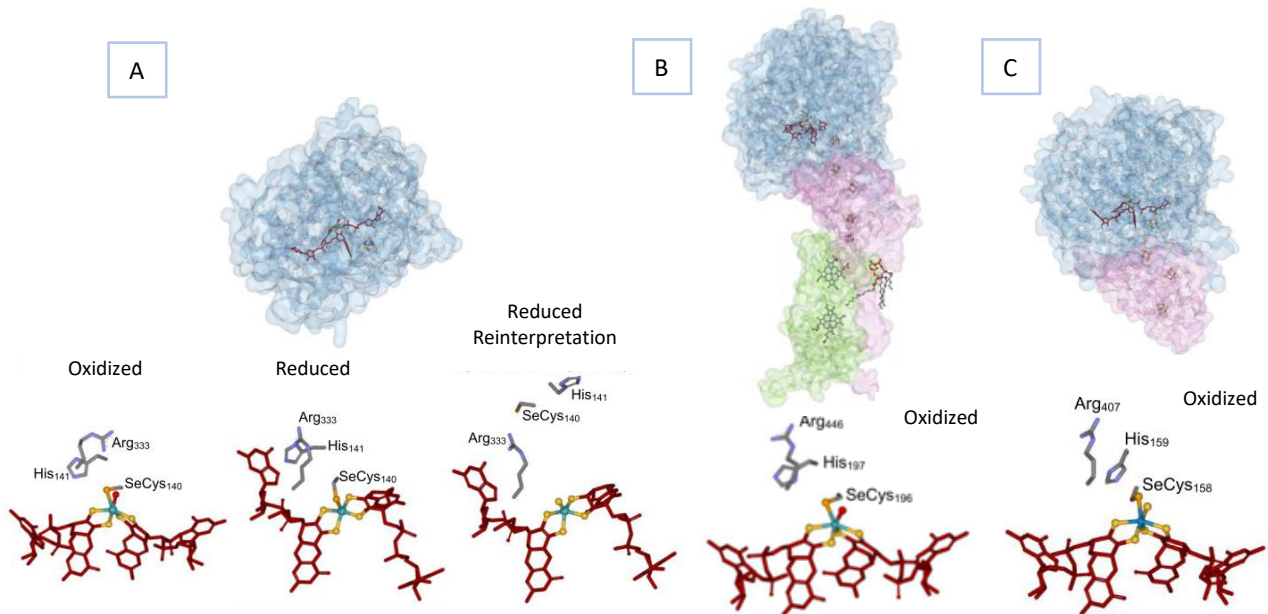
Figure 3: (A) Structure of formate dehydrogenase H from *E. coli*, with the representation of the metal first coordination sphere and highly conserved residues arginine and histidine in oxidized and reduced state (as described by Boyington *et al.* in 1997 and by Raaijmakers *et al.* in 2006). **(B)** Structure of formate dehydrogenase N from *E. coli*, with the representation of the metal first coordination sphere and highly conserved residues arginine and histidine in the oxidized state. **(C)** Structure of formate dehydrogenase from *D. gigas* with the representation of the metal first coordination sphere and highly conserved residues arginine and histidine in oxidized state. Adapted from Maia *et al.* (2016)[40].

The β-subunit exists in the majority of metal-dependent FDHs[39,45,53], even though some enzymes only possess the α-subunit such as the FDH-H[55]. This subunit commonly harbours multiple iron-sulphur clusters, being referred to as an electron transfer subunit. In the case of FDH-N has four conserved cysteine motifs for [4Fe-4S] cluster binding (CxxCxxCx$_n$CP), which are divided into two subdomains (the first near the surface possessing FeS-1 and FeS-3 and the second buried in the protein and possessing FeS-2 and FeS-4, with different hydrogen bonds supporting them, which are thought to influence their redox potentials[45]. In the β-subunit of the FDH from *D. gigas*, it is observable the presence of three [4Fe-4S] clusters and the space and folding to accommodate another cluster if Val$_{116}$ were mutated to Cys, changing this site to the typical cysteine motif which is highly conserved in FDHs[53]. In FdhAB from *D. vulgaris*, the β-subunit possesses three [4Fe-4S] cluster and an overall structure similar to the ones from *D. gigas* and *E. coli*[39]. The α- and β-subunits interact by a mesh of hydrogen bonds, salt bridges, and hydrophobic interactions[39,45,53]. The major differences in terms of subunit interaction are the interactions between the N-terminal of the α-subunit of *D. gigas* that protrudes through the β subunit[53] and in the case of the FdhAB from *D. vulgaris* the loop 58-66 in the β-subunit is one residue longer than in *D. gigas*, allowing an extra hydrophobic interaction with the α-subunit[39]. It is worth noting that most of the Mo/W-bisPGD enzymes also possess this typical β-subunit with ferrodoxin-type domains that hold the iron-sulphur clusters, although rare "exotic" β-subunits exist that carry rieske-type domains (typically harbours [2Fe-2S] clusters) and *c*-type cytochromes or FAD-binding domains[61].

Recently, the structures of the β-subunits from *C. necator* and *R. capsulatus* were resolved, in which it was observable the presence of more "exotic" β-subunits, with both of them harbouring besides the ferredoxin-type domain that binds one [4Fe-4S] cluster, domains for NAD^+ binding and FMN binding[50,51].
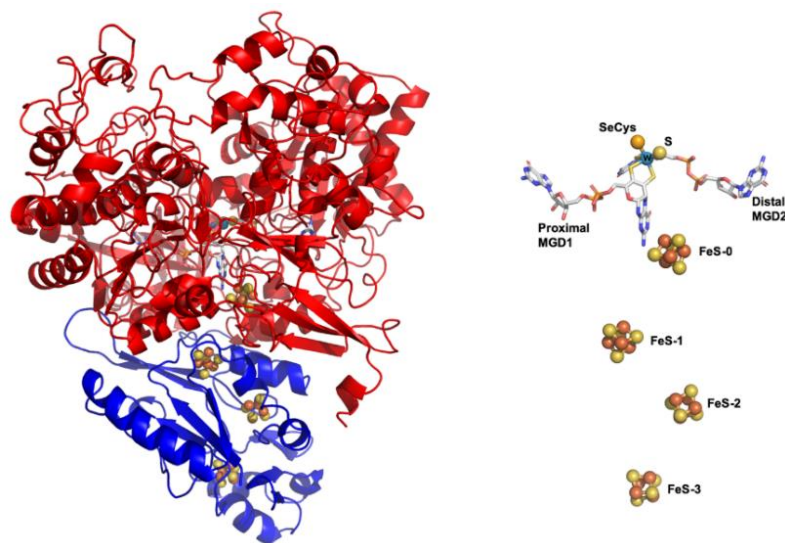
Figure 4: In the left we have the structure of the enzyme FdhAB from *Desulfovibrio vulgaris*, with the alpha subunit being represented in red and the beta subunit in blue. In the right we have the metal first coordination sphere and the "electron transfer apparatus" of FdhAB. Taken from Oliveira *et al.* (2020)[39].

The γ-subunit is the most variable structural element in FDHs, being even absent in some. Furthermore, in the case of membrane complexes, typical the subunit that anchors the protein to the membrane is the γ-subunit. A typical "puzzle piece" of a γ-subunit is a cytochrome[62]. For instance, the γ-subunit of FDH-N is a membrane-bound cytochrome b that contains four transmembrane-helixes and a menaquinone binding site, making the γ-subunit responsible for connecting the soluble α- and β-subunits of FDH-N to the quinone pool[45]. In this case, the γ-subunit is bound to the β-subunit by several hydrogen bonds, being the electrons transferred from the last [4Fe-4S] cluster of the β-subunit to the heme b_p (periplasmic), which transfers them to heme b_c (cytoplasmic) and subsequently to the menaquinone binding site. These 3 groups are coordinated to the polypeptide chain by hydrogen bonds, with several histidine residues playing a part in this connection[45]. The presence of this type of membrane-bound γ-subunit can be considered as an evolution of prokaryotic redox systems to couple "free" cytoplasmic and periplasmic enzymes with larger membrane complexes capable of performing functions such as quinol pool reduction, with the pools being directly involved in proton motive force generation, acquiring a crucial role in prokaryotic bioenergetics[2,45,61]. It is worth noting that even though the γ-subunit, in this case, is a membrane-bound subunit that functions as an interface between formate oxidation and quinone pool reduction, other types of γ-subunits exist. For instance, the γ-subunit from the FDH from *C. necator* is a small subunit containing one [2Fe-2S] cluster[51].

The active site tends to be far away from the surface of the proteins, therefore channels for substrate and product trafficking must exist in all the structures. This channel can be larger or smaller depending on how buried the active site is. For instance, in *D. gigas*, the active site is 25 Å away from the surface[53], while in *D. vulgaris* it is 30 Å away[39]. When observing the tunnels for formate and CO_2, the former is positively charged, with several conserved residues, whilst the latter is a hydrophobic

tunnel, however, the degree of conservation is lower[39]. In FdhAB from *D. vulgaris*, the formate channel has a minimum diameter of 1.4 Å and length of 13.2 Å and the CO_2 channel has a minimum diameter of 1.8 Å and length of 39.7 Å (Figure 5). Curiously, in *D. vulgaris*, the conformational change discussed in His_{193} from the oxidized to the reduced form of the enzyme, affects the tunnels, with the amino acid blocking and unblocking them, acting like a switch[39].
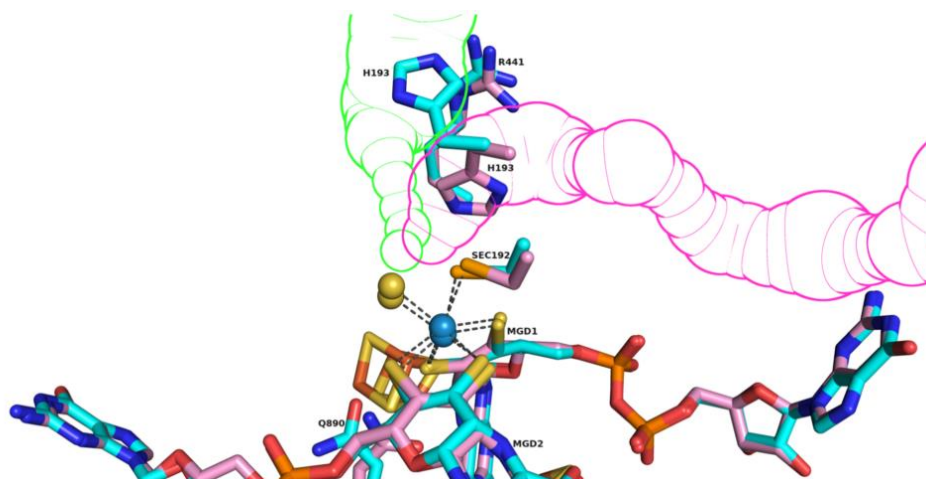
Figure 5: Representation of the active center of FdhAB from *D. vulgaris* with the proposed formate and CO_2 tunnels (in green and pink, respectively) by Oliveira *et al.* in 2020. Taken from Oliveira *et al.* (2020)[39].

In summary, we can look at metal-dependent FDHs as the cell choosing subunits from a toolbox kit to fulfil a given function. This characteristic, that renders them high adaptability, is not only present in metal-dependent FDHs, but also in other enzymes from their superfamily (Mo/W-bisPGD), as described in a review by Grimaldi *et al.* (2013)[61].

The genes encoding these subunits are organized mainly in operons with the diversity in the structure being predicted by the gene organization. Some authors like Nielsen *et al.* (2019) even proposed a classification scheme based on gene organization, cofactor content (which appears to be correlated with gene organization as well), and chaperone genes that are proposed to be necessary for assembly and maturation of the enzyme[60].

2.3.2. Mechanistic Characterization of Formate Dehydrogenases

Initially, the main question was whether this class of FDHs followed the typical oxygen atom transfer mechanism seen in other members of the Mo/W-bisPGD superfamily or if it followed a hydrogen atom transfer. For this, in 1998, Khangulov *et al.* studied the oxidation of ^{13}C-labeled formate in ^{18}O-enriched water using FDH-H from *E. coli*, to see if the CO_2 generated would incorporate the labelled oxygen and proved that it does not happen. The incorporation of oxygen from the water occurs in OAT reactions, therefore, since no incorporation is seen, the enzyme catalysed a HAT reaction[52].

Nonetheless, in 1997, Boyington *et al.* had already proposed a mechanism that potentially could involve a hydride transfer reaction[55]. In its mechanistic proposal, it was assumed that the hydroxyl ligand (which as previously discussed, was thought to complete the coordination sphere of the molybdenum center) was labile and formate would replace it on the coordination sphere, being this interaction stabilized and oriented by Arg_{333} and His_{141} through hydrogen interactions. Subsequently, formate was

oxidized, which was proposed to occur by a direct two-electron transfer through the oxygen of formate to the metal center or by a direct hydride transfer to the metal center. In either case, the α-proton of formate is then given to the SeCys$_{140}$ and then to the His$_{141}$, generating the reduced metal center Mo(IV). Afterwards, the electrons are transferred through the conserved lysine residue (Lys$_{44}$) to the first iron-sulphur cluster one by one (ping-pong mechanism). The first oxidation (Mo(IV) to Mo(V)) leads to hydrogen bonding between the proton given to His$_{141}$ and SeCys$_{140}$, while after the second oxidation (Mo(V) to Mo(VI)) the hydrogen bond between the SeCys$_{140}$ and His$_{141}$ breaks and the proton is released to the solvent. Afterwards, an OH$^-$ (from water) is bound to the metal coordination sphere to regenerate the complex. Several aspects of this proposal were supported by experimental evidence, such as the role of selenocysteine on catalysis and the histidine involvement in catalysis due to the known pH dependency in catalytic activity[63]. The one electron at a time re-oxidation of the Mo center was supported by the identification of a transient [Mo(V),4Fe-4S$_{REDUCED}$] in EPR experiments[52,64].

In the report of the crystallographic structure of FDH-N described by Jormakka *et al.* (2002), the authors proposed that a histidine residue (His$_{197}$ in FDH-N) was directly involved in the α-proton removal rather than the selenocysteine residue (SeCys$_{196}$ in FDH-N), due to slight changes in the orientation of the histidine residue in this structure comparing to the structure of FDH-H[45]. The direct involvement of the histidine residue is supported by EPR studies, where it seems that the energy released upon the *in vitro* oxidation of the Mo(IV) center by the [4Fe-4S] is used for deprotonation of the histidine residue and transfer of the proton to the solvent against the thermodynamical potential (due to the decrease of pK$_A$ of the residue upon metal oxidation)[52].

In 2006, with the reinterpretation of the crystal structure of FDH-H by Raaijmakers *et al.*, a different mechanism was proposed (Figure 6). This was due to the discovery of the mistraced loop and the proposed displacement of SeCys$_{140}$ in the reduced structure, as previously mentioned. Therefore, in this case, formate was proposed to bind directly to Mo (as seen for the previous mechanisms), by substitution of the SeCys$_{140}$ (resulting in a 9 Å displacement of this residue compared to the oxidized structure)[49]. The selenium atom of the SeCys$_{140}$ would be stabilized by Arg$_{333}$. Therefore, the Arg$_{333}$ residue was proposed to stabilize and orient formate in the oxidized state of the enzyme (Arg$_{333}$ points to the active site) and to stabilize the selenium atom in the reduced state (Arg$_{333}$ points away from the active site)[49]. The α-proton of formate would be abstracted by SeCys$_{140}$ being then transferred to His$_{141}$ (as proposed by Boyington *et al.* in 1997)[49,55]. Afterwards, CO$_2$ is directly released or, alternatively, would generate a selenium carboxylated intermediate before release (with the decarboxylation being catalysed by amino acids near, such as Arg$_{333}$, Arg$_{138,}$ or His$_{141}$)[49]. The mechanism proposed by Raaijmakers *et al.* (2006) was supported by several studies. The crucial role of SeCys in its unbound form was supported by studies of inactivation with iodoacetamide (which reacts with ionized selenol and thiol groups), as previously mentioned, showing that loss of activity only occurs when formate is present[63]. Raaijmakers *et al.* (2006) suggests that in this the loop containing the SeCys is solvent-exposed, in the reduced form, making the selenium atom more accessible for alkylation[49]. However, the same experiment in *D.* vulgaris revealed opposite results, showing that even in the presence of formate, the alkylation of SeCys could not be detected[39]. Furthermore, a computational study compared the mechanisms proposed by Boyington *et al.* (1997) and Raaijmakers *et al.* (2006) in order to determine

the most thermodynamical stable mechanism and found that the activation barrier involved in the mechanism where $SeCys_{140}$ is bound is higher than the one where $SeCys_{140}$ is displaced, therefore stating that only the displacement of the selenocysteine is thermodynamically favourable[65]. However this study only focused on the rate-determining step (the proton abstraction from formate), neglecting the influence of the total process (substrate binding, the influence of the histidine and arginine residues and the process release)[66].
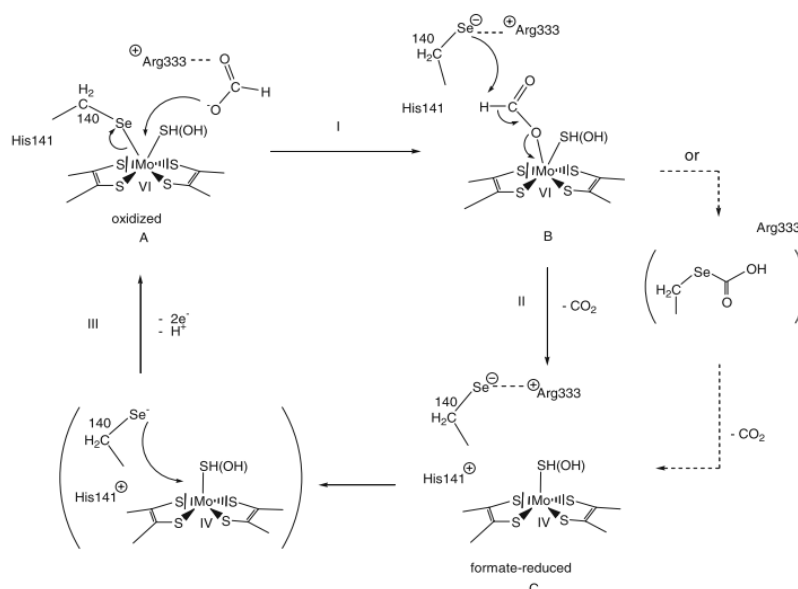
Figure 6: Mechanism proposed by Raaijmakers *et al.* (2006). A: oxidized state of the enzyme; B: intermediate state where the formate binds to the metal center and SeCys is stabilized by the arginine residue; C: enzyme in the formate reduced state. Taken from Raaijmakers *et al.* (2006)[49].

In 2012, Tiberti *et al.* studied, computationally, the ability of several groups to act as proton/hydride acceptors (the selenium atom of SeCys, one of the thiolenes from MGD and the metal itself) considering both the mechanism where the -SH group was substituted by formate in the coordination sphere and the mechanism of SeCys displacement proposed by Raaijmakers *et al.* (2006). In the second mechanism, it also studied the ability of the -SH group to act as an acceptor. With the results, the authors suggested, based on energy consideration, that the mechanisms featuring a direct abstraction of the α-proton of formate by the metal center were energetically more favourable and even more considering the mechanism of SeCys displacement. In this case, the proposal is that the metal would abstract the hydride from the formate molecule, while the displaced SeCys would abstract the proton from the -SH ligand, which renders the sulphur ligand capable of a proton shift with the metal center, leading to the formation of a pentacoordinated Mo(IV) with -SH ligand in an apical position[67].

Mota *et al.* (2011) and Cerqueira *et al.* (2013) proposed a different mechanism named the "sulphur shift"[68,69] (Figure 7). In this mechanism, the active site would require activation (the sulphur shift). In summary, the mechanism proposes that in the inactive state there is a semi-covalent bond between the sulphur atom and the selenium atom, both being coordinated to the metal, which does not allow the catalysis to occur. However, when formate is present, it will induce the displacement of the SeCys from the metal center (first coordination sphere) to the second coordination sphere (selenium shift), being covalently bound to the sulphur atom which moves and occupies the position of the

dissociated residue (sulphur shift), opening a vacant coordination position where the substrate can bind[69]. This inactive state was proposed to be associated with the aerobic purification of an enzyme that works under anaerobiosis, which is further supported by the need of incubating the enzyme with the substrate previously to the addition of the electron acceptor to obtain maximum activity. Upon substrate binding, the enzyme regains its hexacoordinated geometry with the last ligand being the oxygen of the formate molecule, with the conserved arginine residue stabilizing the substrate prior and upon binding to the metal center. Furthermore, the bond between the selenium and the sulphur group breaks, forming a selenide anion that is stabilized by the conserved histidine residue.
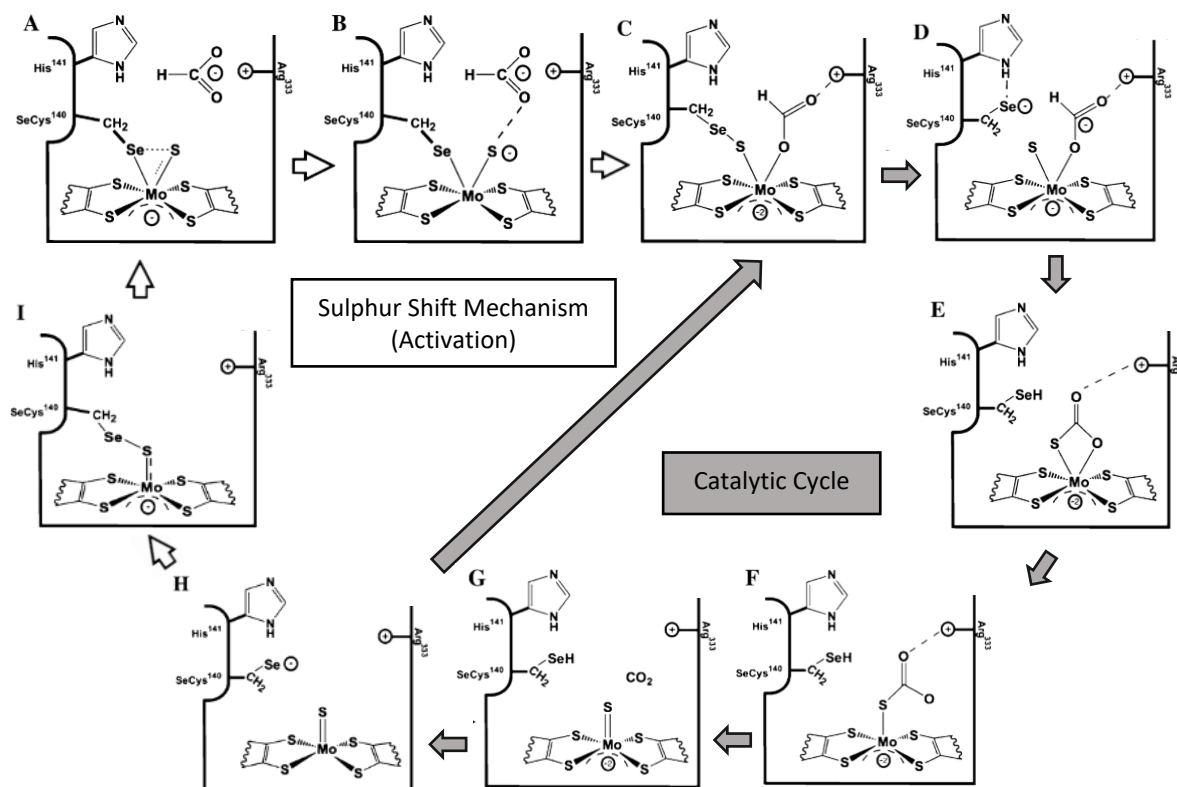
Figure 7: Representation of the mechanistic proposal by Moura *et al.* (2011, 2013). In A, B, C and I we have the metal activation mechanism (sulphur shift), while in D, E, F, G and H we have the catalytic cycle. Adapted from Mota *et al.* (2011)[68].

This mechanism proposes a transfer of 2 electrons to the metal center and the abstraction of the proton by the formed selenide anion, rather than the hydride transfer mechanism. Afterwards, the metal suffers oxidation (by transfer of the electrons to the electron transfer chain), which promotes the release of CO_2. Then two pathways are possible: in the absence of formate, the enzyme returns to the initial state with both the sulphur and the selenium atoms bound to the metal center, or in the presence of formate the catalytic mechanism proceeds without the formation of the hexacoordinated initial complex (Figure 7).

In 2016, Maia *et al.* reanalysed the "sulphur shift mechanism" and pointed out that it would not explain the enzyme being activated by methyl viologen as seen by the group (since the molecule would not enter the active site), nor it would explain the inability of CO_2 to activate the enzyme (knowing that it acts as a CO_2 reductase as well)[40]. Most importantly, the mechanism does not give any catalytic role to the sulphur group that remains bound to the metal center, which is suggested to be crucial for activity

due to FDH being inactivated by cyanide[40] and the need for sulfotransferases for the enzyme to be active[56–58]. Additionally, a study by Niks *et al.* proposed that the hydrogen atom from formate is solvent-exchangeable and is transferred to the molybdenum center, therefore stating that a hydride transfer mechanism occurs with the sulphur group of the metal center acting as the hydride acceptor[70]. Moreover, the pK$_a$ values of the sulphur atom when associated with the metal in the oxidized and reduced form change drastically, with higher oxidation states having a deprotonated sulphur ligand (Mo(VI)=S) and lower oxidation states having a protonated sulphur ligand (Mo(IV)-SH), explaining the suitability of this group to act as both an acceptor and a donor hydride, in formate oxidation and carbon dioxide reduction, respectively[40].

Henceforth, Maia *et al.* suggested that formate is oriented towards the active site through the arginine residue and possibly histidine (through hydrogen bonds to the oxygen atoms of the formate molecule), followed by a direct hydride transfer to the sulphur ligand of the oxidized metal center (Mo(VI)=S), generating the reduced metal and protonated sulphur group (Mo(IV)-SH) and CO$_2$ (which is eventually released). Furthermore, the electrons are transferred to the iron-sulphur clusters regenerating the oxidized center, which favours the deprotonation of the sulphur ligand (Figure 8A). The carbon dioxide reduction follows the same steps but in reverse, meaning that the metal starts in the reduced state with the protonated sulphur ligand, followed by a direct hydride transfer to the CO$_2$ molecule resulting in the oxidation of the metal center and deprotonation of the sulphur, which is further regenerated by the molecule used as an electron donor (Figure 8B).
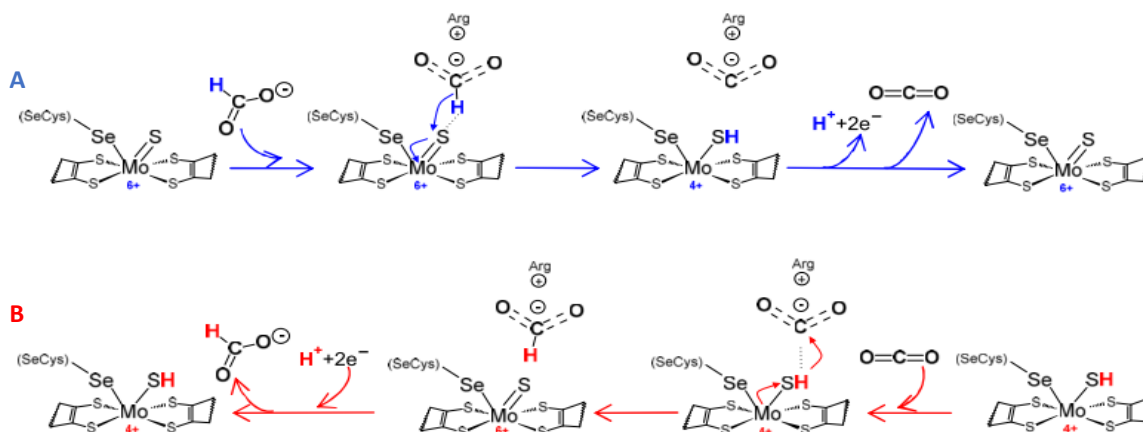
Figure 8: Mechanistic proposal for formate oxidation (A) and for carbon dioxide reduction (B) by formate dehydrogenase proposed by Niks *et al.* (2016) and Maia *et al.* (2016). Taken from Maia *et al.* (2016)[40].

In this mechanism, as can be seen in Figure 8, no alterations to the metal geometry occur (hexacoordinated throughout the reaction), with the importance given to selenocysteine being to hinder oxygen atom transfers (characteristic of the enzymes' present in the superfamily) and to increase the covalency of the Mo-S bond and/or modulate the metal reduction potential[40].

In 2018, Dong *et al.* compared the reaction mechanisms proposed by Boyington *et al.* (1997), Raaijmakers *et al.* (2006), Mota *et al.* (2011) and Cerqueira *et al.* (2013), Tiberti *et al.* (2012), and Maia *et al.* (2016) through computational methods (QM/MM approach) to understand which mechanism is more likely to happen[71]. Their results indicate that formate does not bind directly to the metal center (which was proposed in all mechanisms[55,67–69], except for the last[40]) since it would require the

displacement of either SeCys[49,67–69] or =S/-SH[55]. Recently, the crystallization of formate dehydrogenase from *D. vulgaris* in the presence of formate supports the proposal that the metal coordination sphere is not altered[39]. Therefore, Dong *et al.* study support the mechanism proposed by Maia *et al.* (2016) that suggests formate binding in the second coordination sphere, with the hydride transfer occurring to the sulphur ligand (=S/-SH) and the CO_2 being released only when the metal is fully oxidized (Mo(VI)).

In summary, currently, no consensus exists on the mechanism of metal-dependent FDHs, with several mechanistic proposals, but all with their limitations. The main issues under debate are: (i) whether the selenocysteine/cysteine residue present in the first coordination sphere of the metal dissociates during the reaction, yielding a pentacoordinated metal[49,67–69] or if the metal remains hexacoordinated during the whole process[40]. In the first case, it is proposed that the vacant position is occupied by the substrate while in the second case the substrate is thought to bind the second coordination sphere; (ii) whether the reduction of the metal center occurs through the abstraction of one proton and two electrons from the formate molecule[45,49,68,69] or by a hydride transfer through the metal (directly)[67] or through the sulphur ligand (=S/-SH)[40,71].

Concerning the different characteristics of the enzyme, it is expected that the W/Mo and the SeCys/Cys dichotomies do not translate to any meaningful mechanistic changes. This is further corroborated by the existence of enzymes capable of incorporating both metals[44] and that SeCys to Cys mutants in FDH-H show activity, even though two orders of magnitude lower, which further evidences the important role of this residue in catalysis[63,72]. Nonetheless, it is important to take into consideration that all mechanistic proposals given at the moment are focused on formate oxidation, however, the main goal and hope for the future is to use the enzyme as a CO_2 reductase. Even though the last mechanistic proposal did encompass both an explanation to formate oxidation and CO_2 reduction, little to no information has been published regarding the mechanism of the latter, which requires urgent tackling since there is a possibility that the mechanism is not exactly the same and may have different energetical stepstones.

2.4. FdhAB from *D. vulgaris* Hildenborough

D. vulgaris Hildenborough is the model organism when studying sulphate reducing bacteria (SRB), which are anaerobic prokaryotes that are characterized by using sulphate as the terminal electron acceptor to generate energy in a process named sulphate respiration[73,74]. The genome of this microorganism encodes three different FDHs: two soluble periplasmic enzymes (FdhAB and FdhABC₃) and one membrane-bound enzyme (FdhM) also facing the periplasm[75].

Several enzymes capable of CO_2 reduction have been characterized, but the tungsten- and SeCys-containing FdhAB from *D. vulgaris* Hildenborough is among the most promising, with high rates both for formate oxidation (1310 s^{-1}) and for CO_2 reduction (315 s^{-1})[39]. Besides that, it does have an advantage compared to other anaerobic FDHs, since it can be purified aerobically without compromising its activity. Nonetheless, it does require pre-activation with DTT in order to reach its highest activity[39].

These reaction rates were only surpassed by *Acetobacterum woodii* and *Thermoanaerobacter kivui* for both formate oxidation (1690 and 1335 s^{-1}, respectively) and CO_2 reduction (372 and 2657 s^{-1}, respectively)[30,41], and by two enzymes present in *Syntrophobacter fumaroxidans* for formate oxidation (3380 and 5600 s^{-1})[37,38,43]. Nonetheless, despite the high reaction rates obtained, the enthusiasm

surrounding these microorganisms was hampered by the fact that these are hard to manipulate and that the enzymes are extremely O_2-sensitive, not being realistic to envision a biotechnological application[30,38,41,43].

The interest in FdhAB is also attributed to its operational stability, having formate oxidation activities close to 80% at pH 7 to pH 10 (with an optimal value of 7.6) and CO_2 reduction activities above 60% at pH 4.7 to 7.6 (with an optimal value of 7.1), with its overall stability prompting exciting studies in several applications such as photocatalysis, which will be further described.

All of these characteristics make FdhAB an excellent model system for understanding the mechanisms behind CO_2 reduction and highlight the tremendous potential of the enzyme for converting CO_2 to formate, making it one of the most promising tools to address the two major problems that society will face in the next few decades.

2.5.   Photocatalysis

One of the most encouraging approaches for shifting the way we utilize energy is investing in renewable green fuels, with photocatalysis gaining a center position to achieve this goal. Photocatalysis consists of using light as a source of energy to perform several types of catalytic reactions[76]. When photocatalytic systems are coupled with biological catalysts this approach is generally named photobiocatalysis[76,77].

To present knowledge, only four types of enzymes require light to catalyse a chemical reaction. These enzymes are commonly named photoenzymes and are the photosystem, photolyases, protochlorophyllide-reductases and photodecarboxylases. Nonetheless, several other enzymes do not require light to catalyse but show a different reaction behaviour when irradiated[78]. In recent years, photocatalysis has been employed to activate redox-enzymes, since they often rely on an external electron source to catalyse reactions, opening a vast realm of applications for these processes[76,78].

These types of photobiocatalytic systems often rely on three components: a sacrificial electron donor, a photocatalyst, and the enzyme. In some cases, additional components such as cofactors and mediators might be required for the enzymatic system to function[77].

In order to understand what makes a good photocatalyst, we need to first understand the basics of band theory. According to this theory, materials can be divided into three different types: insulators, semiconductors, and conductors[79]. To perform this division, we need to consider the existence of electronic bands, which basically consists of a collection of energy levels. The valence band corresponds to the collection of energies associated with the outer shell electrons, while the conduction band corresponds to the collection of energies associated with free electrons (meaning they can conduct current). In conductors, the valency band and the conduction band overlap, therefore electrons that exist in the valency band automatically are in the conduction band (meaning the outer shell electrons are free to move). In the case of semi-conductors and insulators, there is something called a "forbidden gap", which is a range of energies that the electron cannot possess (being this the reason why a molecule can only absorb photons of certain wavelengths). In the case of insulators, this energy gap between the two electronic bands is too large, therefore electrons cannot move to the conduction band, however in

the case of semiconductors, this gap is not as big and if energy is supplied they can move from the valency band to the conduction band[79,80].
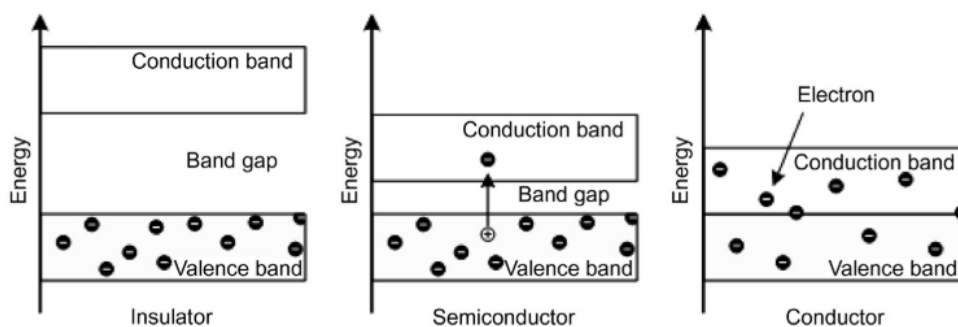
Figure 9: Classification of the different type of materials according to band theory. In the left image, there is a general description of an insulator, where the bandgap is large, making the electrons not accessible to the conduction band. In the center image we have a semiconductor, where an external supply of energy can make an electron reach the conduction band, where it is free to conduct. In the right image, we have a conductor, where there is no gap between the valence and the conduction bands, therefore the electrons are free to conduct. Taken from Ameta *et al.* (2018)[80].

Photocatalysts are typically semiconductors and when they are exposed to photons with the energy needed to overcome the gap, the energy is absorbed and the electron is excited from the valency band to the conduction band[80]. When this happens, there is the formation of a hole (h^+) in the valence band, which can be defined as an empty state even though holes behave as being positively charged[79]. When this occurs, there are four ways that the energy can dissipate: internal conversion (mode of decay where the energy is released in the form of heat), fluorescence (mode of decay where the energy is released in the form of a photon), resonance energy transfer (the excitation is transferred to another molecule) or photooxidation/photoreduction (if the redox level of the substrate is lower than the conduction band of the semiconductor or higher than the valence band of the semiconductor, respectively; it is possible that both occur if the redox level of the substrate is lower than the conduction band and higher than the valence band of the semiconductor and neither if the redox level of the substrate is higher than the conduction band and lower than the valence band of the semiconductor)[80,81]. The first two processes are generally named recombination processes, meaning that there is a recombination between two opposite charged particles, an electron and a hole, with the energy being somehow dissipated[80–82]. These recombination processes lead to lower efficiencies of the photocatalytic reactions since the energy absorbed is not being used for the intended process[82]. Even though the photocatalyst can directly perform photooxidation or photoreduction, typically it is used another molecule (referred to as co-catalyst), where the desired reaction takes place and, in this case, the primary role of the semiconductor is to absorb photons and to generate electron-hole pairs, transforming the photon energy in electrons accessible to the catalyst[82].

When choosing a semiconductor, we must take into consideration two general parameters: the wavelength it absorbs and the energy of the conduction/valence bands. The former, ideally, we would want a photocatalyst capable of absorbing in the visible region, while the latter is dependent on the reaction we want to perform to see if it is thermodynamically feasible. It is worth noting that, from a kinetic standpoint, it is important that an overpotential exists when performing the reaction, therefore the

edge of the bandgap needs to be higher or lower than the potential of the reaction, depending whether we are performing photooxidation or photoreduction[82].

A commonly used semiconductor in studies focused on photocatalytic CO_2 reduction is titanium oxide, mostly due to its high abundance in nature, being non-toxic, stable, and durable[83]. It has mainly three different polymorphors, brookite, rutile, and anatase, but only the latter two show photocatalytic effects[84]. The differences in the crystal structures result in different bandgaps for anatase and rutile (3.2 eV and 3.0 eV, respectively which correspond to the minimum energy of photons needed of 389 nm and 413 nm, respectively, to initiate an interband transition)[82,84]. A lot of effort has been made to determine which of the polymorphs is better for photocatalytic applications and it has been shown that mixed-phase TiO_2 has enhanced efficiencies due to complementary effects between both phases[82]. The major drawback of TiO_2 is that, even though it has high efficiency for photocatalysis under UV irradiation, it does not respond to visible light due to its wide bang gap (Figure 10)[83]. This is a major limitation because only 4% of the sunlight is in the UV region, whilst 48% is in the visible region (and the remaining in the IR region), resulting in a lot of wasted light, limiting efficiency[83].
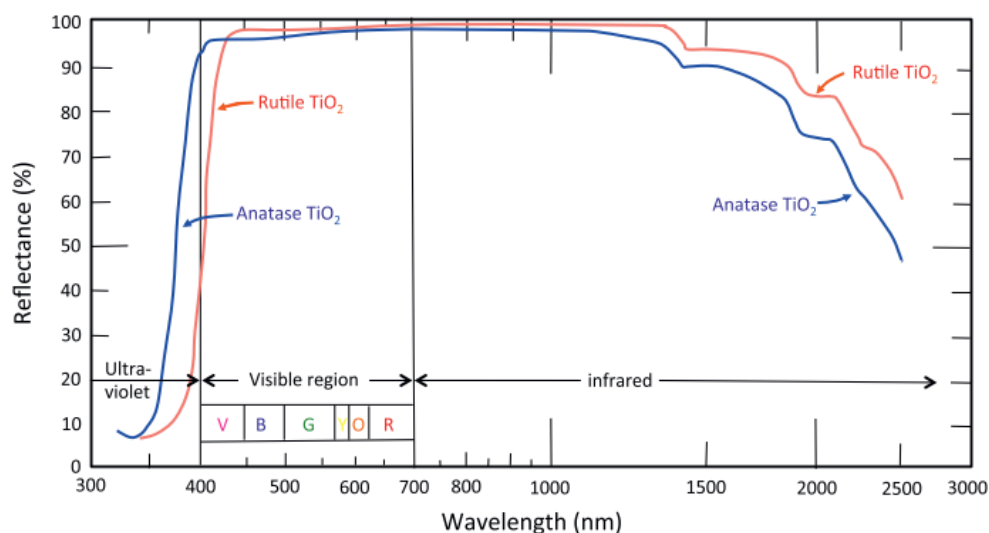
Figure 10: Reflectance (%) of both rutile and anatase in function of the wavelength of the incident radiation. Lower reflectance corresponds to higher absorptions of energy at that wavelength. Taken from Lan, Y. *et al.* (2013)[84].

Several modifications can be performed in a semiconductor to adapt its characteristics and properties for the wavelength of light required and for the reaction being performed. These approaches can be divided into essentially two categories: bandgap engineering or sensitization. In the former, metal doping and non-metal doping are two commonly used methodologies, whereas titanium atoms are replaced by other metal ions, giving rise to a new conduction band, or oxygen atoms are substituted by other atoms such as nitrogen, carbon, sulphur, giving rise to a new valence band[80,83]. These modifications decrease bandgap size and increase charge-separation, which allow visible region light absorption and decrease recombination processes[80,83]. The main issue with this approach is that particularly in reactions where the solvent is water, doped TiO_2 with other metals is subjected to photocorrosion and leads to the deactivation of the doped complex, and doping with both metals and non-metals leads to several reproducibility problems due to variation in loading and location[83].

On the other hand, sensitization is a promising modification method since it allows the use of organic dyes (such as porphyrins, xanthenes, Ru complexes, flavins) and colloidal dispersions of

nanomaterials (such as quantum dots, carbon nitrides, and carbon dots) that are able to absorb visible light and inject the electrons in the conduction band of the semiconductor, initiating the photocatalytic cascade[76,80]. Furthermore, a crucial parameter for the efficiency of the system will be the type (e.g. physisorption or chemisorption) and the strength of the interaction between the dye and the surface of the semiconductor[85]. Xanthenes, such as Eosin Y, are being widely used, mostly because of their low cost, good visible absorption properties, commercial availability, and being metal-free[76]. Furthermore, Eosin Y has a LUMO (lowest unoccupied molecular orbital) with a potential of -0.92 V, which is higher than the conduction band of TiO_2 (-0.5 V), therefore allowing from a thermodynamical perspective the electron injection[86–88]. After electron injection, the dye must be regenerated. Utopically, CO_2 reduction would be coupled with water oxidation (due to its abundancy) or with advanced oxidation of wastewaters from industrial effluents (which would allow simultaneous wastewater treatment)[89]. However, since photocatalytic CO_2 reduction by itself is already a difficult task, the focus at this point is only the reduction half-reaction, so there is the need to use sacrificial electron donors (SEDs). A good SED has to fulfil three basic conditions: needs to be thermodynamically adequate for the molecule we want to regenerate (meaning that the redox potential of the SED needs to be lower than the redox potential of the HOMO – highest molecular occupied orbital – from the dye), needs to be irreversible transformed when oxidized into an inert molecule and, if required, have faster degradation kinetics than the recombination kinetics of the photoexcited dye. Therefore, the choice of the SED will have a great impact on the overall efficiency of the process[89]. Nonetheless, apart from the thermodynamically considerations, it is not easy to rationalize and predict how a SED will behave in a system, since the environment will influence its behaviour[76,89]. We can look at a SED as a mean to an end, eventually the use of SEDs will be abandoned since they are not viable for large scale applications, but rather work as proofs-of-concept for laboratory scale experiments. For these reasons, the choice of the SED for testing a photocatalytic system is mostly focused on it not limiting the system and it not being detrimental for the performance of the system[89].

Until now, photocatalytic CO_2 reduction makes use of mediators, which are typically not efficient, are toxic (e.g. methyl viologen), and are expensive (e.g. NAD^+)[90,91]. Only one photocatalytic system with a highly active FDH has been described that does not use these mediators[92]. In this study, FdhAB from *D. vulgaris* was evaluated with the use of two systems containing TiO_2 nanoparticles and a metal-containing dye (RuP – ruthenium tris-2,2'-bipyridine complex) or a metal-free dye (DPP – diketopyrrolopyrrole). Under visible-light-driven excitation, a turnover of 11 s^{-1} with RuP and 5 s^{-1} with DPP were obtained for CO_2 reduction. Furthermore, it was shown that FdhAB and TiO_2 nanoparticles interact strongly, even when increasing ionic strength showing more than electrostatic interactions. The authors suggest that this strong interfacial interaction plays a critical role in the high stability and activity of the systems, with turnovers significantly higher than other studies with the use of soluble redox mediators, highlighting the potential of TiO_2 when employed with this enzyme and that the route for improving photocatalytic applications of the enzyme needs to take into consideration the interactions between the enzyme-material interface[92].

2.6. Computational Stabilization of Enzymes

Protein folding is a quite marvellous and complex thematic that essentially consists in the organization of the residues in a three-dimensional structure with the lowest free energy. For this to

occur, the interactions in the folded structure need to be strong in order to surpass the entropic cost of folding[93]. Furthermore, for a protein to remain folded, the free-energy of the native stated must be lower than the free-energy of unfolded or misfolded states, being this gap in energy the broad definition of protein stability[94]. In 1961, it was shown that the small protein ribonuclease (124 amino acids) is able to spontaneously refold to its native structure, which led to the theory that folding is dependent on amino acid sequence (thermodynamic hypothesis of folding)[95]. If we assume three degrees of freedom per residue, the number of unfolded states of a protein with 100 residues would be 5.15 x 10[47], so if proteins folded randomly, the time for the process would be longer than the age of the universe[96,97]. Therefore, as proposed by Levinthal, folding intermediates and pathways must exist[96]. Even though some proteins are able to spontaneously fold, the so-called two-state folders (only two states exist, the folded and unfolded, with no stable intermediates), larger proteins typically possess stable intermediates (local structures) that can lead to inactive states of the enzyme (misfolded or aggregate)[94].

In natural proteins, the difference in energy between these inactivate states and the native state is quite low, being often said that proteins are only marginally stable, which often means that they are sensible to mutations or changes in the environment[94]. There are essentially four explanations to the existence of marginally stability in proteins: (i) in some activities, there is a need for a compromise between structural and functional features (for instance, mutations in catalytic residues tend to stabilize proteins, even though decreasing activity – the so-called stability-activity trade-off[98,99]); (ii) the number of non-functional states far exceeds the native state by several orders of magnitude, which means that the entropic penalty for folding is high and each stabilizing contact in the folded state is small, therefore many of these contacts are needed; (iii) it can be a mechanism of regulation since a highly stable protein would be harder to degrade; (iv) the pressure for evolution decreases when a certain level of stability is achieved[94]. For these reasons, there is a high potential for improvement in the stability and functional features of proteins.

In the last two decades, with the appearance of next-generation sequencing technologies and with the development of the field of protein engineering, we have the information and the means for protein tailoring[100]. Nonetheless, the (re)design of a protein is not a trivial task, with all methodologies having their pros and cons. The simplest technique is direct evolution, which is an experimental method that consists of supplying pressure for evolution to generate a library of mutants, that can then be screened for the intended characteristic[100,101]. Even though simple, this technique is quite laborious and demands high throughput (mostly in the screening phase)[100]. The immensity of a protein sequence and structure completely prohibits the evaluation of all possible combinations of mutations. With this in mind, rational design has emerged as a strategy capable of significantly reducing experimental effort by using knowledge to guide the mutations[100]. Several computational approaches have been developed in order to predict activity-enhancing or stabilizing mutations, with the methods being divided essentially into four types: energy-based, structure-based, phylogeny-based, and hybrid-design techniques[94].

Energy-based methods use force fields (collection of bonded and non-bonded interaction terms) to estimate the potential energy of a system[100]. There are three main methods for estimating the energy: statistical effective energy functions (SEEFs – energies are derived from the frequency of residue or atom contacts in the protein database), empirical effective energy functions (EEEFs – use empirical

data from experiments), and physical effective energy functions (PEEFs – such as the ones based on molecular dynamics and Monte Carlo simulations)[102]. For an energy function to fall in the category of SEEF, EEEF, or PEEF is strictly related to the structural detail it encompasses[102]. EEEFs are widely used, such as FOLD-X[103] and PremPS[104], since they can provide a good balance between accuracy and computational cost, in comparison to SEEFs and PEEFs. Nonetheless, one of the big disadvantages of EEEFs is that they have been parametrized for single-point-mutation datasets, therefore when applying to multiple-point mutants the errors increase, which is further exacerbated if there are epistatic effects[100].

Several software tools have been developed that rely on energy functions, with one of the most popular being Rosetta, which is highly versatile comprising a growing number of modules that started with structure prediction and now ranges from macromolecular modelling to design[100,105]. The Rosetta energy function computes the energy of the system (ΔE_{total}) as a linear combination of several weighted energy terms that are calculated in function of the chemical identity and of the geometric degrees of freedom[106]. The first version of the Rosetta Energy Function used statistical potentials (SEEF)[107], however, it has been significantly improved with the latest version being Rosetta Energy Function 2015 (ref2015) which combines physic-based and statistic-based potentials, comprising in total 19 energy terms that describe atom-pair interactions (Van der Waals interactions, electrostatic interactions, hydrogen bonding and disulphide bonding), protein backbone and side-chain torsions, special-case torsions (such as the ones caused by prolines), terms for modelling nonideal bond lengths and angles and terms for protein design. A detailed explanation of each term can be found in [106]. In the past, the energy derived from Rosetta used a generic unit called Rosetta Energy Unit (REU), nonetheless nowadays with ref2015 energy function, the values obtained are a stronger approximation of energies in kilocalories per mole[106]. For prediction of protein stability, Rosetta estimates the free energy of a given mutation, which is commonly referred to as $\Delta\Delta G$ of mutation, and consists in the energy difference between the wild-type structure and the single-point mutant with the use of the energy function[106,108]. If this $\Delta\Delta G$ satisfies a given defined cut-off, the mutation is accepted as being stabilizing[106,108]. For multi-point mutants, Rosetta combinatorial sequence design can be used to find an optimal combination of the mutations that were considered as potentially stabilizing, through sophisticated search algorithms based on statistical methods[109].

Structure-based methods focus on a given aspect of protein stability, in order to eliminate the uncertainty associated with conformational sampling and the need to use energy functions[94]. They focus on given parts of the structure and aim to improve a given structural characteristic of that part (e.g., increasing solubility[110], rigidifying backbone[111]). Even though they have been successful, they are not general methods for increasing stability, only focusing on a defined subset of situations[94].

Phylogeny-based can be centred on consensus design or ancestral sequence reconstruction[100]. Consensus design has at its core that, in a family of homologous proteins, in any position, the most prevalent amino acid is likely to be the most stabilizing, therefore supporting the protein folding and function. On the contrary, amino acids that are rare or absent in a given family might be deleterious for folding and function[94,100]. Ancestral sequence reconstruction is mostly used as a method to detect primordial mutations, which have been proved to be a benefit for thermostability engineering[100].

Hybrid-design approaches try to combine the advantages of energy, structure and phylogeny-based methods[94]. Two main methods of hybrid-design are FireProt[112,113] and Protein Repair One Stop Shop (PROSS)[109], which were pioneers in highly stable multiple-point mutant design. Both methods require a solved structure and evolutionary information that can be obtained by multiple sequence alignments (MSA). PROSS has the advantage of designing the whole protein, with the exception of the active site while FireProt only focuses on regions of the protein that are likely to be the ones causing stability issues.

PROSS protocol starts with the generation of a position-specific substitution matrix (PSSM), which represents the log-likelihood of observing any of the 20 canonical amino acids at a given position. Mutations with a score (log-likelihood) ≥ 0 are considered favourable, using the reasoning that deleterious mutations are rare or absent due to natural selection and that the most frequent amino acids at a given position may indicate stability (consensus effect – evolution-based approach). In a second stage, Rosetta calculations are used to obtain the energy difference between the wild-type and a given single-point mutant (ΔGG_{calc}), for all single-point mutations considered by the first step (energy-based approach). Depending on the energy threshold considered, we obtain a different set of single-point mutants. For instance, if we use a strict *cutoff* ($\Delta GG_{calc} \leq -2.0$) we will obtain fewer single-point mutants than considering a less strict cutoff ($\Delta GG_{calc} \leq +0.5$), nonetheless, in the last model, it will likely be obtained a higher number of false positives. The authors advise using a *cutoff* of $\Delta GG_{calc} \leq --0.45$, which in combination with the scoring function given by the PSSM, is able to eliminate 99.6% of all destabilizing mutations while retaining one to two-thirds of experimentally validated stabilizing mutations. The mutations that are validated, surpassing both stages of selection, are then used by Rosetta combinatorial sequence design, which finds the optimal combination of mutations (from all the single-point mutants), generating a final model that needs to be experimentally validated. PROSS has been used to stabilize human acetylcholinesterase, resulting in an increase of 2000-fold in bacterial expression without negatively impacting activity and increasing heat tolerance by 20 °C[109]. Furthermore, it has been used to stabilize PfRH5, which is a malaria invasion protein that is a promising vaccine candidate, without altering immunogenicity or ligand binding. This allowed the expression using *E. coli* and an increase in thermal stability[114]. PROSS is also available as a user-friendly webserver in https://pross.weizmann.ac.il/step/pross-terms/ [109].

Hybrid methods are more robust and allow the generation of multiple-point mutants with a high degree of confidence. Furthermore, even though the methods described are based on increasing stability, more recently hybrid methods for predicting catalytic activity and/or changing substrate specificity, such as FuncLib[115], are being developed. The future has never been brighter for the world of protein engineering.

3. Motivation and Objectives

As previously seen, FDHs are highly heterogenous, mostly in quaternary structure and redox center composition[45,49]. Nonetheless, their active site is well-conserved, with a Mo or W center coordinated to two PGDs forming a metal-dithiolene complex. The remaining two positions of the coordination sphere are occupied by a sulphide group (=S/-SH) and a cysteine/selenocysteine residue from the polypeptide chain[39,45,49,53]. Even though the active site structure is widely accepted, there is no consensus around the mechanism of catalysis. The main issues under debate are: (i) whether the SeCys/Cys residue present in the first coordination sphere of the metal dissociates during catalysis, yielding a pentacoordinated metal[49,67-69] or if the metal remains hexacoordinated during the whole process[40]. In the first case, it is proposed that the vacant position is taken by the substrate, while in the second case the substrate is thought to bind the second coordination sphere; (ii) whether the reduction of the metal center occurs through the abstraction of one proton and two electrons from the formate molecule[45,49,68,69] or by a hydride transfer through the metal (directly)[67] or through the sulphur ligand[40,71].

Either way, several enzymes capable of efficient CO_2 reduction have been characterized, but FdhAB from *D. vulgaris* is among the most promising. It possesses two subunits: a catalytic (α) subunit with the active center and one [4Fe-4S] cluster and a small electron-transfer (β) subunit with three [4Fe-4S] clusters. Interestingly, the catalytic subunit of FdhAB has some structural resemblance to the FDH-H from *E. coli*, with the biggest difference between them being the presence of an additional long C-terminal helix in the latter[39]. FDH-H is a Mo-containing SeCys-containing FDH, with a simple quaternary structure composed of just the catalytic subunit and harbouring one iron-sulphur cluster, which is close to the protein surface. It works as a component of the formate hydrogen lyase complex, but it can be isolated on its own, which allowed for its crystallization[55] and kinetic characterization[116]. The similarity between the α-subunit of FdhAB and FDH-H and the success in obtaining an active FDH-H separate from the FHL complex prompted the hypothesis that the α-subunit of FdhAB could be produced and characterized in an isolated form. It is known that DvFdhAB has a high operational stability[39] and, if this stability is maintained with the removal of the β-subunit it could be possible to obtain a highly active simplified version of FdhAB, which could be of great relevance for mechanistic studies and future biotechnological applications. Moreover, it would allow us to confirm whether the β-subunit serves only for electron transfer or if has additional functions that could contribute to the overall activity of the enzyme. A similar study has already been carried with a [Ni-Fe] hydrogenase, where the electron transfer subunit was removed and the authors were still able to obtain the catalytic subunit with full cofactor loading, nonetheless with significant decreases in activity for hydrogen-deuterium exchange (H_2 splitting capacity)[117].

In terms of future applications, one of the most promising approaches to solve the energy crisis humanity faces is to invest in renewable green fuels, with light being an unlimited source of energy[76]. With this in mind, developing better photocatalytic systems and testing enzyme variants both with traditional methods (typical activity assays) and more applied ones (in this case, in a photocatalytic system) is crucial. A substantial amount of studies focused on photocatalytic CO_2 reduction are centred on the use of TiO_2 nanoparticles[83]. Nonetheless, systems up to now rely heavily on the use of undesired and inefficient mediators[90,91]. Knowing that FdhAB strongly interacts with TiO_2 nanoparticles[92] and that

Eosin Y, a cheap commercially available dye, and TiO_2 have already been coupled successfully for hydrogen generation[86], a setup that combines these components would allow us to have an efficient, commercially available system to test future enzyme variants.

With all of this in mind, the first part of the work comprises essentially three distinct goals: **(i) production of a simplified version of the protein composed only of the catalytic subunit**; **(ii) expression and purification of both the simplified FdhA and the FdhAB enzyme for characterization and comparison**; **(iii) testing different photocatalytic setups with a combination of SEDs, TiO_2 nanoparticles and Eosin Y for CO_2 reduction with FdhAB**.

Due to the existence of marginal stability in natural proteins, the great majority of them have room for improvement, in terms of both stability and functionality[94]. Moreover, anticipating that the removal of an entire subunit can lead to several stability issues, as seen in the example of the [Ni-Fe]-hydrogenase[117], a computational optimization of both the FdhAB and the simplified FdhA was performed. Therefore, the second part of the work consists of achieving a fourth goal: **(iv) *in silico* stabilization of FdhAB and FdhA mutant using PROSS**[109].

Understanding how to improve FdhAB stability and its variants may bring important information that could potentiate the biotechnological use of this enzyme as a biocatalyst for CO_2 reduction.

4. Materials and Methods

4.1. Generation of a pRec-FdhA-Strep Plasmid

To generate the strain expressing only FdhA, the plasmid pRec-FdhAB-Strep[39] (Figure 26 in the Annex Section) was mutated to eliminate the *fdhB* sequence. For this, an *Eco*RI restriction site was introduced by site-direct mutagenesis. Primers containing the *Eco*RI restriction site were designed to flank the KanR cassette near the ribosomal binding site of KanR. The region downstream *fdhB* gene had already a restriction site for *Eco*RI. The primers were manufactured by Invitrogen and their sequence is present in Table 12 in the Annex section. Mutagenesis was performed following the NZYMutagenesis Kit protocol, using *E. coli* DH5α competent cells. After transformation, the cells were plated onto LB Agar (25 g/L LB Broth with 15 g/L Agar) plates with spectinomycin (100 µg/mL) and were grown at 37 ºC overnight. Colonies were collected and grown on LB medium with spectinomycin (100 µg/mL) at 37 ºC overnight. The plasmid was then extracted using the plasmid DNA Miniprep NZYTech Kit and stocks of the culture were made with 10% glycerol. Digestion of the plasmid extracted was performed with *Eco*RI for 37 ºC and 1 hour, followed by an inactivation step of 20 minutes at 65 ºC. The product of the reaction was analysed using an agarose gel. The linearized plasmid, with the expected weight of 7875 bp was then extracted from the agarose gel using the GeneJET Gel Extraction Kit from Thermo Fisher, and the extracted plasmid was ligated using T4 DNA Ligase (BioLabs Inc., New England) for 2.5 hours at room temperature, followed by an inactivation stage of 10 minutes at 65 ºC. The generated pRec-FdhA-Strep was then transformed into *E. coli* DH5α competent cells, with the cells plated onto spectinomycin (100 µg/mL) LB Agar plates and grown at 37 ºC overnight. A colony PCR was performed to confirm the presence of the pRec-FdhA-strep plasmid using primers that hybridize in the regions flanking the desired deletion, yielding in positive colonies a band with 3893 bp. The selected colonies were then grown at 37 ºC overnight in LB medium with spectinomycin (100 µg/mL). The plasmid DNA was extracted, using the same protocol as before, and sequenced by GATC Biotech, Germany.

4.2. Generation of a DvFdhA Mutant

A *D. vulgaris ΔfdhAB* deletion strain was grown in MOYLS medium pH 7.2 (8 mM MgCl$_2$, 20 mM NH$_4$Cl, 0.6 mM CaCl$_2$, 2 mM K$_2$HPO$_4$-NaH$_2$PO$_4$, 0.6% (v/v) Trace Elements (0.5 g/L MnCl$_2$•4H$_2$O, 0.3 g/L CoCl$_2$•4H$_2$O, 0.2 g/L ZnCl$_2$, 0.05 g/L Na$_2$MoO$_4$•4H$_2$O, 0.02 g/L H$_3$BO$_3$, 0.1 g/L NiSO$_4$•6H$_2$O, 0.09 g/L NiCl$_2$•6H$_2$O, 0.002 g/L CuCl$_2$•2H$_2$O, 0.006 g/L Na$_2$SeO$_3$•5H$_2$O, 0.008 g/L Na$_2$WO$_4$•2H$_2$O), 0.06 mM FeCl$_2$, 0.12 mM EDTA, 30 mM Tris-HCl 2 M, 0.1% (v/v) Thauers Vitamins 10 X, 1.2 mM Thioglycolate, 1 g/L Yeast Extract, 30 mM Lactate 1 M, 30 mM Sulphate). When the cells reached an OD of 0.2, they were electroporated, as described in [118], with the constructed plasmid at 1250 V, 250 Ω, and 25 µF (Gene Pulse Xcell, Bio-Rad) and left anaerobically in MOYLS medium overnight at 37 ºC. Cells were then plated using MOYLS medium with Agar (15 g/L) containing spectinomycin (100 µg/mL) inside of an anaerobic chamber. The cells were then left to grow for 6 days in anaerobic conditions at 37 ºC. The selected colonies were then grown in liquid MOYLS. When cultures reached maximum growth (OD of 0.6-0.8) stocks in 10% glycerol were made. Colonies were inoculated in Postgate Medium C (3.7 mM KH$_2$PO$_4$, 18.7 mM NH$_4$Cl, 17.6 mM Na$_2$SO$_4$, 0.4 mM CaCl$_2$•2H$_2$O, 0.24 mM MgSO$_4$•7H$_2$O, 0.4 g/L Yeast Extract, 26 µM FeSO$_4$•7H$_2$O, 1 mM sodium citrate tribasic dihydrate, 0.57 mM L-ascorbic acid, 0.88 mM sodium thioglycolate, 40 mM sodium formate, 10 mM sodium DL-lactate and 100 mM Tris base,

supplemented with 10 μM of NiCl$_2$•6H$_2$O/Na2SeO3•5H$_2$O and 10 μM of Na$_2$O$_4$W•2H$_2$O). After 24 hours, cells of the different colonies were recovered by centrifugation and stored at -20 ºC. The plasmid was extracted following the previously described procedure and sequenced by GATC Biotech, Germany.

4.3.   Protein Purification

4.3.1.   Protein Purification by Affinity Chromatography

D. vulgaris Hildenborough cells expressing the proteins of interest were subjected to a large-scale growth (30 L in the case of the cells expressing FdhA and 300 L in the case of the cells expressing FdhAB) in Postgate Medium C. The cells were then collected and stored at – 80 ºC.

When protein purification was performed, the cells expressing the protein of interest were retrieved from – 80 ºC and resuspended in binding buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 10 mM NaNO$_3$, 10% (v/v) glycerol). During cell resuspension, DNAse was added. The resuspended cells were then disrupted in a French Press. The cellular extract was subjected to centrifugation at 8000 g for 20 minutes at 4 ºC, with the supernatant being recovered and subjected to ultracentrifugation at 109 000 g for 2 hours at 4 ºC. The supernatant was recovered, and protease inhibitor was then added, obtaining the soluble fraction. For FdhAB purification, the reduction potential was further measured for 45 minutes at 4 ºC and the protein extract was then left to oxidize for 2 overnights at 4 ºC under mild agitation. The reduction potential was then re-measured, confirming its oxidation (only in samples from FdhAB). Before purification, the soluble fraction was centrifuged at 48 000 g for 30 minutes at 4 ºC. The protein extract (soluble fraction) was then loaded on a Strep-tactin gravity flow affinity column (IBA Lifesciences, Germany; column volume of 1.5 mL) equilibrated with binding buffer. Afterwards, 5 washing steps were performed with binding buffer. The elution stage was then performed by 8 elution steps using elution buffer (binding buffer supplemented with 2.5 mM D-desthiobiotin). Two elution samples were recovered: the first two elution steps were pooled together (Elution A) and the same for the following 6 elution steps (Elution B). The fractions that were collected during the loading and washing steps (flow-through) were then subjected to another round of affinity chromatography, as described. The buffer of the eluted fractions was then changed to sample buffer (20 mM Tris-HCl pH 7.6, 10 mM NaNO$_3$, 10% (v/v) glycerol) using a 30-kDa cut-off ultracentrifugation unit (Amicon Ultra-15 30K NMWL, Millipore) with the samples being stored at 4 ºC for immediate use or at -20 ºC for longer periods. In samples purified anaerobically, all steps were performed anaerobically and the affinity chromatography setup was performed inside a COY Anaerobic Chamber under an atmosphere of 2% H$_2$/98% N$_2$ at room temperature. In samples purified aerobically, all the procedures were performed under air. The purity of each sample was evaluated by a 12% SDS-polyacrylamide gel. The UV-visible absorption spectrum for all samples obtained from purification was measured in the Nanodrop ND2000C from 250 to 500 nm.

4.3.2.   FdhA Purification by Ion Exchange Chromatography

Samples were centrifuged at 17000 g for 20 minutes at 4 ºC, with the supernatant being then loaded on a 6 mL Q-Resource^{TM} column (Pharmacia Biotech), equilibrated with sample buffer (20 mM Tris-HCl pH 7.6 with 10% (v/v) glycerol and 10 mM NaNO$_3$). The column elution was performed at a flow rate of 1 mL/min with stepwise increments of NaCl (25-50 mM NaCl), until 200 mM NaCl, then a

linear gradient was applied until a concentration of 1 M NaCl or a two-step increment with 500 mM NaCl and 1 M NaCl were performed. To evaluate the purification, a 12% SDS-polyacrylamide gel was run.

4.4.   Protein Concentration

The samples obtained after all steps of purification were quantified by Nanodrop ND200C. In the case of FdhAB purification, the Lambert-Beer Law was used with the $\varepsilon_{410\ nm}$ = 43.25 mM^{-1}.cm^{-1}. In the case of FdhA, the following rationale was used:

$$When\ Abs_{260\ nm} > Abs_{280\ nm} \rightarrow [Protein](mg/mL) = 1.55 \times Abs_{280\ nm} - 0.76 \times Abs_{260\ nm}\ (eq.\ 4)$$

$$When\ Abs_{260\ nm} < Abs_{280\ nm} \rightarrow [Protein](mg/mL) = Abs_{280\ nm} \hspace{3cm} (eq.\ 5)$$

For the kinetic analysis of FdhA, the protein was quantified by BCA (BCA Protein Assay Kit from Novagen) or by the estimated $\varepsilon_{410\ nm}$ determined for FdhA. For circular dichroism spectroscopy, both FdhA and FdhAB were quantified using BCA. For the thermal-shift assay, quantification of FdhA and FdhAB was performed using their respective $\varepsilon_{410\ nm}$.

4.5.   Western-Blot

The soluble extract was obtained by cell disruption using BugBuster Protein Extraction Reagent (Novagen, Inc.). The sample was then run on a 12% polyacrylamide gel and transferred using a Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad) to a polyvinylidene difluoride membrane (PVDF membrane) during 1 h at 350 mA (150 V, 100 W) at 4 °C. The membrane was then left to dry overnight. On the next day, the membrane was blocked with PBS blocking buffer (PBS Buffer with 3% BSA, 0.5% (v/v) Tween 20) for 1 h at room temperature and mild agitation. Afterwards, three washing steps were performed (5 minutes each) with PBS Buffer with 0.1% Tween 20. The membrane was then incubated with Biotin Blocking Buffer (IBA Lifesciences, Germany) for 10 minutes and then treated with Strep-Tactin antibody labelled with alkaline phosphatase (IBA Lifesciences, Germany) diluted 1:4000 for 1 hour at room temperature and mild agitation. Two washing steps were then performed (1 minute each) with PBS Buffer with 0.1% Tween 20. Immunodetection was performed, protected from light, in 100 mM Tris HCl pH 8.0, 100 mM NaCl and 5 mM MgCl$_2$ with nitro-blue tetrazolium and 5-bromo-4-chloro-3'-indolyphosphate (NBT-BCIP) (Sigma-Aldrich) during 2h30 at room temperature and mild agitation. To stop the reaction, distilled water was used to wash the membrane.

4.6.   Activity-stained Native Gel

A soluble fraction from FdhAB and FdhA (50 µg) were run in a native polyacrylamide gel (7.5%) containing 0.1% Triton X-100 to evaluate formate oxidation activity. The gel was incubated in a solution of 50 mM Tris-HCl buffer pH 7.6, 20 mM sodium formate, 0.908 mM DTT, and 0.625 mM benzyl viologen, under a N$_2$ atmosphere. When it was possible to visualize bands in the gel, corresponding to the reduction of benzyl viologen, the bands were fixed by adding 10 mM of 2,3,5-triphenyltetrazolium chloride solution.

4.7. Activity Assays and Kinetic Analysis

Formate dehydrogenase activity assays were performed anaerobically in a COY Anaerobic Chamber under an atmosphere of 2% H_2/98% N_2 at room temperature under constant stirring, using a UV-1800 Shimazu spectrophotometer. The assays for formate oxidation activity and CO_2 reduction activity were performed with or without DTT activation. With activation, the enzyme was pre-incubated with 50 mM DTT for 5 minutes. For formate oxidation, the mixture was then added to a cuvette with 50 mM KPi pH 7.6, 1 mM DTT, and 2 mM benzyl viologen. After 30 seconds the reaction was initiated by the addition of 20 mM of formate. For CO_2 reduction, the reaction was started by adding the pre-incubated enzyme to a cuvette with 50 mM KPi pH 6.9, 1 mM DTT, Zn-reduced methyl viologen (adjusted to an absorbance of 1), and 50 mM sodium bicarbonate. In the absence of activation, no DTT incubation step was performed and no DTT was added to the reaction cuvette. Formate oxidation was followed by the reduction of benzyl viologen at 555 nm ($\varepsilon_{555\ nm}(BV^+) = 12$ mM^{-1}.cm^{-1}). CO_2 reduction activity was followed by the oxidation of methyl viologen at 578 nm ($\varepsilon_{578\ nm}(MV^+) = 9.7$ mM^{-1}.cm^{-1}). One enzymatic unit (U) was defined as the amount of FDH capable of oxidizing (or reducing) 1 µmol of formate (or CO_2) per minute. For FdhAB activity, 1.4 nM of the enzyme was used (final concentration in the assays), whilst for FdhA activities reported, 2.8 nM of the enzyme (initial assays and O_2 exposure assays) or 0.9 nM of the enzyme (final assays; in these the pre-incubation step was performed with 500 mM DTT) were used.

4.8. O_2 Exposure Assays

Anaerobic purified FdhA in aerobic sample buffer (20 mM Tris-HCl pH 7.6, 10 mM NaNO$_3$, 10% (v/v) glycerol) with a concentration of 2.8 µM was exposed to air for 8 h at room temperature. At 0 h, 1 h, 2 h, 4 h, 6 h, and 8 h samples were collected, and formate oxidation and CO_2 reduction activities were measured as described in the Activity Assays and Kinetic Analysis subsection.

4.9. Circular Dichroism Spectroscopy

FdhA and FdhAB in a concentration of 0.04 mg/mL and 0.20 mg/mL respectively, in 20 mM Tris HCl Buffer pH 7.6 with 10 % (v/v) glycerol were analysed using a 0.1 cm path length quartz cuvette in a Jasco J-815 CD spectropolarimeter from 200 nm to 250 nm with a data pitch of 0.1 nm and accumulation of 2 scans, using a scanning speed of 50 nm/min. To the results obtained, the spectra of the buffer isolated was subtracted. The results were analysed using Beta Structure Selection webserver (BeStSel)[119,120] for secondary structure prediction.

4.10. Thermal Shift Assay

The samples recovered from the circular dichroism spectroscopy were concentrated to values of 0.05 mg/mL and 0.45 mg/mL for FdhA and FdhAB, respectively, in 20 mM Tris HCl Buffer pH 7.6 with 10 % (v/v) glycerol. The Applied Biosystems Protein Thermal Shift Dye kit was used to determine the melting temperatures of both proteins, with the enzyme being mixed with the dye diluted in the same buffer, with the melting curve being recorded from 25 to 99 °C on the QuantStudio 7 Flex Real-time PCR System from Applied Biosystems.

4.11. Photocatalytic Assays

The photocatalytic reactions were carried out in 11 mL flasks with a working volume of 5 mL, sealed using Aluminium Crimp Cap and grey butyl rubber stoppers. Reactions were assembled under anaerobic conditions inside a COY Anaerobic Chamber under an atmosphere of 2% H_2/98% N_2 or using syringes. The reaction buffer used was 20 mM Tris HCl pH 7. TiO_2 nanoparticles (Acros Organics) (0.4 – 2.7 mg/mL) were dispersed in reaction buffer using a vortex. When Eosin Y was used in combination with TiO_2 nanoparticles, the dye was added after the dispersion of the nanoparticles in the solution, and a 15-minute incubation under agitation was done in the dark. When TiO_2 nanoparticles were not used, these last two steps were not performed, and Eosin Y was directly added to the reaction flask. Eosin Y concentration used varied between 0.22 mM and 0.88 mM, as discriminated in the results section. Furthermore, 100 mM TEOA (pH 1.9) or 100 mM of a mixture of Cystein/HCl-Cystein (6:1 ratio) was added, followed by 100 mM of sodium hydrogen carbonate, as CO_2 source. The enzyme (12 nM), previously incubated with 50 mM of DTT for 5 minutes (final concentration of DTT in the assays 0.016 mM) was the last component added before light exposure. The flasks were then exposed to the light source which was either LEDs that emit light at a wavelength of 470 nm with an intensity of 7 W/m^2 or a Solar Simulator that consists of a 300 W Xenon Lamp (Sirius-300P, Zolix) that emits light at wavelengths from 320-2500 nm (with an intensity of 7000 W/m^2 from 320-450 nm and of 35 000 W/m^2 above 400 nm). When IR light attenuation and slight attenuation of UV light was performed, a SCHOTT KG1 optical filter glass was used, with a transmittance of light of 62% at 320 nm, raising to values above 90% at 350 nm up to 610 nm, declining afterwards. The experiments using LEDs were performed inside a 4 °C room, keeping the temperature of the systems at 25 °C during the whole experiment. In the case of the Solar Simulator, the refrigeration was done using a water bath set to 25 °C, but during the light incidence, the temperature rose to 35 °C. The flasks were under continuous stirring during the whole process.

At 0, 3, 6, and 24 h for the LEDs and at 0, 1.5, and 3 hours for the Solar Simulator, 200 μL of each reaction were retrieved and centrifuged at 17 000 g for 10 minutes at 4 ºC and frozen at − 20 °C. Formate concentration was then determined by high-performance liquid chromatography (HPLC), using a Waters Alliance 2695 chromatographer (Waters Chromatography, Milford, MA) connected to a Waters 2996 Photodiode Array Detector set at 206.5 nm. Chromatography separation was undertaken using an Aminex HPX-87H column (300 x 7.8 mm), 9 μm particle size (Bio-Rad, Hercules, California) and set at 60 °C. Elution was carried out isocratically, at a flow rate of 0.5 mL.min^{-1}, with 0.005 M of H_2SO_4 and the volume injected was 100 μL. The formate retention time obtained was 15.9 min. Data acquisition was accomplished with the Empower 2 software (Waters Chromatography).

4.12. *In silico* Stabilization of the Enzyme

To perform the computational stabilization of FdhA and FdhAB an unpublished x-ray structure of FdhAB is used. The full structure was used to design the FdhAB variants, while for the FdhA simplified version, the β-subunit was removed from the structure. The stabilization of the enzyme can be divided into three parts as previously done in PROSS protocol[109]: preparation of the structure, filterscan, and design. In all steps, the energy function ref2015 was used with harmonic backbone coordinate restraints

set to the coordinates observed in the crystal structure[121]. In the case of the design step, a hard-repulsive version of ref2015 was combined with a soft-repulsive version[121].

The preparation step included the modelling of the active site as a rigid body. For this, the active center was removed from the structure (MGDs, W, and S atoms), hydrogens were added at pH 7.2 and the active site structure was relaxed using Avogadro[122]. The conserved selenocysteine (SeCys$_{192}$) was modelled as a Cysteine. Rosetta commons was used for all the procedures stated if not said otherwise. For Rosetta to be able to recognize the active center, a parameters file was generated. The active site was then again added to each structure and the structures were scored and renumber, followed by the relaxation of each structure being carried using the Rosetta relax protocol.

The filterscan protocol was performed as described in PROSS[109], with the PSSM used being generated through a MSA using PSI-BLAST[123] with refseq_selected database. The sequences were clustered using CD-HIT[124] to eliminate redundant sequences. The computational mutation scanning applied used the Filterscan mover in Rosetta, with the mutations allowed having a score in the PSSM ≥ 0, with eight energy thresholds being applied: ΔGG ≤ +0.5, ΔGG -0.45, ΔGG ≤ -0.75, ΔGG ≤ -1.0, ΔGG ≤ -1.25, ΔGG ≤ -1.5, ΔGG ≤ -1.8 and ΔGG ≤ -2.0 kcal/mol, using the ref2015 score function[121]. The design protocol was performed using Rosetta combinatorial sequence design[109]. It is worth noting that in the case of the FdhAB optimization, the filterscan and design steps were performed in each subunit separately, however in the presence of the whole complex. The results obtained for a given threshold for the α-subunit design were then subjected to β-subunit design using the same threshold. Furthermore, in both cases, the residues at 4 Å from the active center and from the iron-sulphur clusters were fixed and restricted, with the selection being carried using Visual Molecular Dynamics (VMD) software[125].

The analysis of each design was done using FeaturesReporter suite[126]. For the determination of the core region of the protein, residue burial was used, with a residue being defined as a core residue if it has >22 and >75 neighbouring non-hydrogen atoms within 10 Å and 12 Å, respectively (as defined in [109]). The isoelectric points were calculated using ExPASy webserver[127]. The alignment of the simplified FdhA and of the FdhAB with their respective designs was done using JALVIEW[128].

5. Results

5.1. Generation of a FdhA-expressing *D. vulgaris* Mutant

Since the catalytic center of FdhAB is present in the α-subunit, which also harbours an iron-sulphur cluster, it can be theorized that the large subunit of the enzyme harbours the minimal number of centers crucial for catalysis.

To generate a FdhA-expressing *D.vulgaris* mutant, a restriction site for *Eco*RI was added flanking the Kan_RBS downstream the stop codon of *FdhA* (Figure 26 in Annex section) by direct mutagenesis. The plasmid already contained an *Eco*RI restriction site downstream the *FdhB* gene, upstream the pBG1 gene (origin of replication in *D. vulgaris*). This allowed to excise the coding sequence of the small subunit from the expression plasmid. Then, digestion with *Eco*RI was performed, with a subsequent ligation step. The generated plasmid was transformed into *E. coli DH5α* cells and grown overnight at 37 ºC. A colony PCR was performed with the resulting amplifications being shown in the agarose gel in Figure 11.
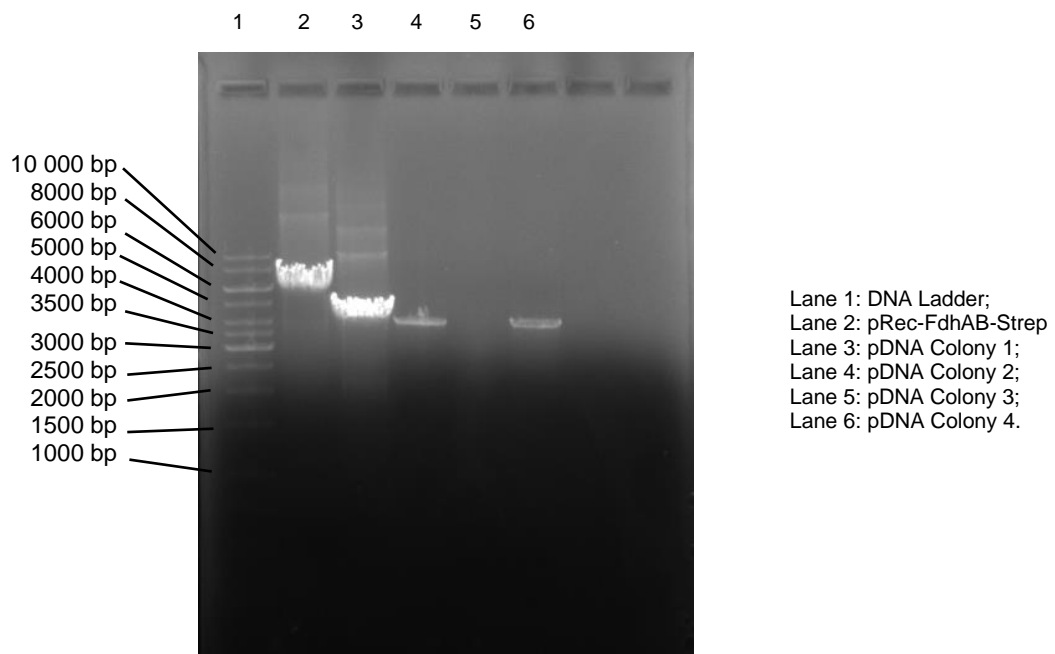
Figure 11: Agarose gel (0.8%) to analyse the selected colonies to confirm the plasmid identity. Lane 1: Thermo Fisher GeneRuler 1 kb DNA Ladder (10 000 bp, 8000 bp, 6000 bp, 5000 bp, 4000 bp, 3500 bp, 3000 bp, 2500 bp, 2000 bp, 1500 bp, 1000 bp); Lane 2: pRec-FdhAB-Strep; Lane 3: pDNA Colony 1; Lane 4: pDNA Colony 2; Lane 5: pDNA Colony 3; Lane 6: pDNA Colony 4.

The primers used flanked the region where the removed segment was situated, therefore, a product with 7875 bp was expected on the original plasmid, and in the positive colonies, a product with 3893 bp was expected. When observing the gel, the behaviour of both the original plasmid and colony 1 was not the expected one, with a band that did not possess the typical shape and was not compacted. The band that corresponds to the product of PCR from the amplification of the pRec-FdhAB-Strep plasmid was broad, being found between the 8000 bp marker and the 6000 bp marker, with the expected size being 7875 bp. In colony 3, no amplification was observed. When comparing the bands for colonies 1, 2, and 4, it was possible to observe that the first had a different behaviour and size. Colonies 2 and 4 had sizes near the 4000 bp marker, which suggested that our intended product was present. For

colony 1, since the band was above the 4000 bp marker and its migration had a strange behaviour, we could not conclude that the plasmid matched the product we sought to obtain.

As such, colonies 2 and 4 were selected for sequencing and the results confirmed that the pRec-FdhA-Strep plasmid (the obtained plasmid can be found in Figure 27 in the Annex Section) was successfully obtained.

Afterwards, the plasmid from colony 4 was successfully transformed into a *ΔfdhAB D. vulgaris* Hildenborough, generating a mutant strain expressing the recombinant FdhA. To confirm that no mutation was inserted in the plasmid during the transformation in *D. vulgaris* cells, the plasmid was sequenced. The result confirmed that the *fdhA* gene was correct in the mutant strain (results not shown).

The new strain was grown, and the cells were disrupted using BugBuster Protein Extraction Reagent to obtain the soluble fraction. Furthermore, in order to evaluate the expression of FdhA in the soluble fraction, a Western-blot against Strep-tag and an activity-stained native gel were performed (Figure 12A and Figure 12B, respectively).
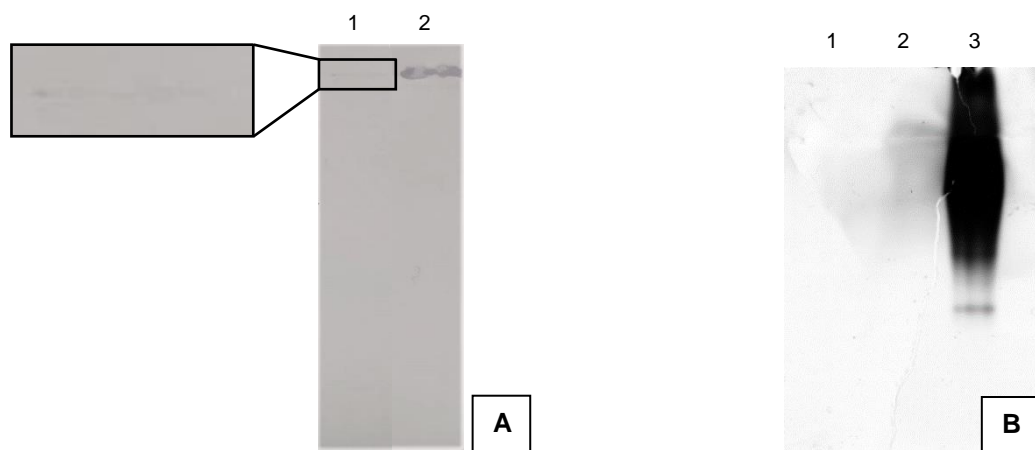
Figure 12: (A) Western-blot of the soluble extract collected from *D. vulgaris* Hildenborough strain expressing the recombinant FdhA protein (Lane 1 – 10 µg) and a strain expressing the recombinant FdhAB protein (Lane 2 - 10 µg) **(B)** Native-polyacrylamide gel (12%) revealed for formate oxidation (reduction of benzyl viologen) with a soluble extract from a the strain expressing the FdhA protein (Lane 1 – 50 µg) and the one expressing the FdhAB protein (Lane 3 – 50 µg).

The results from the Western-blot in Figure 12A show an extremely tenuous band in lane 1 that was only visible when zooming in the section corresponding to the molecular weight of the α-subunit (108.8 kDa), which indicates that the mutant protein is being expressed at a low level. Furthermore, in lane 1 of Figure 12B no formate oxidation activity was detected for the FdhA protein, even when protein quantities were increased (up to 200 µg – results not shown). This clearly contrasts with the band corresponding to the FdhAB protein in the Western-blot (Figure 12A – Lane 2) and in the native gel (Figure 12B – Lane 3), where the expression is nitid and a highly active enzyme is obtained.

With these results, it is possible to conclude that even though FdhA is being expressed, its expression and/or degradation kinetics are altered in comparison to the FdhAB protein. When protein expression is somewhat impaired it can indicate that the protein has stability issues, something that will be further discussed. Nonetheless, with the result from the native gel, it was not possible to conclude that FdhA has very low or no activity. The negative result in the activity-stained gel can be due to

increased oxygen susceptibility of the simplified enzyme, therefore losing activity when exposed during purification or if the activity is too low, it can be below the detection limit of the technique.

5.2.   Purification and Characterization of FdhAB and FdhA

Since at this point the exact reasons behind the low expression of FdhA and the lack of activity in the gel were unknown, we decided to purify the mutant to do a closer evaluation. To study if oxygen was damaging the protein, we performed both anaerobic and aerobic purifications. Furthermore, we purified the strep-tagged version of FdhAB aerobically, which is the standard procedure, in order to compare both proteins. For this, *D. vulgaris* cells expressing the strep-tagged version of FdhA and the strep-tagged version of FdhAB were grown. The cells were retrieved and broken in a French Press Cell as described in the method section to obtain the soluble fraction. This fraction was purified using a Strep-tactin column, eluted with D-desthiobiotin.

The UV-visible spectrum of the elutions (Elution A and B) for the FdhAB protein and the mutant (both aerobically and anaerobically) were analysed (Figure 13).
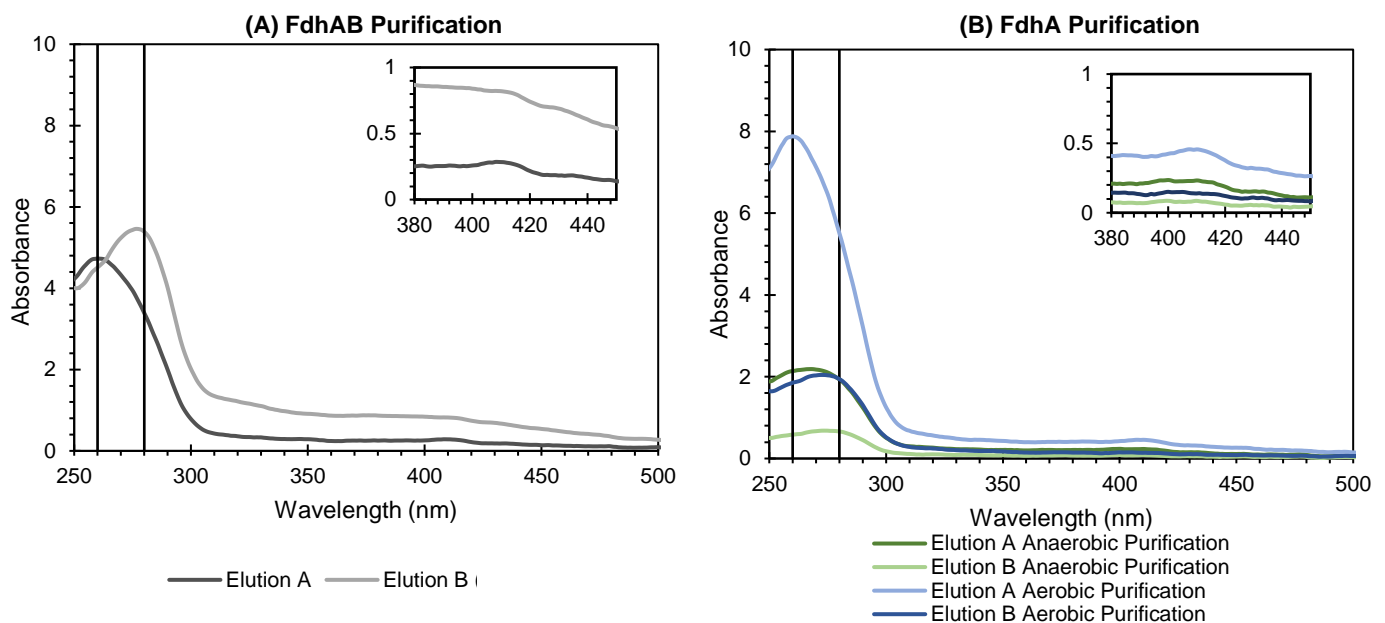
Figure 13: In **(A)**, the UV-visible spectrum of elution A (dark grey) and elution B (light grey) of the aerobic purification of the FdhAB enzyme. In **(B)**, the UV-visible spectrum of elution A and B of the aerobic purification of the FdhA protein (light blue and dark blue, respectively) and of the anaerobic purification (dark green and light green, respectively). In both cases, the signal between the 380 and the 450 nm is highlighted on the top right of the graphical representation. Two black lines intersect the x axis at 260 nm and 280 nm.

The spectrum of Elution A from FdhAB purification (Figure 13A) and of the aerobic and anaerobic purification from FdhA (Figure 13B) presented a peak of absorbance that deviated from the typical 280 nm, being closer to the 260 nm mark. In opposition, the peak in Elution B was closer to the 280 nm wavelength that characterizes protein absorption[129]. This shows that in all purifications, Elution A was contaminated with nucleic acids (that absorb at 260 nm[129]), explaining the rationale behind this separation. In proteins with iron-sulphur clusters, a characteristic absorbance is expected at 410 nm due to the presence of these groups[130]. In the case of FdhAB, the protein has four iron-sulphur clusters (1 in the α-subunit and 3 in the β-subunit), while FdhA only has the one from the α-subunit. These groups generate a characteristic shoulder at 410 nm that was easily identified when observing the zoomed

representation correspondent to the Elution B of FdhAB purification. In the case of FdhA, in both aerobic and anaerobic purifications, this shoulder was not easily detected, possibly due to the lower amount of protein obtained (due to the decreased expression of the mutant) and the presence of fewer groups that absorb at 410 nm in comparison to FdhAB. In Elution A of the FdhAB purification and of the aerobic purification of FdhA, at 410 nm the presence of a peak was more noticeable than of a shoulder, which could indicate that there was contamination with cytochromes that absorb strongly at 410 nm[131], further supporting this separation. For the FdhAB protein, the absorbance at 410 nm was used to calculate protein concentration using the Lambert-Beer Law with the known molar absorption coefficient ($\varepsilon_{410\ nm}$ = 43.25 mM^{-1}.cm^{-1}). For the mutant protein, since no $\varepsilon_{410\ nm}$ has been characterized, protein concentration was calculated with the BCA method. The results can be found in Table 1.

Table 1: Volume obtained after purification of each sample. Representation of the protein concentration in mg/mL and µM obtained by Lambert-Beer Law for FdhAB (elution A and B) and with BCA method for FdhA, in both elution A and B purified aerobically and anaerobically. Protein content for each sample is also represented along with the yield (amount of protein retrieved per amount of cells).

Sample		Volume (mL)	[FdhAB] (mg/mL)	[FdhAB] or [FdhA] (µM)	Protein (mg)	mg protein/ g cells
FdhAB	Elution A	0.50	0.905	0.453	6.54	1.17
	Elution B	0.35	12.4	4.34	89.7	16.0
FdhA Aerobic Purification	Elution A	0.35	2.11	0.739	19.0	2.24
	Elution B	0.47	5.86	2.75	52.8	6.21
FdhA Anaerobic Purification	Elution A	0.40	-*	-*	-*	-*
	Elution B	0.58	3.09	1.79	27.9	5.58

*Data not available.

The degree of purity of each sample was analysed by SDS-polyacrylamide gels for Elution A and Elution B of the FdhAB and of the mutant protein (Figure 14).
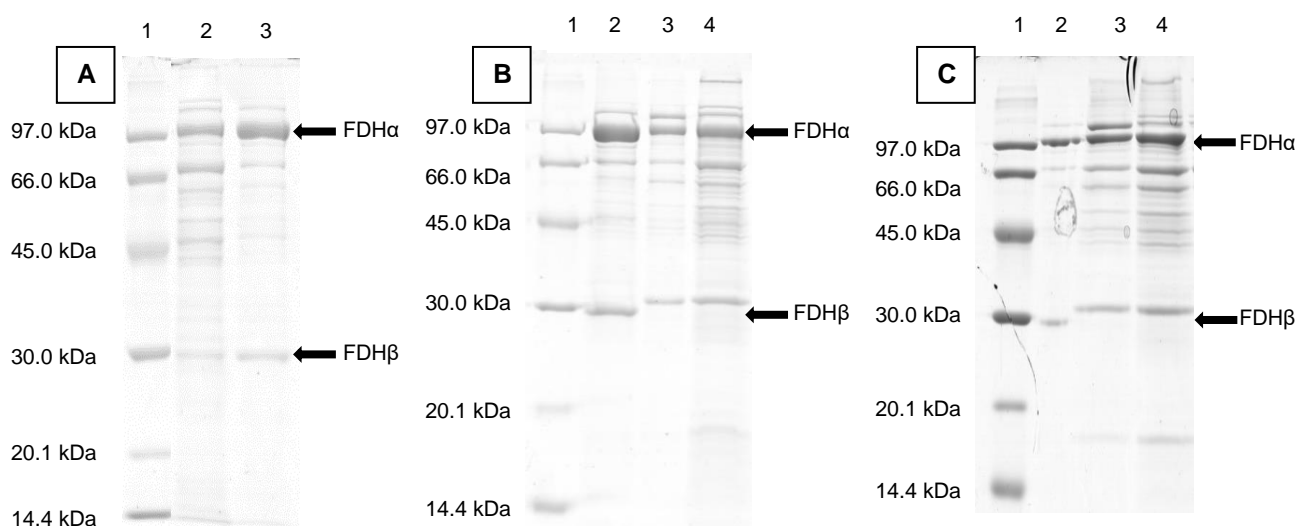
Figure 14: SDS-polyacrylamide gel (12%) to analyse the elutions obtained after aerobic purification of FdhAB **(A)**, aerobic **(B)** and anaerobic **(C)** purification of FdhA. In **(A), (B) and (C)**: Lane 1: Amersham Low Molecular Weight Calibration Kit for SDS Electrophoresis (97.0 kDa, 66.0 kDa, 45.0 kDa, 30.0 kDa, 20.1 kDa and 14.4 kDa). In **(A)**: Lane 2: 10 µg Elution A FdhAB purification; Lane 3: 10 µg Elution B FdhAB purification. In **(B)**: Lane 2: 10 µg Elution B FdhAB purification; Lane 3: 10 µg Elution B FdhA purified aerobically; Lane 4: 10 µg Elution A FdhA purified aerobically. In **(C)**: Lane 3: 4 µg Elution B FdhA purified anaerobically; Lane 4: 6.55 µg Elution A FdhA purified anaerobically.

The FdhAB protein with the strep-tag has a molecular mass of 138.32 kDa, with the α-subunit having 108.8 kDa and the β-subunit 26.4 kDa, which was consistent with the bands observed in the lanes of the Elutions A and B (Figure 14A). When comparing Elution A and B (lane 2 and 3, respectively), it was evident that Elution A showed a higher number of contaminants and with higher intensities comparing to Elution B, besides having nucleic acid contamination, as previously seen in the UV-vis spectrum. Therefore, Elution B was purer and is hereon named FdhAB, being the one used for the kinetic characterization of the FdhAB enzyme.

Regarding FdhA, in both aerobic and anaerobic purification, it was possible to observe that the protein was successfully obtained in Elution A and B (Figure 14B and Figure 14C). There was a band close to the 30 kDa marker, which was near the molecular mass from β-subunit, nonetheless, it was above the band from the marker and, in comparison with the FdhAB enzyme in lane 1 from both Figure 14B and Figure 14C, it was nitid that it was a contaminant, which was not present in the FdhAB protein purification. The profile in terms of contaminants both in Elution A and Elution B was quite similar in the sample purified aerobically and anaerobically. As seen for the FdhAB enzyme, this separation between Elution A and B helps us obtain a purer sample of the protein. Nonetheless, it is clear that Elution B from both aerobic and anaerobic purification of the mutant enzyme was not as pure as the one from the FdhAB (present in lane 2). For these reasons, from hereon, Elution B of the mutant was named FdhA purified aerobically or anaerobically and was the fraction used for the following kinetic analysis.

To analyse the kinetic behaviour of FdhA for both formate oxidation and CO_2 reduction and to compare its performance to the FdhAB enzyme, formate oxidation activity was followed by the reduction of benzyl viologen at 555 nm and CO_2 reduction activity by the oxidation of methyl viologen at 578 nm. The assays were performed with and without DTT activation. In the former, the enzyme was incubated with DTT for 5 minutes before the addition to the assay and of the activity measurement. In Table 2, it is possible to find the k_{cat} (s^{-1}) as well as the specific activity of both FdhAB and the simplified version of the enzyme, in the two conditions mentioned above.

Table 2: Kinetic characterization of FdhAB and the mutant FdhA for both formate oxidation and CO_2 reduction, with and without incubation with DTT (50 mM). For both reaction directions, catalytic constant (k_{cat} (s^{-1})) and specific activity (U/mg) are represented.

Sample		Formate Oxidation		CO_2 Reduction	
		k_{cat} (s^{-1})	Specific Activity (U/mg)	k_{cat} (s^{-1})	Specific Activity (U/mg)
FdhAB	Without DTT	103 ± 2	45 ± 1	2.85 ± 0.98	1.24 ± 0.42
	With DTT	961 ± 84	417 ± 36	234 ± 13	102 ± 6
FdhA Aerobic Purification	Without DTT	-*	-*	-*	-*
	With DTT	0.24 ± 0.03	0.13 ± 0.02	1.51 ± 0.03	0.81 ± 0.06
FdhA Anaerobic Purification	Without DTT	ND**	ND**	0.55 ± 0.00	0.30 ± 0.02
	With DTT	1.70 ± 0.05	0.92 ± 0.03	1.00 ± 0.03	0.54 ± 0.02

*Activities were not measured without DTT from the sample purified aerobically due to limitations in the quantity of protein obtained.

**Activity not detected.

The k_{cat} was determined both with a pre-incubation with DTT and addition of DTT to the assay, or without any use of DTT. The assay where no DTT was used functions as a control since it is typical

of formate dehydrogenases to require incubation with reduction agents in order to gain full activity when purified aerobically[39,40]. This behaviour was quite evident for the FdhAB enzyme where formate oxidation with DTT was more than 9-fold higher than without DTT (961 ± 84 s^{-1} and 103 ± 2 s^{-1}, respectively). In the case of CO_2 reduction, this was even more exacerbated with the activity with DTT being 82-fold higher than without DTT (234 ± 13 s^{-1} and 2.85 ± 0.98 s^{-1}, respectively).

When comparing the activities obtained for FdhAB and for FdhA, formate oxidation and CO_2 reduction was significantly decreased both in the case of aerobic and anaerobic purification of the simplified enzyme. This suggests that the β-subunit, even though being just an electron transfer subunit, which in theory is not crucial for catalysis, can have an important structural and functional role besides the transfer of electrons.

When comparing the aerobic and anaerobic purification of FdhA, even though for both the activity was significantly decreased in comparison to FdhAB, some changes were seen between them. When the activity was measured in the presence of DTT, it was possible to register a formate oxidation activity of 0.24 ± 0.03 s^{-1} for the aerobic purification of the enzyme. On contrary, this value for the anaerobic purification was higher, with an activity of 1.70 ± 0.05 s^{-1} being registered. Curiously, in the case of CO_2 reduction, when the protein was purified exposed to oxygen the activity was higher than when the protein was purified anaerobically (1.51 ± 0.03 s^{-1} and 1.00 ± 0.03 s^{-1}, respectively). It seems that O_2 exposure can somehow impact activity, with the effects being different if the enzyme is catalysing formate oxidation or CO_2 reduction.

Furthermore, it is known that when FdhAB is purified anaerobically, the activity with or without DTT remains unchanged (data not published), but this was not the case for the anaerobic purification of FdhA, where no activity for formate oxidation was observed (ND and 1.70 ± 0.05 s^{-1}, without and with DTT, respectively) and with a 2-fold decrease for CO_2 reduction (0.55 ± 0.00 and 1.00 ± 0.03 s^{-1}, without and with DTT, respectively). The exact role of DTT is not yet elucidated, nonetheless, this experimental result adds another layer of complexity to its understanding.

All these results show that the elimination of the β-subunit significantly hampers activity, implying that its presence is important for the enzyme. When piecing together these results with the lower expression of the FdhA protein, it can indicate that the α-subunit faces major stability issues in the absence of the β-subunit. Furthermore, purification in the presence of oxygen and its absence leads to different and unexpected effects on catalysis.

To understand if the simplified enzyme is less tolerant to oxygen, the anaerobically purified FdhA was subjected to oxygen at room temperature, with the activity of the enzyme being measured for both formate oxidation and CO_2 reduction to analyse the effect of oxygen over time (Figure 15).

Regarding the formate oxidation rate (blue circles), the turnover remained fairly constant during 8 hours of exposure, indicating that no inactivation of the enzyme was occurring. Concerning CO_2 reduction, there was a decrease from 0 h to 2 h, but afterwards, there was an increase in the reaction rate (from 4 h to 8 h). This fluctuation may not be significant and occurs due to low reproducibility, further exacerbated by the high variability of the data at 0 h. For these reasons, it is probably more likely that CO_2 reduction was also not significantly affected during the time of the experience. Even though this indicates that the protein might not be susceptible to O_2 in the oxidized state, the data analysed only

accounts for an 8 h exposure, while purification tends to take more than 2 days. Nonetheless, the results indicate that the enzyme is quite stable and allows aerobic handling and purification, as seen for FdhAB in a previous study[39].
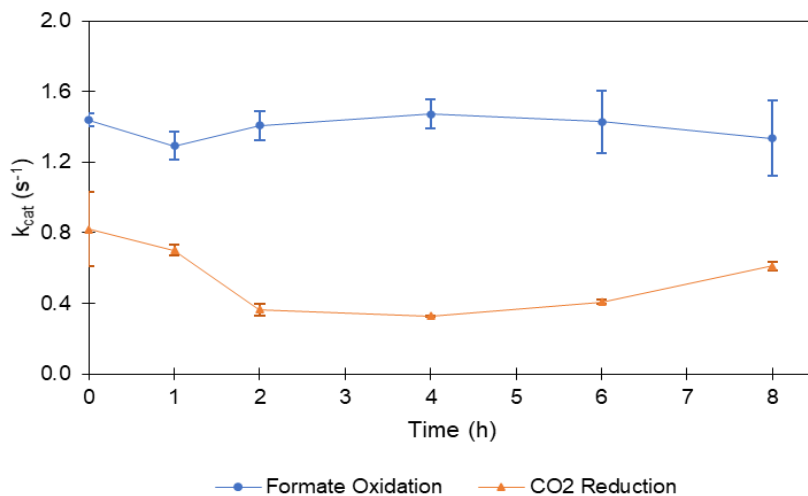
Figure 15: FdhA purified anaerobically was incubated aerobically at room temperature in sample buffer. At several timepoints, the protein was retrieved to measure formate oxidation activity (blue circles) and CO_2 reduction activity (orange triangles). The assays were performed in triplicate with the error bars representing the standard deviation.

To structurally characterize FdhA, we needed to obtain purer samples, therefore a second purification step of the aerobically purified FdhA was performed. Both Elutions A and B were loaded, individually, on a Q-resource™ column, equilibrated with sample buffer. The elution was done by increasing NaCl concentration (chromatograms can be found in Figure 28 and Figure 29 in the Annex Section, for Elution A and B, respectively). The eluted samples are divided into several fractions, quantified, and analysed by SDS-PAGE gel to identify FdhA (Figure 16A and B, for Elution A and B purification, respectively).
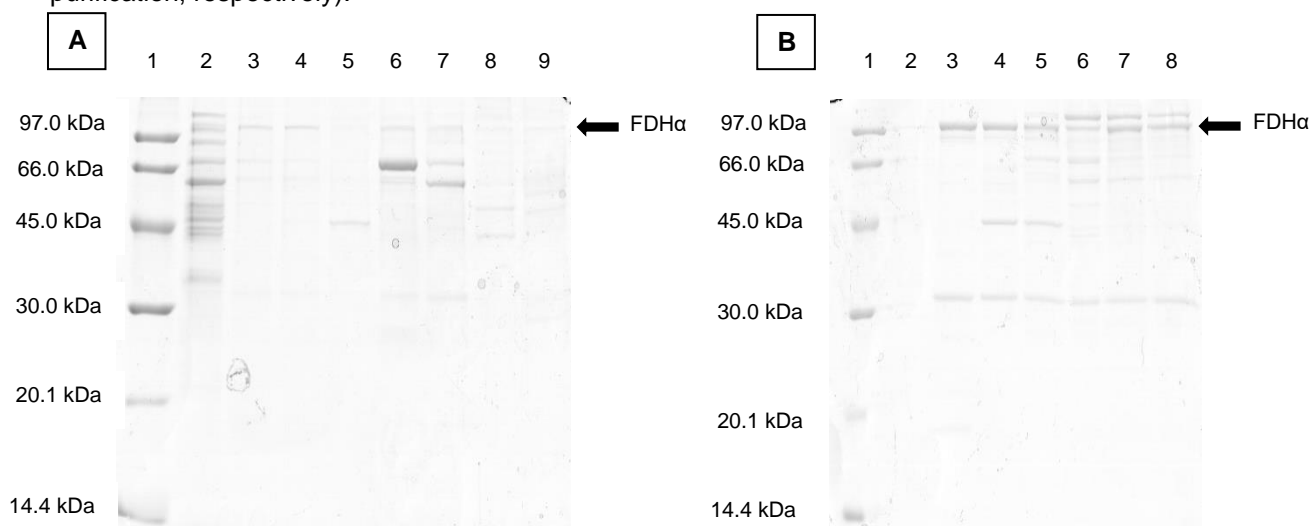
Figure 16: SDS-polyacrylamide gel (12%) to analyse the fractions eluted from the second step of aerobic purification of elution A **(A)** and elution B **(B)** of FdhA. In **(A) and (B)**: Lane 1: Amersham Low Molecular Weight Calibration Kit for SDS Electrophoresis (97.0 kDa, 66.0 kDa, 45.0 kDa, 30.0 kDa, 20.1 kDa and 14.4 kDa) (5 µg). In **(A)**: Lane 2: Fraction 0 (8.42 µg); Lane 3: Fraction 1 (5 µg); Lane 4: Fraction 2 (4.22 µg); Lane 5: Fraction 3 (2.01 µg); Lane 6: Fraction 4 (7.24 µg); Lane 7: Fraction 5 (4.97 µg); Lane 8: Fraction 6 (10.04 µg); Lane 9: Fraction 7 (0.66 µg). In **(B)**: Lane 2: Fraction 0 (1.25 µg); Lane 3: Fraction 1 (5 µg); Lane 4: Fraction 2 (5 µg); Lane 5: Fraction 3 (5 µg); Lane 6: Fraction 4 (5 µg); Lane 7: Fraction 5 (5 µg); Lane 8: Fraction 6 (5 µg).

The chromatographic profile was quite different for both elutions. In Elution A, the majority of the contaminants eluted in Fraction 0 (Lane 2 in Figure 16A), whilst for Elution B this did not occur (in both cases, this corresponds to 0% NaCl). For both elutions, FdhA was widespread in several fractions even though at different intensities (in Elution A, it was present in Lanes 2, 3, 4, 6, and 7, and in Elution B in Lane 3 up to Lane 8), which led to protein loss during this purification step. FdhA was more enriched in Fractions 3 and 4 in Elution A, which were pooled together to generate the purified FdhA and was used for further structural characterization, and in Fraction 3 from Elution B, which was used for kinetic analysis.

Since at this point, protein quantity was a major limitation to perform further assays, we sought to estimate an $\varepsilon_{410\,nm}$ in order to minimize protein used in quantification methods. For this, the pure sample generated from the purification of Elution A was quantified by BCA and was used to estimate the $\varepsilon_{410\,nm}$ of FdhA. The results can be found in Table 3.

Table 3: Quantification by BCA of the pooled fractions 1 and 2, with FdhA concentration obtained in mg/mL and µM. Estimated $\varepsilon_{410\,nm}$ using the known concentration obtained by BCA and the absorbance at 410 nm obtained.

Sample	[FdhA] (mg/mL)	[FdhA] (µM)	$\varepsilon_{410\,nm}$ (mM^{-1}.cm^{-1})
P1+P2	0.137	1.24	91.13

Furthermore, to analyse whether the FdhA mutant was properly folded, the same sample was used for circular dichroism spectroscopy (CD) for comparison with the FdhAB protein previously purified. CD is widely used for rapid and simple prediction of secondary structure when it is not possible to obtain high-resolution structures (for instance, X-ray crystallography or solution NMR). This technique essentially measures the differential absorption between left circularly polarized light and right circularly polarized light, which is dependent on the conformation and environment surrounding peptide bonds, therefore allowing secondary structure prediction[120]. The spectrum obtained for FdhA and FdhAB were quite different (Figure 17).
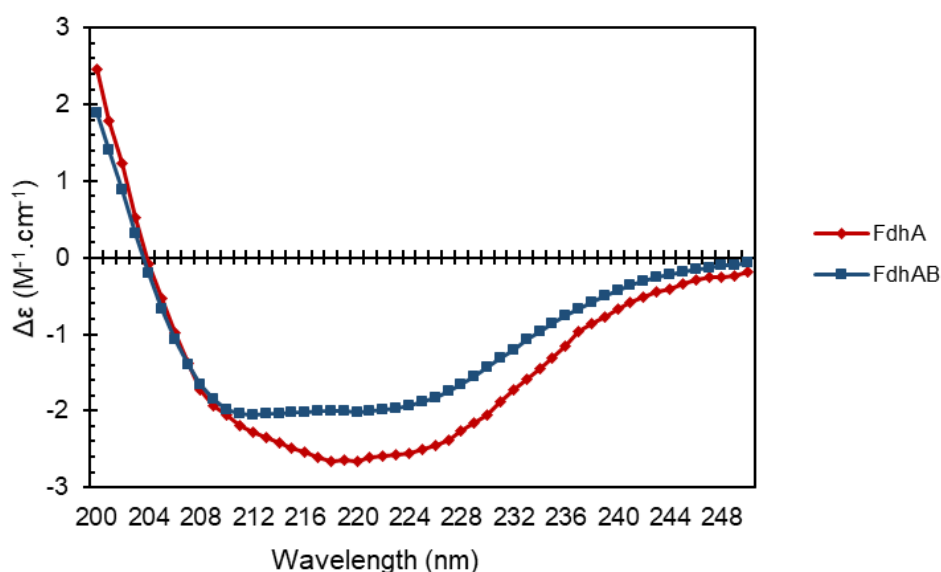
Figure 17: Graphical representation of the delta epsilon (Δε) of FdhA (red diamonds) and FdhAB (blue squares) in function of the wavelength of light used.

To estimate the % of each type of secondary structure using CD data, several algorithms have been developed. In this case, BeStSel (Beta Structure Selection) was used to analyse and compare the folding of FdhAB and FdhA (Figure 18), since it provides secondary structure prediction in a more detailed manner, with higher accuracy mostly in the prediction of β-sheets, which usually have a higher spectral diversity causing limitations in their estimation[119,120].
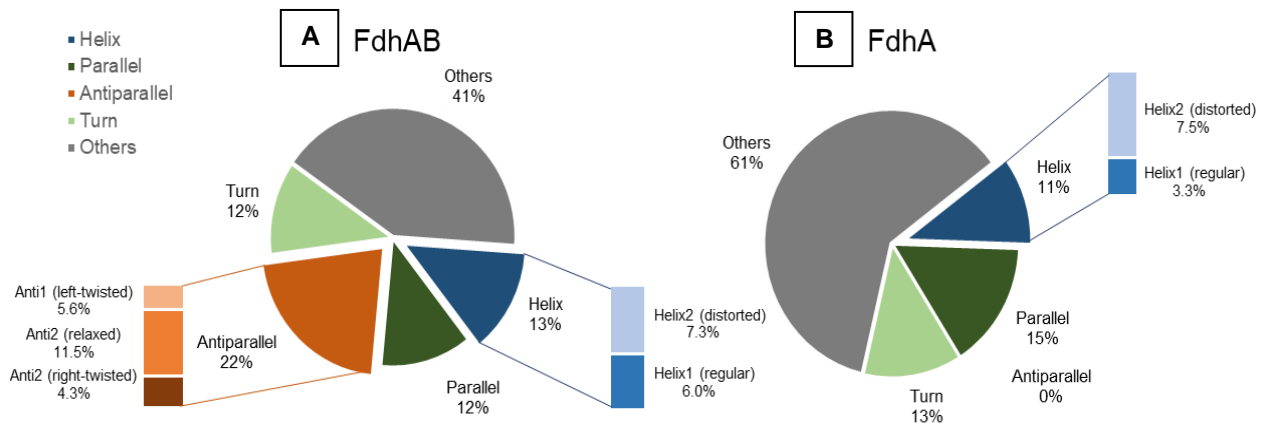
Figure 18: Predicted secondary structure composition (%) for FdhAB **(A)** and FdhA **(B)** obtained using BeStSel[117,118]. The secondary structure elements are divided in α-helixes (which are further divided in regular α-helixes – middle part – and distorted α-helixes – the two residues at each end of the helix –), parallel β-strands, antiparallel β-strands (which are further divided in left-twisted, relaxed and right-twisted antiparallel β-strands), turns, and others (where structures such as 3_{10}-helix, π-helix, β-bridge, bends, loop/irregular and invisible regions of the structure are assigned).

BeStsel is able to assign eight different types of secondary structures: α-helixes (which are further categorized in Helix1 and Helix2, that consist in the amino acids in the middle part – regular – and the two amino acids of each terminal of the helix – distorted –); parallel β-sheets; antiparallel β-sheets (which are further categorized depending on the degree and orientation of the twist in Anti1, Anti2 and Anti3, corresponding to left-twisted, relaxed and right-twisted, respectively); turns; others (which are all elements that BeStsel cannot discriminate or are unorganized such as 3_{10}-helix, π-helix, β-bridge, bends, loop/irregular and invisible regions). This analysis shows that some elements of the predicted secondary structures of FdhA and FdhAB had little variances between them, such as α-helixes, turns, and parallel β-sheets (13%, 12% and 12% for FdhAB and 11%, 13% and 15% for FdhA, respectively). Even though the general proportion of α-helixes varied only slightly (2% between the two structures) when observing Helix1 and Helix2 the difference can be attributed to less regular α-helixes in the structure of FdhA (3.3%) in comparison to FdhAB (6.0%). The presence of the β-subunit would not explain this difference since from the oxidized x-ray crystal structure of FdhAB (PDB code:6SDR) it is possible to obtain the % of each secondary structure component and FdhB has even a smaller % of helix-α in comparison to FdhA (about 1% difference).

The major difference in secondary structure was seen in antiparallel β-sheets where there was a complete absence of this type of secondary structure in FdhA, while it comprehended about 22% of secondary structure composition in the FdhAB protein. This difference was quite notorious and, since there was an increase as well in about 20% in the category defined as others for FdhA, it is quite probable that the majority of antiparallel β-sheets is disorganized in the FdhA protein, with some of these

possibly acquiring a different type of secondary folding (for instance, there was an increase in about 3% of parallel β-sheets in FdhA in comparison to FdhAB).

Even though CD data only allows a rough estimation of secondary structure content, with these results it is quite notorious that FdhA has clearly adopted a different folded conformation in comparison to FdhAB. Nonetheless, it is not possible to analyse if the difference is located in a given region of the protein (for instance, the interface between β- and α-subunits) or if it is a more global folding problem.

To further evaluate protein stability, the samples used in the CD analysis were recovered and concentrated in order to perform a Thermal Shift Assay, with the results being presented in Table 4 (the raw data for FdhAB and FdhA can be found in the Annex Section in Figure 30 and Figure 31, respectively).

Table 4: Thermal shift assay of FdhAB and FdhA, with the melting temperate (T_M, temperate at which 50% of the protein is denaturated) being represented as an average value of two replicates, with the error representing the standard deviation.

Sample	T_M (ºC)
FdhAB	78.9 ± 0.1
FdhA	66.3 ± 1.0

The melting temperature (T_M) of FdhAB (78.9 ± 0.1 ºC) was significantly higher than the one of FdhA (66.3 ± 1.0 ºC), further supporting the presence of important structural differences between the two proteins. This is a considerable difference (12.6 ºC) that, in combination with the poor expression and the CD spectroscopy results further supports the hypothesis of a major stability issue.

In order to revaluate the kinetic parameters from FdhA, the purer sample from the purification of Elution B (Figure 16B) was quantified using the $\varepsilon_{410\,nm}$ and new activity assays were performed, with the results being present in Table 5.

Table 5: Kinetic characterization of the FdhA variant for both formate oxidation and CO_2 reduction, incubated with 500 mM DTT. For both reaction directions, catalytic constant (k_{cat} (s^{-1})) and specific activity (U/mg) are shown.

Sample	Formate Oxidation		CO_2 Reduction	
	k_{cat} (s^{-1})	Specific Activity (U/mg)	k_{cat} (s^{-1})	Specific Activity (U/mg)
FdhA (P2)	5.28 ± 0.55	2.86 ± 0.30	1.34 ± 0.08	0.73 ± 0.04

Before this polishing step, an activity of 0.24 ± 0.03 s^{-1} and 1.51 ± 0.03 s^{-1} had been attained for formate oxidation and CO_2 reduction, respectively (Table 2). This polishing step led to a 22-fold increase in formate oxidation activity (5.28 ± 0.55 s^{-1}), while it led to an unexpected but slight decrease in CO_2 reduction activity (1.34 ± 0.08 s^{-1}). The increase in formate oxidation activity may be associated to a more accurate estimation of the protein concentration used in the assay, something that was not observed for CO_2 reduction possibly due to some sort of mild inactivation of the enzyme due to the stress caused by the additional purification step. Furthermore, it is worth noting that to obtain full activity, a higher concentration of DTT (500 mM) was required to activate the protein in comparison with the regular concentration used (50 mM), possibly due to the longer period of oxygen exposure due to the additional purification step, which might have caused some reversible damage to the protein.

In summary, all these results show that it is possible to obtain the simplified version of the protein, only possessing the α-subunit, but with a significant decrease in expression and activity. These observations, combined with the results obtained from CD spectroscopy indicate that the isolated FdhA variant might have a stability problem, related to unfolded or misfolded regions of the protein. The difficulties associated with obtaining enough protein in a pure sample hampered further additional experimental tests that could be performed to get a more concrete notion of the problem.

5.3. Photocatalytic Reduction of CO_2 with FdhAB

One of the most promising applications of FDHs is in photocatalysis, with the use of an endless resource, light, to supply the energy for the catalysed reaction. Furthermore, to gain more insight into relevant mutations that can drive the performance of a system, an easy and commercial photocatalytic setup was created and tested. Therefore, in this section of the work, we aimed to study two different catalytic setups, one using isolated TiO_2 nanoparticles coupled with FdhAB (Figure 19A) and another with the TiO_2 nanoparticles being sensitized by Eosin Y (a commercial and cheap dye) that is known to work well with TiO_2 (Figure 19B). The major difference is that the setup without dye sensitization, a Solar Simulator was used as the source of light, which irradiates light in the UV, visible, and IR range, however, the system is only able to absorb UV light, which is a disadvantage for a future application (since UV light is a minor fraction of the total amount of light that hits the Earth's crust). The dye-sensitized system can be used with visible light, with an LED emitting light at 470 nm being used.
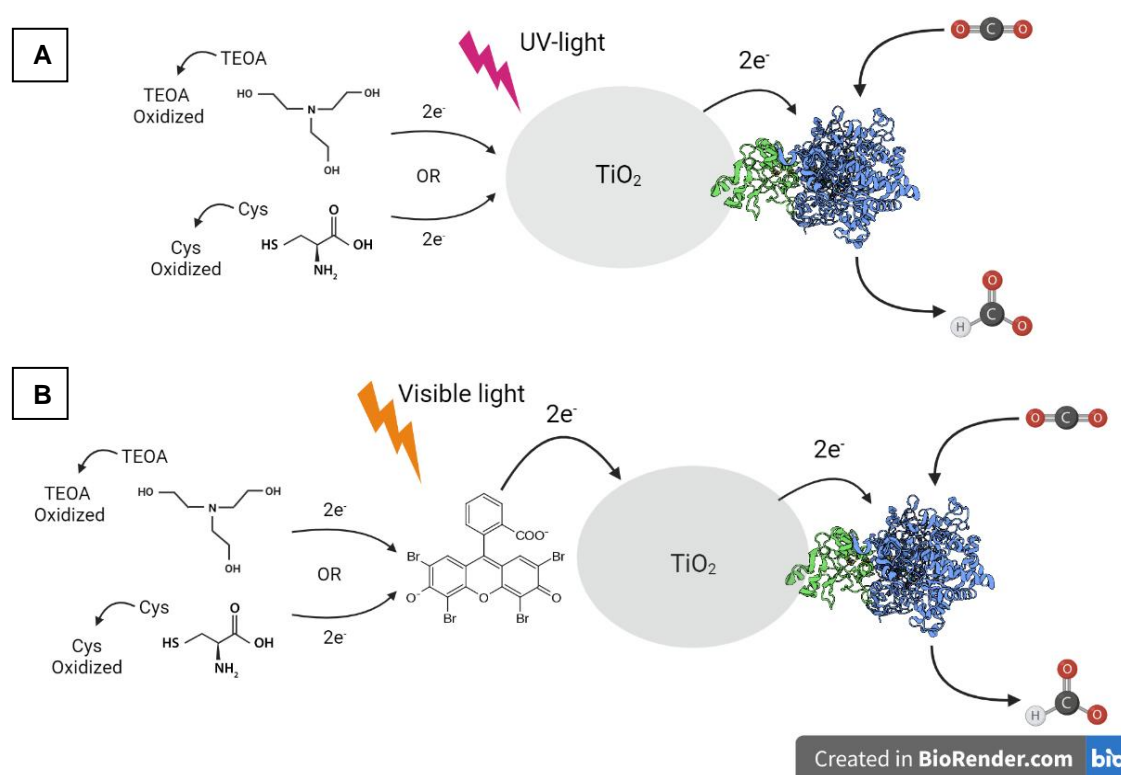
Figure 19: Photocatalytic setups for CO_2 reduction with the highly active FdhAB from *D. vulgaris*. In each system, (A) and (B), two sacrificial electron donors (SED) are used (Cys or TEOA) to regenerate the photoexcited molecule that, in the case of system (A) is TiO_2 nanoparticles, whilst in system (B) is Eosin Y (a commercially available dye). The light, UV for system (A) attained using a Solar Simulator and visible for system (B) using LEDs at a wavelength of 470 nm, excite electrons in the photocatalyst generating an electron-hole pair, with the electron being injected directly to FdhAB (system A) or indirectly through the TiO_2 nanoparticles (system B). The image was created with BioRender.com.

Initially, the goal was to test which SED could be successfully coupled to each of the systems, therefore Cysteine (Cys) and Triethanolamine (TEOA) were experimentally tested. For this, photocatalysis with the system consisting of FdhAB and TiO_2 nanoparticles sensitized with Eosin Y plus either SED (Cys or TEOA) was performed for 24 hours using LEDs that emit radiation at 470 nm. In the case of the system without sensitization, again both SEDs were tested with the TiO_2 nanoparticles being directly excited using a Solar Simulator for 3 hours. The results can be found in Figure 20.
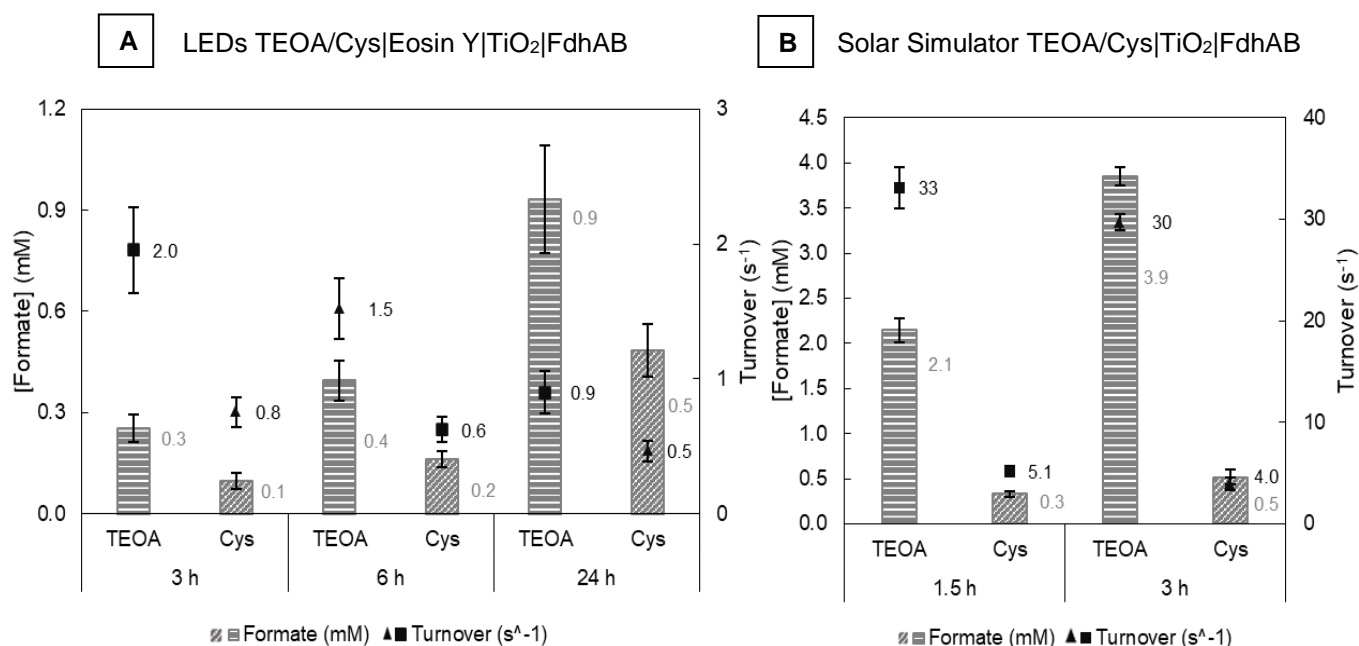
Figure 20: (A) LEDs were used to power the system composed of either Cys or TEOA with Eosin Y, TiO_2 nanoparticles and FdhAB for CO_2 reduction for 24 hours (samples taken at 0, 3, 6 and 24 hours). **(B)** A Solar Simulator was used to the system composed of either Cys or TEOA with TiO_2 nanoparticles and FdhAB for CO_2 reduction for 3 hours (samples taken at 0, 1.5 and 3 hours). In **(A) and (B)** at each timepoint the formate value measured at 0 hours was subtracted. Formate concentration (mM) is represented on the left axis for either TEOA (grey and white horizontal stripes) or Cys (grey and white diagonal stripes). Turnover (s^{-1}) is represented on the right axis, for both TEOA (black squares) and Cys (black triangles). The data for the turnover is represented in black, while for formate production is represented in grey. The errors bars represent the standard deviation of three replicates.

When analysing the performance of each system with the only variable being the SED used, it was quite noticeable that TEOA significantly outperforms Cys, with higher formate production and, consequently, a higher turnover. Using the LEDs, on average the system with TEOA had a two-fold increase in formate production in comparison to the system with Cys at all timepoints measured. In the Solar Simulator, the difference was even higher, with the system using TEOA as a SED outperforming the system using Cys around 7-fold for formate production.

With these results, it is clear that TEOA is able to regenerate Eosin Y (Figure 20A) and TiO_2 (Figure 20B) more efficiently than Cys. For these reasons, TEOA was selected as the SED for both systems, and their optimization was further pursued.

Firstly, in an attempt to optimize the TiO_2 concentration to be used, the concentration of this component was varied (ranging from 0.44 to 2.175 mg/mL). It is known that the enzyme and TiO_2 interact strongly, therefore the ratio of nanoparticles for the enzyme used could influence highly the efficiency of the system.
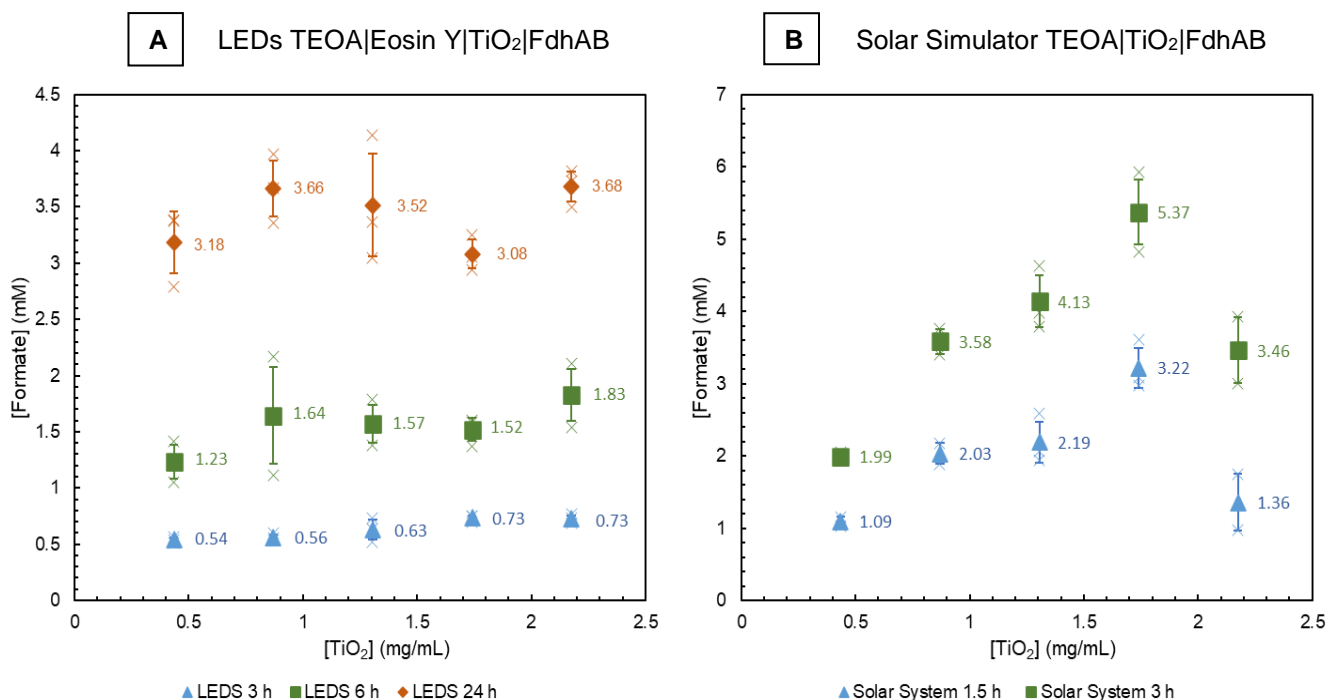
Figure 21: Formate production as a function of TiO_2 concentration used for the system with TEOA|Eosin Y|TiO2 nanoparticles|FdhAB using LEDs **(A)** and for the system with TEOA|TiO_2|FdhAB using the Solar Simulator **(B)**. At all timepoints, the formate concentration at time 0 was subtracted. Replicates are represented in crosses with the colour corresponding to the timepoint and the standard deviation is given by the error bars.

The two systems reacted differently to variations in TiO_2 concentration (Figure 21). In the case of the sensitized system, the changes in TiO_2 concentration did not result in significant changes in formate production. For instance, after 3 h, for the two highest concentrations used (1.74 and 2.18 mg/mL of TiO_2) 0.73 mM of formate was produced, and the lowest concentration used (0.44 mg/mL) achieved 0.54 mM of formate. Despite the variability between replicates for the 6 h and 24 h data being quite high, an increase in TiO_2 concentration does not seem to be justified. In the case of the Solar Simulator, there was a clear increase in formate production with the increase of TiO_2 up to 1.74 mg/mL (where a value of 5.37 mM was obtained after 3 h), with the highest TiO_2 concentration used (2.175 mg/mL) resulting in an abrupt drop in production (3.46 mM after 3 h). From these results, it is possible to conclude that TiO_2 is not the limiting factor in the system using sensitization, but it is limiting the system without sensitization, with the concentration of 1.74 mg/mL resulting in the highest production of formate after 1.5 h and 3 h (3.22 and 5.37 mM, respectively).

Similarly, we tried to evaluate the effect of using different concentrations of the dye (Eosin Y), while maintaining the original concentration of TiO_2 nanoparticles used (0.87 mg/mL). Furthermore, we wanted to analyse if the dye could directly supply electrons to FdhAB, so a system without TiO_2 nanoparticles was also tested (TEOA|Eosin Y|FdhAB). In both cases, Eosin Y concentration was varied from 0.22 mM to 0.88 mM to evaluate its effect on formate production (Figure 22).
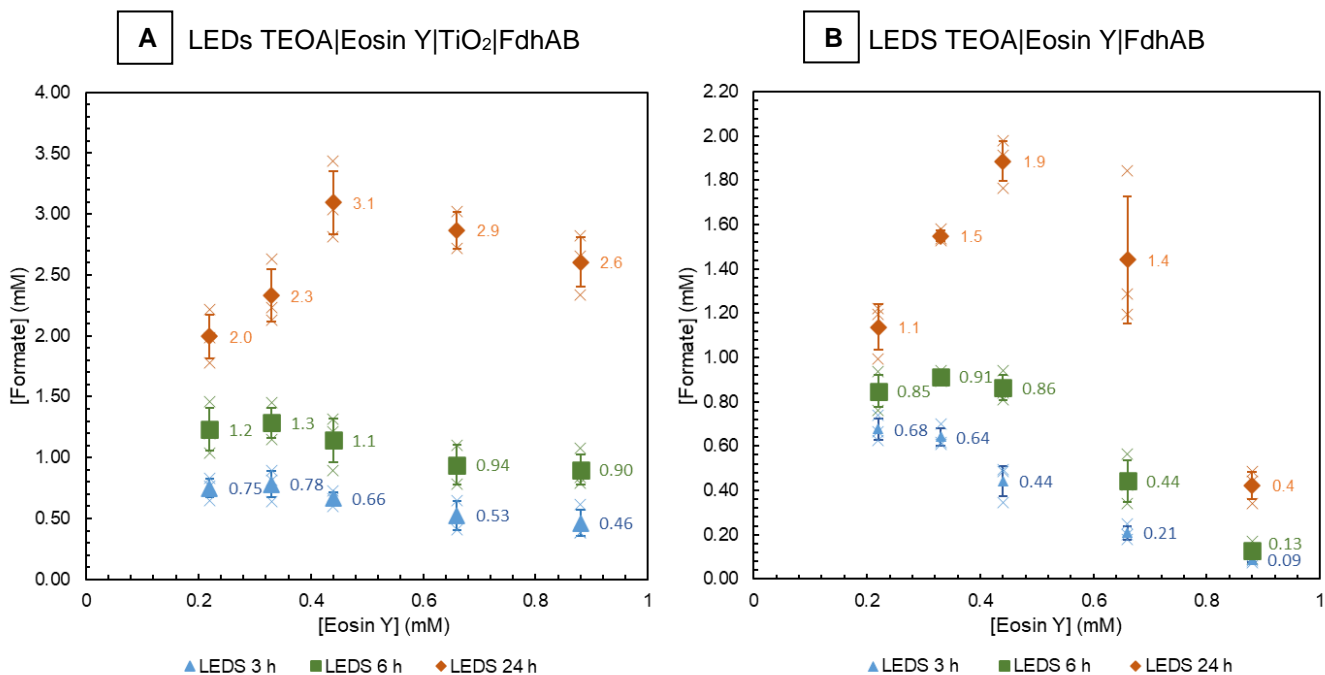
Figure 22: Formate production as a function of Eosin Y concentration for the system with TEOA|Eosin Y|TiO2 nanoparticles|FdhAB **(A)** and for the system with TEOA|Eosin Y|FdhAB **(B)**, both using LEDS (470 nm). At all timepoints, the formate concentration at time 0 was subtracted. Replicates are represented in crosses with the colour corresponding to the timepoint and the standard deviation is given by the error bars.

The first conclusion that can be taken from the results obtained is that it is possible to directly inject electrons to FdhAB from Eosin Y, which corresponds to a completely metal-free system. When comparing both systems it was possible to identify different trends depending on the timepoint analysed. For instance, at 3 h and 6 h, systems with lower concentrations of Eosin Y had slightly higher production of formate. The difference between lower and higher concentrations at these timepoints seemed to be more exacerbated in the system without TiO_2, but in terms of formate production in all concentrations, the sensitized system with TiO_2 outperformed the system without.

A different behaviour was observed for the 24 h data, where in both systems, the total amount of formate produced increased with the increase in Eosin Y concentrations up to 0.44 mM Eosin Y, decreasing afterwards. Interestingly enough, the decrease was more abrupt in the system without TiO_2 than in the system with TiO_2, which suggests that the metal somehow has a stabilizing effect in Eosin Y or in the enzyme. At 24 h, both systems with 0.44 mM of Eosin Y outperformed the remaining concentrations in the study.

Even though at 3 h and 6 h the use of 0.33 mM Eosin Y seemed to slightly outperform the use of 0.44 mM Eosin Y for both systems, at 24 h 0.44 mM Eosin Y showed a significant increase in formate production in comparison to 0.33 mM Eosin Y (3.1 and 2.3 mM in the system with TiO_2 and 1.9 and 1.5 mM in the system without TiO_2, respectively).

To validate the three systems in the study, several controls were performed using the optimized concentrations. In the systems using Eosin Y, with and without TiO_2 nanoparticles, controls without light, without TEOA, or without FdhAB were tested, with no formate production being obtained (results not shown), showing that the system only works with the presence of all the components.

In the case of the systems without sensitization, the controls mentioned above were performed as well, including and an additional control without TiO_2 nanoparticles, with the ones where formate production was identified being present in Table 6.

Table 6: Controls that showed formate production in the TEOA|TiO_2|FdhAB system using the Solar Simulator. For each timepoint formate concentration in mM is given with the error representing the standard deviation.

Controls Solar Simulator	[Formate] (mM)	
	1.5 h	3 h
Without TEOA	1.02 ± 0.13	1.83 ± 0.24
Without FdhAB	0.20 ± 0.00	0.23 ± 0.01

Regarding the TEOA|TiO_2|FdhAB system without TEOA in the Solar Simulator, a production of formate of 1.02 ± 0.13 mM and 1.83 ± 0.24 mM at 1.5 and 3 h, respectively, was observed. This production should correspond to the non-regenerated reaction, where the electrons TiO_2 injects into FdhAB to perform the reaction are not being supplied by TEOA. It is known that DTT can also act as a SED, therefore the limited amount of DTT present can be repairing the initial holes[92], but not enough to prevent recombination processes that lead to a decrease in the production of formate. In the case of the system without FdhAB, the formate production was quite low and can be due to TiO_2 being able to catalyse the reaction non-specifically[82], therefore yielding 0.20 ± 0.00 and 0.23 ± 0.01 mM of formate at 1.5 and 3 h, respectively.

The systems using LEDs have a low energy input and are only a single-wavelength light source, therefore we wanted to test its performance with a higher intensity of energy being supplied and in the full visible spectrum. For this, the Solar Simulator with a filter that attenuates IR light and some degree of UV light was tested. In the case of the LEDs, an intensity of 7 W/m^2 at a wavelength of 470 nm is being supplied, whilst the Solar Simulator without attenuation supplies the full visible spectrum with an intensity of 35 000 W/m^2, which is around 5000-fold higher. Unfortunately, it was not possible to measure the intensity of the light being supplied with attenuation. The results obtained in combination with the previous results obtained with the optimized conditions are summarized in Table 7. The graphical representation of the three systems can be found in Figure 23.

Table 7: Overall performance in terms of formate production (mM) and turnover (s^{-1}) of each system in study in different light set-ups. The results correspond to all the timepoints studied.

		TEOA|TiO2|FdhAB	TEOA|Eosin Y|TiO_2|FdhAB		TEOA|Eosin Y|FdhAB	
		Solar Simulator	Solar Simulator with Filter	LEDS 470 nm	Solar Simulator with Filter	LEDS 470 nm
1.5 h	[Formate] (mM)	3.2 ± 0.3	0.2 ± 0.0	-*	ND**	-*
	Turnover (s^{-1})	50 ± 4	2.4 ± 0.1	-*	ND**	-*
3 h	[Formate] (mM)	5.4 ± 0.5	0.3 ± 0.0	0.6 ± 0.1	ND**	0.44 ± 0.07
	Turnover (s^{-1})	42 ± 4	4.6 ± 0.0	5.1 ± 0.4	ND**	3.4 ± 0.5
6 h	[Formate] (mM)	-*	-*	1.1 ± 0.2	-*	0.87 ± 0.06
	Turnover (s^{-1})	-*	-*	4.4 ± 0.7	-*	3.3 ± 0.2
24 h	[Formate] (mM)	-*	-*	3.1 ± 0.3	-*	1.9 ± 0.1
	Turnover (s^{-1})	-*	-*	3.0 ± 0.2	-*	1.8 ± 0.1

*Not determined **Not detected

Considering the system of Eosin Y-sensitized TiO_2, it appears that the increase in light intensity was not beneficial and in the case of the system without TiO_2 no formate production was registered. The TEOA|Eosin Y|TiO_2|FdhAB was able to produce formate but at lower rates (0.3 ± 0.0 mM with a turnover of 4.6 ± 0.0 s^{-1} after 3 hours) in comparison with the LEDs (0.6 ± 0.1 mM with a turnover of 5.1 ± 0.4 s^{-1} after three hours). The lack of formate production in the sensitized system without TiO_2 versus formate production in the sensitized system with TiO_2 further supports the argument that TiO_2 somehow acts as a stabilizer to Eosin Y or to FdhAB. It would be interesting to test light intensities between 7 W/m^2 and 35 000 W/m^2 to see what the optimal intensity is, however, we only had the possibility to increase intensity above 35 000 W/m^2, something that these experimental results indicate that would not be beneficial.
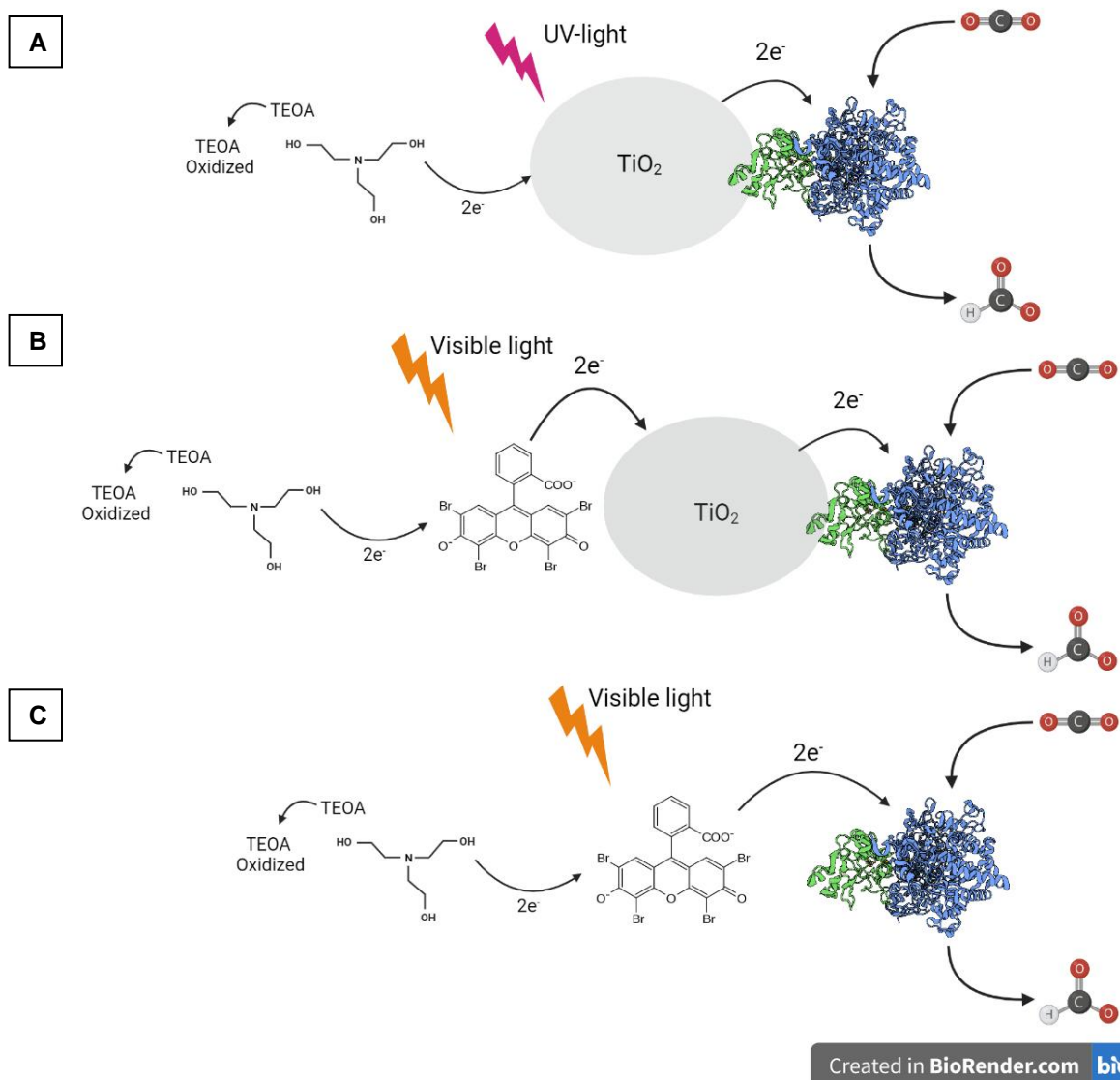
Figure 23: Photocatalytic setups for CO_2 reduction with the highly active *D. vulgaris* FdhAB. All systems used TEOA as a SED, with the excited material being the dye (Eosin Y) by visible light in systems **(B) and (C)** and TiO_2 by UV-light in system **(A)**. The electron is then injected from the excited material into the enzyme indirectly (through TiO_2 – system B) or directly (system (A) and (C)). Image was created with BioRender.com.

Overall, it is clear that the TEOA|TiO_2|FdhAB system in the Solar Simulator outperformed the remaining two systems, resulting in the production of 3.2 ± 0.3 mM of formate with a turnover of 50 ± 4

s^{-1} after 1.5 h and 5.4 ± 0.5 mM with a turnover of 42 ± 4 s^{-1} after 3 h. However, it is important to note that in this case, the full light spectrum and high light intensity are being used. In the case of the TiO$_2$-sensitized system in the LEDs, a formate production rate of 0.6 ± 0.1 mM with a turnover of 5.1 ± 0.4 s^{-1} and 3.1 ± 0.3 mM with a turnover of 3.0 ± 0.2 s^{-1} was attained after 3 and 24 h, respectively. Nonetheless, only a wavelength of 470 nm and a light intensity of 7 W/m^2 is used, which means that when taking into account system performance and light usage this system outperforms the one without sensitization. Furthermore, this system is even effective in a completely metal-free setup, although with lower production rates.

In summary, we were able to develop three different catalytic setups with commercially available components, with exception of the enzyme. These setups allow us to test FdhAB variants with different light wavelengths and intensities, with and without the usage of metal nanoparticles.

5.4. Computational Optimization of FdhA and FdhAB

FdhA, the simplified version of the protein appears to have stability issues. For this reason, the stabilization of the full α-subunit was studied using the PROSS algorithm[109]. Furthermore, natural proteins are only marginally stable[94], so it would be of high relevance to make FdhAB also more robust and stable than it already is. Furthermore, this could make FdhAB more tolerant to mutations in the active site, which could be further explored in order to improve this highly promising biocatalyst.

Even though PROSS is available in a webserver, for more control the protocol was applied with some alterations using the xml scripts that the authors made available[109]. Two major inputs are required, a structure of the protein, already available, and a PSSM matrix.

Therefore, for the stabilization of FdhA, the β-subunit with its three iron-sulphur clusters were removed from the FdhAB structure, whilst for FdhAB stabilization the full structure was used. The structure was renumbered being the number one position considered the first amino acid in the structure (therefore the first 34 amino acids in the α-subunit that correspond to the signal peptide are not being used). Furthermore, since the structure available did not identify the strep-tag and the last 22 amino acids of the β-subunit, these were also not used in the optimization. With this in mind, it is worth noting that the enumeration that is presented here for the α-subunit does not match the enumeration from the sequence. In the case of the β-subunit the number of the amino acid in the sequence matches the one in the structure. The structures were then prepared by following the Rosetta relaxation protocol.

Besides the structure, PROSS algorithms requires a position-specific scoring matrix (PSSM) that represents the log-likelihood of each of the 20 amino acids being in a given position. The PSSM is obtained by a Multiple Sequence Alignment (MSA) using PSI-BLAST, with CD-HIT being used to cluster the structures to remove redundant ones. A PSSM was obtained for each of the subunits since they are two different proteins. The PSSM is used as a first filter, where only mutations with a favourable score (≥ 0) are considered as acceptable, which has mainly two objectives: minimizing the risk of deleterious mutations being added and restricting the number of mutations tested by Rosetta (which has a high computational cost).

To avoid loss of activity, some residues had to be restricted and fixed (not allowing design nor repacking of the sidechains). In FdhA, this meant residues at 4 Å of the tungsten metal center, the MGDs binding site, the metal sulphide ligand and the iron-sulphur cluster. In FdhAB, the same restrictions were applied, with the restriction in the iron-sulphur cluster being extended to the ones in the β-subunit. The residues that were restricted and fixed can be found in Table 13 in the Annex Section.

For the filterscan a total of 8 different thresholds were used: $\Delta\Delta G < - 2.0$ (design 1); $\Delta\Delta G < - 1.8$ (design 2); $\Delta\Delta G < - 1.5$ (design 3); $\Delta\Delta G < - 1.25$ (design 4); $\Delta\Delta G < - 1.0$ (design 5); $\Delta\Delta G < - 0.75$ (design 6); $\Delta\Delta G < - 0.45$ (design 7); $\Delta\Delta G < + 0.5$ (design 8). The mutations applied in designs with lower energy are more conservative than in designs with higher energy, however the increased risk with increasing the threshold of energy also increases the stabilizing potential.

The mutations that pass both filters are then subjected to Rosetta combinatorial sequence design which will find the optimal combination of mutations. The sequence of FdhA and FdhAB aligned with their respective designs can be found in Figure 32 and Figure 33 in the Annex Section, respectively. To analyse the results obtained from the design step, Rosetta Features Reporter was used.

For FdhA, the number of mutations in each design, along with the mutations in the core of the protein, its overall energy score (given by the Rosetta energy function), and its isoelectric point can be found in Table 8. For FdhAB, the same information can be found in Table 9.

Table 8: FdhA and FdhA designs using PROSS algorithm. Energy values were obtained using Rosetta ref2015 score function with the values being represented in kcal.mol^{-1}. The isoelectric point for each protein was obtained using the ExPASy webserver[127].

Protein	Number of Amino acids	Mutations	Mutations in Core*	Energy (kcal.mol^{-1})	Isoelectric Point (pI)
FdhA		-	-	-2950	8.61
dFdhA1		124 (13%)	42 (34%)	-3195	8.65
dFdhA2		131 (14%)	42 (32%)	-3179	8.74
dFdhA3		147 (15%)	49 (33%)	-3204	8.70
dFdhA4	970	153 (16%)	51 (33%)	-3210	8.73
dFdhA5		156 (16%)	48 (31%)	-3208	8.61
dFdhA6		163 (17%)	50 (31%)	-3226	8.65
dFdhA7		174 (18%)	56 (32%)	-3237	8.84
dFdhA8		200 (21%)	62 (31%)	-3233	8.42

*In brackets, the % of mutations in the core are represented as a function of the total number of mutations.

Table 9: FdhAB and FdhAB designs using PROSS algorithm. Energy values were obtained using Rosetta ref2015 score function with the values being represented in kcal.mol^{-1}. The isoelectric point for each protein was obtained using the ExPASy webserver[127].

Protein	Number of Amino acids	Mutations	Mutations in Core*	Energy (kcal.mol^{-1})	Isoelectric Point (pI)
FdhAB		-	-	-3618	8.36
dFdhAB1		90 (8%)	22 (24%)	-3857	8.05
dFdhAB2		93 (8%)	22 (24%)	-3852	7.95
dFdhAB3		100 (8%)	26 (26%)	-3867	7.83
dFdhAB4	1184	105 (9%)	27 (26%)	-3871	7.95
dFdhAB5		119 (10%)	31 (26%)	-3875	8.03
dFdhAB6		127 (11%)	37 (29%)	-3877	8.11
dFdhAB7		143 (12%)	42 (29%)	-3892	8.30
dFdhAB8		172 (15%)	53 (31%)	-3888	7.93

*In brackets, the % of mutations in the core are represented as a function of the total number of mutations

As expected, when applying a more permissive energy threshold, the number of mutations that the algorithm incorporates in the design increases. When comparing both designs, it is noteworthy that the number of mutations added in FdhA is higher than the number of mutations added in FdhAB for a given threshold, which is curious since FdhAB is a bigger protein, therefore one would expect a higher number of mutations. On the other hand, FdhAB is more stable, since the β-subunit somehow stabilizes the large subunit. Furthermore, the restrictions applied in the α-subunit for FdhA and FdhAB designs had only slight alterations (Table 13) due to the fact that when the structures were relaxed, the presence of the β-subunit changes the environment of the α-subunit causing slight variations in the conformational

structure. In the case of FdhA, 75 residues were restricted from design, whilst for FdhAB in total 111 were restricted (70 in the α-subunit and 41 in the β-subunit). Remarkably, in FdhAB no mutations were added to the β-subunit in all designs, even in the most permissive one (dFdhAB8), as can be seen by the sequence alignment (Figure 33 in Annex Section). Therefore, all FdhAB designs only have mutations in the α-subunit. Even though FdhAB was optimized one subunit at a time, the presence of the β-subunit clearly influences the design that PROSS conducts in the α-subunit.

In the case of the FdhA designs, the increase in the number of mutations, leads to a linear increase in the incorporation of mutations in the protein core, with the % of core mutations as a function of the total of mutations being fairly constant from dFdhA1 to dFdhA8. In FdhAB, the % of core mutations does not vary proportionally with the increase in the number of mutations, with a higher number of mutations being incorporated in terms of % in the core of the protein as the design gets more permissive (31% mutations in the core in dFdhAB8, while only 24% in dFdhAB1).

For all designs, there is a decrease in energy in comparison to the original sequence. The profile is similar for the designs of both proteins, with the energy decreasing with the increase in the energetic threshold permissiveness. In the case of design 8, since the algorithm allows the incorporation of mutations with a ΔΔG between 0 and 0.5 kcal.mol^{-1}, there is a slight increase in energy. Nonetheless, in both cases design 7 is the one that has a lower energy, with an increase of around 10% in dFdhA7 in comparison to FdhA (-3237 kcal.mol^{-1} and -2950 kcal.mol^{-1}, respectively) and 8% in dFdhAB7 in comparison with FdhAB (-3892 kcal.mol^{-1} and 3618 kcal.mol^{-1}, respectively). The isoelectric point (pI) varies with no identifiable pattern for the designs, nonetheless it seems that the variations are more abrupt in the designs from FdhAB, with values ranging from 7.83 to 8.30 (with FdhAB having a value of 8.36) than in FdhA, that ranges from 8.42 to 8.84 (with FdhA having 8.61).

To better understand how each design is being stabilized and to understand the variations in pI it is important to analyse some common features, namely the difference in terms of stabilizing interactions, such as H-bonds, salt-bridges, mutations to proline that can stabilize protein backbone, as well as charge interactions. For this, Rosetta Features Reporter was used, and the results can be found in Table 10 and Table 11 for FdhA and FdhAB, respectively.

Table 10: Important structural features of the designed FdhA proteins in comparison to FdhA. The differences in the number of hydrogen bonds and salt-bridges, mutations to proline, negatively charged (D/E) or positively charged (K/R) residues are represented. Positive values represent more interactions in the designed proteins and negative values representing fewer interactions, versus FdhA.

Protein	H-Bonds	Salt Bridges	X <-> Pro	Negatively Charged	Positively Charged
dFdhA1	12	-11	3	0	1
dFdhA2	16	-6	3	-1	2
dFdhA3	11	-8	4	1	3
dFdhA4	15	7	4	4	8
dFdhA5	19	-5	3	4	6
dFdhA6	7	0	4	6	8
dFdhA7	9	+8	4	3	10
dFdhA8	1	-6	4	8	6

Table 11: Important structural features of the designed FdhAB proteins in comparison to FdhAB. The differences in the number of hydrogen bonds and salt-bridges, mutations to proline, negatively charged (D/E) or positively charged (K/R) residues are represented. Positive values represent more interactions in the designed proteins and negative values representing fewer interactions, versus FdhAB.

Protein	H-Bonds	Salt Bridges	X <-> Pro	Negatively Charged	Positively Charged
dFdhAB1	10	21	-4	10	6
dFdhAB2	10	20	-4	9	4
dFdhAB3	3	29	-4	11	5
dFdhAB4	4	22	-3	11	6
dFdhAB5	7	29	-2	11	7
dFdhAB6	5	28	0	11	8
dFdhAB7	12	18	1	9	9
dFdhAB8	7	12	0	11	6

Interestingly, in FdhA each design optimizes different types of interactions. While for the more conservative thresholds (from 1 to 5) the algorithm focuses on increasing H-bonds, for the more permissive designs, it focuses more on the electrostatic interactions, with mutations to negatively charged or positively charged amino acids (from design 4 to 8). In all cases, the mutations from any amino acid to proline remain fairly constant, with the algorithm attempting to increase backbone rigidity. Furthermore, salt-bridges vary with no recognizable pattern, but dFdhA7 which by Rosetta energy calculations appears to be the most stable has a significant increase in salt-bridges (8). In dFdhA7, the optimization appears to be more general instead of focusing on a given interaction, with an overall increase in H-bonds (9), backbone rigidity (4 mutations to proline), and electrostatic interactions (3 and 10 mutations to negatively and positively charged amino acids, respectively).

Surprisingly, in FdhAB design the algorithm increases in all models the number of salt-bridges, from 12 to 29 versus FdhAB. There is also a great focus on increasing negatively charged and positively charged amino acids. Furthermore, in the most conservative designs (from dFdhAB1 to dFdhAB5) there is a decrease in prolines, which should contribute to increase the flexibility in certain areas of the protein. On the contrary, the most permissive designs did not change this parameter (dFdhAB6 and dFdhAB8), with dFdhAB7 being an exception with the increase of 1 proline.

Noteworthy, the presence of the β-subunit alters the energy of the protein, therefore it changes how the algorithm designs the protein, with more focus being given to salt-bridges and charged amino acids. When comparing the two best designs in terms of energy, dFdhA7 and dFdhAB7 (Figure 24 and Figure 25, respectively and Table 14 in the Annex Section), all types of interactions studied seem to be optimized, possibly explaining why overall these designs seem to have the lower energy, being the most stable.
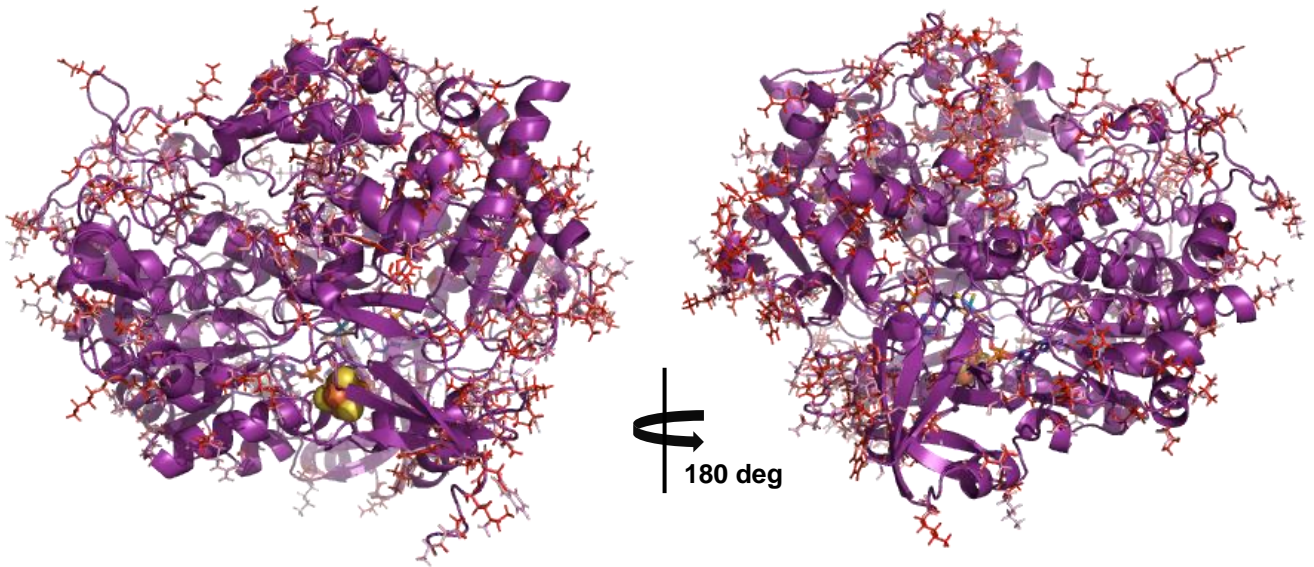
Figure 24: Structural representation of the relaxed structures of the α-subunit of FdhA in purple cartoon. In pink lines, we have the positions mutated in the FdhA protein in dFdhA7, with the designed amino acids in red lines. Images were created with PYMOL.
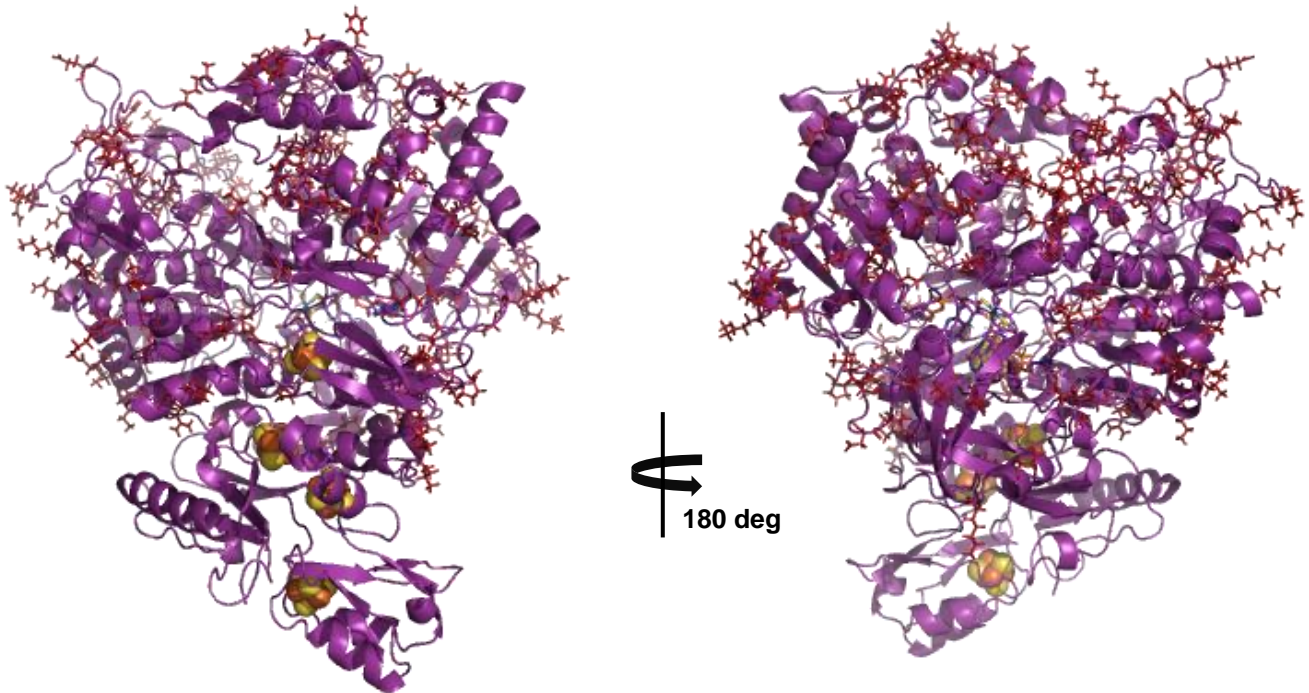
Figure 25: Structural representation of the relaxed structures of FdhAB in purple cartoon. In pink lines, we have the positions mutated in the FdhAB protein in dFdhAB7, with the designed amino acids in red lines. Images were created with PYMOL.

In summary, the PROSS algorithm was able to design several optimized models for FdhA and FdhAB, with more stability by computational calculations, which can be experimentally tested to create a stable version of the FdhA simplified version of the enzyme, and also to increase even further FdhAB stability.

6. Discussion

As humanity walks towards the future our consciousness of the impact of CO_2 emissions and the need to develop cleaner alternatives to power our society is rising. Climate change is proving to us day-by-day that change is urgently required. CO_2 capture and utilization associated with hydrogen generation is a promising couple that could become a relevant strategy for solving this challenge. Formate can be the bridge between this couple, with CO_2 reduction generating formate, which is one of the most promising hydrogen-storage materials, addressing several safety concerns associated with its transportation and usage. Formate dehydrogenases (FDHs) are one of the most efficient ways of performing the two electron-reduction of CO_2 to formate. Therefore, understanding the way FDHs perform CO_2 reduction can lead us to design better catalysts and get us closer to a future application.

The FdhAB from *D. vulgaris* Hildenborough offers high activity and great overall stability, being among the most promising enzymes for this task. Therefore, developing a simplified version that allows a more detailed study of the catalytic subunit can help us elucidate the mechanism.

In this work, the generation of a simplified version comprising only the alpha subunit was successfully performed. Furthermore, the simplified version (FdhA) was compared to FdhAB to evaluate its stability and kinetic capabilities. The FdhAB purified in this work is a highly active enzyme, capable of catalysing formate oxidation and CO_2 reduction at a rate of 961 ± 84 s^{-1} and of 234 ± 13 s^{-1}, respectively. These values are similar to the ones previously reported for this enzyme (1310 ± 50 s^{-1} and 315 ± 28 s^{-1} for formate oxidation and CO_2 reduction, respectively)[39]. However, the simplified version without the β-subunit (FdhA), even though able to catalyse both reactions, has much lower turnovers, with decreases of around 200-fold for both formate oxidation and CO_2 reduction. This shows that even though, in theory, the catalytic subunit is the only one required for catalysis, the presence of the β-subunit makes a world of a difference in terms of catalytic efficiency, showing that the β-subunit is more than a simple electron-wire. These results are similar to the ones obtained with the [NiFe] hydrogenase from *Ralstonia eutropha*, which has a catalytic subunit and an electron transfer subunit with three iron-sulphur clusters in a similar configuration as FdhAB. There the authors removed the small electron wire subunit, leading to a 200-fold decrease in H/D exchange activity comparing to the wild-type enzyme[117]. The authors suggested that this occurred due to a decrease in affinity for H_2 and a higher competition of the metal center for H_2O due to the lack of the confining and hydrophobic environment caused by the small subunit[117]. In FdhAB, no competition occurs between solvent and substrate for the active site and, even though the affinity of the substrate for the active center was not evaluated in the simplified version of FdhA, this would not explain everything. In our case, the abrupt decrease in activity is accompanied by protein expression being impaired (seen by Western-blot and SDS-gel), something that in the [NiFe] hydrogenase does not occur. It is worth noting that expression is being defined as the whole process needed to obtain a mature protein, with the translocation of the protein to the periplasm being encompassed in our definition.

There is a high number of factors that can contribute to reduced protein expression, but the most common is a protein stability issue. The stability issue hypothesis is further supported by circular dichroism data and by a thermal shift assay. Surprisingly, CD results show a complete abolishment of antiparallel β-strands in FdhA, which are a common secondary structure found in FdhAB (above 20%

predominance). Furthermore, a decrease from 78.9 ± 0.1 °C to 66.3 ± 1.0 °C is seen without the β-subunit, which indicates a notorious change in protein stability.

Little is known about the maturation process of metal-dependent FDHs. In the case of FdhAB, it is known that the α-subunit interacts with the β-subunit in the cytoplasm because only the α-subunit has the signal peptide required for the protein to be translocated to the periplasm where it performs its physiologic function. This signal peptide is common in FDHs that are translocated across the membrane, such as FDH-N from *E. coli* and FdhAB where the 33 residues from the N-terminal are missing in the mature protein and in crystallographic structures. This signal peptide contains a twin-arginine that works as a translocation signal (for the twin-arginine translocation pathway – TAT pathway)[132,133]. Furthermore, the TAT pathway exports fully folded proteins from the cytoplasm to the periplasm, having a folding quality control intrinsic to the process[132,134]. It is known that the maturation of the FDH from *R. capsulatus* occurs through a concerted process, where first subunit assembly occurs and only after the cofactors and the iron-sulphur clusters are inserted through the help of chaperones, followed by a final folding step after insertion[66,135]. With this in mind, the low expression can be explained by two main hypotheses: (i) the β-subunit could be required for α-stabilization and without the presence of the β-subunit proper folding is not acquired and the protein is not translocated to the periplasm due to problems with the folding quality control mechanism and is marked for degradation, explaining the low yields or (ii) proper insertion of the cofactors might be required prior to the formation of the dimer due to the need of having a "pre-folded" protein that has the cofactor binding pocket, which would explain the low activity seen. It would be of high interest to perform metal quantification to analyse whether cofactors are being fully loaded, however, the low protein yields hampered this quantification up to now. In the case of the [NiFe] hydrogenase, the authors were able to obtain a fully loaded protein, confirmed by metal content determination, UV-vis, and IR spectroscopy, however, they did not perform any structural characterization that could indicate if the protein is properly folded[117]. Besides that, the size of the active center is not comparable to the one from FdhAB that has two bulky MGDs bound to the W atom, and it is known that the insertion of the metal only happens if it is assembled with the cofactors[66,135].

This work further confirms the importance of the β-subunit, probably having a structural function besides electron transfer, which is consistent with what was seen in the work with the [NiFe] hydrogenase[117]. Nonetheless, the authors of that study suggested that the small subunit is responsible for creating the hydrophobic environment needed for H_2 affinity and for activating the catalytic center in a redox-dependent manner[117]. In our work, the β-subunit appears to be of structural relevance for proper folding of the catalytic subunit and for efficient catalysis, with its absence leading to lower yields of protein, something that was not seen for the hydrogenase.

Photocatalysis aims to utilize the endless resource that is light as a source of energy to drive catalytic reactions, being of high potential for the area of sustainability. Several studies have already been performed using FDHs for CO_2 reduction powered by light, including with FdhAB[92]. Since this highly promising field of research can be applied for CO_2 reduction, it is important to understand how a mutation in FdhAB can affect its performance in a future application setting. Due to the unfortunate loss of expression in FdhA, the photocatalytic approach was only explored using FdhAB. For this, a system using TEOA|TiO_2|FdhAB was developed leading to high productions of formate (3.2 ± 0.3 mM) and high

turnover (50 ± 4 s^{-1}) being obtained after 1.5 h. As previously reported, TiO$_2$ interacts strongly with FdhAB, with more than 60% of FdhAB remaining absorbed to the nanoparticle even after washing with 3 M of KCl, possibly due to the interaction between TiO$_2$ and negatively charged amino acids at the surface of FdhAB[92]. However, this system suffers from the disadvantage that TiO$_2$ is only able to absorb UV radiation, with the high production rates and turnover being obtained with the supply of high intensity of light (7000 W/m^2 from 320-450 nm). Therefore, we successfully sensitized TiO$_2$ with Eosin Y obtaining 0.6 ± 0.1 mM of formate with a turnover of 5.1 ± 0.4 s^{-1} with the energy used being supplied at a single wavelength in the visible region, 470 nm, and in a much lower intensity (7 W/m^2), which clearly makes the process more energetically efficient. Furthermore, in comparison to the previously described system with TEOA|DPP|TiO$_2$|FdhAB where a turnover of 5 ± 0.6 s^{-1} was obtained with 100 W/m^2 light-intensity above 420 nm[92], a similar turnover was obtained using Eosin Y, which is commercially available and with a much smaller input of energy. Even further, Eosin Y has the advantage of being able to form multilayer structures by complexation with metal ions with 3 different coordination sites[136]. A group successfully constructed a multilayer Eosin Y structure with Fe^{3+} coupled to TiO$_2$ for H$_2$ evolution with increases of more than 3-fold in production when equimolar concentrations of Eosin Y and Fe^{3+} are used, with this effect attributed to Fe^{3+} being capable of enhancing dye absorption to TiO$_2$ and suppressing electron back transfer[136].

Eosin Y was even able to transfer electrons directly to FdhAB in a completely metal-free version of the photocatalytic setup, even though at lower rates (0.44 ± 0.07 mM of formate and a turnover of 3.4 ± 0.5 s^{-1}), showing the versatility of Eosin Y. The two systems were also tested in the Solar Simulator that provides higher energy, and covers the full visible light spectrum to analyse if the light was limiting the process, since we were using a single wavelength which is not optimal to photoexcite Eosin Y. However, no production was seen in the completely metal-free setup and a decrease in production occurred in the system where the TiO$_2$ nanoparticles were used, possibly due to dye degradation or photobleaching since we were not able to limit the UV radiation emitted by the Solar Simulator, even though these processes typically occur under aerobic conditions due to the formation of reactive oxygen species (ROS)[137–139].

Initially, our goal was to test the simplified version of the enzyme generated, FdhA, in the photocatalytic system. However, the low expression and low activity, coupled with the results obtained from the CD spectroscopy and from the thermal shift assay, lead us to believe that the α-subunit is not stable alone. Due to the abrupt effect that the removal of the β-subunit caused in the stability of the α-subunit, the stabilization of the full α-subunit was attempted using PROSS algorithm[109]. This is considered a hybrid method since it improves stability by using phylogeny and energy considerations, being the algorithm especially applicable to alleviating problems of low expression levels and misfolding. Furthermore, it has shown wide success in several studies using different types of proteins[109,114,140]. In parallel, the protocol was also employed to FdhAB since natural proteins often suffer from marginal stability[94].

For each protein, 8 optimized models for each protein were obtained. In both cases, due to the high number of mutations considered even in more strict thresholds, the most permissive threshold (model 8) can be riskier to implement experimentally due to the addition of possible false positives

(mutations that PROSS identified as stabilizing, even though experimentally are destabilizing). It would only make sense to consider the implementation of model 8 if the number of mutations obtained with the stricter designs were quite low (< 6 to 8 %). In the case of FdhAB, fewer mutations are added in comparison with FdhA, which was unexpected since even though FdhAB is more stable (both experimentally and by computational energy scoring), it is a larger protein, so one would expect that more mutations would be added, especially in the more permissive threshold. The lack of mutations in the β-subunit, possibly due to large restrictions being applied near the three iron-sulphur clusters, coupled to the change in the environment of the α-subunit due to the presence of the β-subunit, led to a significant change in the number of mutations for a given threshold.

Even though the authors of PROSS suggest that model 7, using a threshold of -0.45 is in most cases the ideal one, due to the low risk of false positives[109], the high number of mutations being incorporated in the design variants with this threshold adds some uncertainty regarding the final outcome. In the case of FdhA, the protein is suspected to be unstable, therefore it is expected that a higher number of alterations are required for stabilization. In the case of FdhAB, we are dealing with a protein that is shown to be operationally stable, with a high T_M as seen in this study and previously reported by our lab[39]. Nonetheless, being able to increase the yields of protein obtained and increase stability to implement a functional design protocol for improved activity[115] would be highly beneficial.

With this in mind, along with the models dFdhA7 and dFdhAB7 it would be of great interest to test experimentally two additional models to reduce the risk of the high number of mutations proposed in these models. For instance, models dFdhA1 and dFdhAB1, which are the stricter ones generated, and models dFdhA4 and dFdhAB4 would be interesting to test as well. It is hard to predict how each model will behave experimentally, however, PROSS has been proven to work, even in far worse settings than this one[114,140]. For instance, in the case of *Plasmodium falciparum* reticulocyte-binding protein homolog 5 (PfRH5), which is a malaria vaccine candidate, the number of homologous sequences available was quite low, therefore the search was extended below the so-called "twilight zone" of sequence identity (< 25%) in order to obtain a MSA to generate the PSSM. The pPfRH5 had limited stability at high temperatures and was not produced correctly in microbial expression hosts, nonetheless, the designed model obtain with PROSS had increased expression in the current leading system for the expression of this protein (*Drosophila melanogaster* Schneider 2 cell line) and was successfully produced in *E. coli* as well. Furthermore, the designed model had an increase of 10 to 20 ˚C in thermal stability versus the wild-type protein[114]. In a different study, an attempt to increase the stability of a cytokine (IL-24), which is believed to be important for immunotherapy, was done by using homology models instead of an actual experimental 3D structure, and PROSS still presented a viable model with increased stability[140]. Therefore, even with the two crucial inputs required being far from ideal, PROSS was still able to optimize the protein in study, increasing its stability.

7. Concluding Remarks and Future Perspectives

In this work, a simplified version of FdhAB, with only the large subunit and the minimal number of centers crucial for catalysis, was successfully generated even though it presents low stability. Unfortunately, the low yields of protein did not allow extensive characterization of this intriguing protein. However, it was possible to conclude that the α-subunit without the presence of the β-subunit is much less stable and cannot perform efficient catalysis. Understanding how the β-subunit affects and stabilizes the catalytic subunit can help to obtain more insights about FdhAB. Furthermore, if stabilization of the simplified FdhA is successfully achieved experimentally with the models obtained by PROSS, this would help us study with more in more detail the mechanisms behind CO_2 reduction. For example, this would allow us to study the potential of the metal center and the proximal sulphur-cluster without the presence of the remaining three iron-sulphur clusters, which end up blurring their signal. Furthermore, if the loss of activity is only due to the loss of stability associated with the absence of the β-subunit, there is hope that the activity lost in the simplified version could be restored. In case of success creating the models suggested by PROSS, it is possible that producing the designed protein heterologously using *E. coli* becomes more feasible, nonetheless for this to occur the problems in terms of cofactor insertion in the protein in *E. coli* need to be addressed as well. This is because typically PROSS is employed to optimize the protein stability for expression in *E. coli*, with the models generated often achieving high expression using this expression system. This would be of high interest since *E. coli* is easier to manipulate and a lot of industry knowledge already exists, something that is not true for *D. vulgaris*. Moreover, the generation of stable variants of either FdhA or FdhAB could drive forward the implementation of computational protocols that aim to improve functionality, which would be of high relevance for improving the catalytic performance of the enzyme and to make this enzyme a tool for fighting climate change and the energetic transition required.

Furthermore, the photocatalytic systems were successfully developed, which allows us to test protein variants in a more practical setting, coupled to highly used TiO_2 nanoparticles. Additionally, this work shows the real application of FDH powering CO_2 reduction with an endless source of energy, solar light. A system that is also productive using less intense sources of light at a single wavelength (470 nm), to power formate production.

This work can be an important stepstone in unveiling several features of FdhAB, without ever losing sight of the ultimate goal, developing a state-of-the-art system for CO_2 reduction for wide implementation in the future.

8. References

1. Dupuis, M. *et al.* Frontiers, Opportunities, and Challenges in Biochemical and Chemical Catalysis of CO_2 Fixation. *Chem. Rev.* **113**, 6621–6658 (2013).

2. Maia, L. B., Moura, I. & Moura, J. J. G. Molybdenum and tungsten-containing formate dehydrogenases: Aiming to inspire a catalyst for carbon dioxide utilization. *Inorganica Chim. Acta* **455**, 350–363 (2017).

3. IEA. Data and Statistics in CO_2 Emissions. https://www.iea.org/data-and-statistics/.

4. NOAA. Trends in Atmospheric Carbon Dioxide. http://www.esrl.noaa.gov/gmd/ccgg/trends/.

5. Masson-Delmotte, V., P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. S., A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J.B.R. Matthews, Y. Chen, X. Zhou, M. I. G. & E. Lonnoy, T. Maycock, M. Tignor, and T. W. *Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change,. Summary for PolicyMakers In: Global Warming of 1,5 °C* (2018).

6. Sanz-Pérez, E. S., Murdock, C. R., Didas, S. A. & Jones, C. W. Direct Capture of CO_2 from Ambient Air. *Chem. Rev.* **116**, 11840–11876 (2016).

7. United Nations. United Nations Framework Convention on Climate Change. 1–33 (1992).

8. United Nations. Kyoto Protocol to the United Nations Framework Convention on Climate Change. 1–24 (1997).

9. United Nations. Paris Agreement to the United Nations Framework Convention on Climate Change. (2015).

10. United Nations. *Emissions Gap Report 2019. Emissions Gap Report 2019* (2019).

11. European Comission. The European Green Deal. (2019).

12. European Comission. REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL establishing the framework for achieving climate neutrality and amending Regulation (EU) 2018/1999 (European Climate Law). (2020).

13. Metz, B. & O. Davidson, H. C. de Coninck, M. Loos, and L. A. M. *The IPCC special report on carbon dioxide capture and storage. IPCC Special Report on Carbon Dioxide Capture and Storage* (2005).

14. Sekera, J. & Lichtenberger, A. Assessing Carbon Capture: Public Policy, Science, and Societal Need. *Biophys. Econ. Sustain.* **5**, 1–28 (2020).

15. U.S. Department of Energy, O. of B. E. S. W. Basic research needs: Catalysis for Energy. (2007).

16. Voldsund, M., Jordal, K. & Anantharaman, R. Hydrogen production with CO_2 capture. *Int. J. Hydrogen Energy* **41**, 4969–4992 (2016).

17. Enthaler, S., Von Langermann, J. & Schmidt, T. Carbon dioxide and formic acid - The couple for environmental-friendly hydrogen storage? *Energy Environ. Sci.* **3**, 1207–1217 (2010).

18. Fasihi, M., Efimova, O. & Breyer, C. Techno-economic assessment of CO_2 direct air capture plants. *J. Clean. Prod.* **224**, 957–980 (2019).

19. Tanaka, R., Yamashita, M. & Nozaki, K. Catalytic hydrogenation of carbon dioxide using Ir(III)-pincer complexes. *J. Am. Chem. Soc.* **131**, 14168–14169 (2009).

20. Brilman, W. *CO$_2$ removal from air. Advances in Carbon Capture* (Elsevier Inc., 2020).

21. Koytsoumpa, E. I., Bergins, C. & Kakaras, E. The CO$_2$ economy: Review of CO$_2$ capture and reuse technologies. *J. Supercrit. Fluids* **132**, 3–16 (2018).

22. Huang, C. H. & Tan, C. S. A review: CO$_2$ utilization. *Aerosol Air Qual. Res.* **14**, 480–499 (2014).

23. Aresta, M., Dibenedetto, A. & Angelini, A. The changing paradigm in CO$_2$ utilization. *J. CO$_2$ Util.* **3**–**4**, 65–73 (2013).

24. Wang, W.-H., Feng, X. & Bao, M. *Transformation of CO$_2$ to Formic Acid or Formate Over Heterogeneous Catalysts*. (2018).

25. Sordakis, K. *et al.* Homogeneous Catalysis for Sustainable Hydrogen Storage in Formic Acid and Alcohols. *Chem. Rev.* **118**, 372–433 (2018).

26. Hull, J. F. *et al.* Reversible hydrogen storage using CO$_2$ and a proton-switchable iridium catalyst in aqueous media under mild temperatures and pressures. *Nat. Chem.* **4**, 383–388 (2012).

27. Filonenko, G. A., Van Putten, R., Schulpen, E. N., Hensen, E. J. M. & Pidko, E. A. Highly efficient reversible hydrogenation of carbon dioxide to formates using a ruthenium PNP-pincer catalyst. *ChemCatChem* **6**, 1526–1530 (2014).

28. Kanega, R. *et al.* CO$_2$ Hydrogenation and Formic Acid Dehydrogenation Using Ir Catalysts with Amide-Based Ligands. *Organometallics* **39**, 1519–1531 (2020).

29. Inoue, Y., Izumida, H., Sasaki, Y. & Hashimoto, H. Catalytic Fixation of Carbon Dioxide To Formic Acid By Transition-Metal Complexes Under Mild Conditions. *Chem. Lett.* **5**, 863–864 (1976).

30. Schwarz, F. M., Schuchmann, K. & Müller, V. Hydrogenation of CO$_2$ at ambient pressure catalyzed by a highly active thermostable biocatalyst. *Biotechnol. Biofuels* **11**, 1–11 (2018).

31. Shi, J. *et al.* Enzymatic conversion of carbon dioxide. *Chem. Soc. Rev.* **44**, 5981–6000 (2015).

32. Alissandratos, A. & Easton, C. J. Biocatalysis for the application of CO$_2$ as a chemical feedstock. *Beilstein J. Org. Chem.* **11**, 2370–2387 (2015).

33. Maia, L. B., Moura, J. J. G. & Moura, I. Molybdenum and tungsten-dependent formate dehydrogenases. *J. Biol. Inorg. Chem.* **20**, 287–309 (2015).

34. Popov, Vladimir; Lamzin, V. NAD$^+$-dependent formate dehydrogenase. *Biochem. J.* 625–643 (1994).

35. Vinals, C., Depiereux, E. & Feytmans, E. Prediction of structurally conserved regions of D-specific hydroxy acid dehydrogenases by multiple alignment with formate dehydrogenase. *Biochemical and Biophysical Research Communications* vol. 192 182–188 (1993).

36. Matelska, D. *et al.* Classification, substrate specificity and structural features of D-2-hydroxyacid dehydrogenases: 2HADH knowledgebase. *BMC Evol. Biol.* **18**, 1–23 (2018).

37. De Bok, F. A. M. *et al.* Two W-containing formate dehydrogenases (CO$_2$-reductases) involved in syntrophic propionate oxidation by *Syntrophobacter fumaroxidans*. *Eur. J. Biochem.* **270**, 2476–2485 (2003).

38. Reda, T., Plugge, C. M., Abram, N. J. & Hirst, J. Reversible interconversion of carbon dioxide and formate by an electroactive enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 10654–10658 (2008).

39. Oliveira, A. R. *et al.* Toward the Mechanistic Understanding of Enzymatic CO$_2$ Reduction. *ACS*

Catal. **10**, 3844–3856 (2020).

40. Maia, L. B., Fonseca, L., Moura, I. & Moura, J. J. G. Reduction of Carbon Dioxide by a Molybdenum-Containing Formate Dehydrogenase: A Kinetic and Mechanistic Study. *J. Am. Chem. Soc.* **138**, 8834–8846 (2016).

41. Schuchmann, K. & Müller, V. Direct and reversible hydrogenation of CO_2 to formate by a bacterial carbon dioxide reductase. *Science (80).* **342**, 1382–1385 (2013).

42. Gonzalez, P. J. *et al.* Periplasmic nitrate reductases and formate dehydrogenases: Biological control of the chemical properties of Mo and W for fine tuning of reactivity, substrate specificity and metabolic role. *Coord. Chem. Rev.* **257**, 315–331 (2013).

43. Maia, L. B., Moura, I. & Moura, J. J. G. Carbon Dioxide Utilisation - The Formate Route. in *Enzymes for Solving Humankind's Problems: Natural and Artificial Systems in Health, Agriculture, Environment and Energy* (ed. Moura, José J.G., Moura, Isabel, Maia, L. B.) (Springer International Publishing, 2020).

44. Da Silva, S. M., Pimentel, C., Valente, F. M. A., Rodrigues-Pousada, C. & Pereira, I. A. C. Tungsten and molybdenum regulation of formate dehydrogenase expression in Desulfovibrio vulgaris Hildenborough. *J. Bacteriol.* **193**, 2909–2916 (2011).

45. Jormakka, M., Törnroth, S., Byrne, B. & Iwata, S. Molecular basis of proton motive force generation: Structure of formate dehydrogenase-N. *Science (80).* **295**, 1863–1868 (2002).

46. Da Silva, S. M., Pacheco, I. & Pereira, I. A. C. Electron transfer between periplasmic formate dehydrogenase and cytochromes C in *Desulfovibrio desulfuricans* ATCC 27774. *J. Biol. Inorg. Chem.* **17**, 831–838 (2012).

47. Ragsdale, Stephen W., Pierce, E. Acetogenesis and the Wood-Ljungdahl Pathway for CO_2 Fixation. *Biochim. Biophys. Acta* 1873–1898 (2008).

48. Worm, P., Stams, A. J. M., Cheng, X. & Plugge, C. M. Growth- and substrate-dependent transcription of formate dehydrogenase and hydrogenase coding genes in *Syntrophobacter fumaroxidans* and *Methanospirillum hungatei*. *Microbiology* **157**, 280–289 (2011).

49. Raaijmakers, H. C. A. & Romão, M. J. Formate-reduced *E. coli* formate dehydrogenase H: The reinterpretation of the crystal structure suggests a new reaction mechanism. *J. Biol. Inorg. Chem.* **11**, 849–854 (2006).

50. Radon, C. *et al.* Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in *Rhodobacter capsulatus* formate dehydrogenase. *Nat. Commun.* **11**, 1–9 (2020).

51. Young, T. *et al.* Crystallographic and kinetic analyses of the FdsBG subcomplex of the cytosolic formate dehydrogenase FdsABG from *Cupriavidus necator*. *J. Biol. Chem.* **295**, 6570–6585 (2020).

52. Khangulov, S. V., Gladyshev, V. N., Charles Dismukes, G. & Stadtman, T. C. Selenium-containing formate dehydrogenase H from *Escherichia coli*: A molybdopterin enzyme that catalyses formate oxidation without oxygen transfer. *Biochemistry* **37**, 3518–3528 (1998).

53. Raaijmakers, H. *et al.* Gene sequence and the 1.8 Å crystal structure of the tungsten-containing formate dehydrogenase from *Desulfovibrio gigas*. *Structure* **10**, 1261–1272 (2002).

54. Laukel, M., Chistoserdova, L., Lidstrom, M. E. & Vorholt, J. A. The tungsten-containing formate

dehydrogenase from *Methylobacterium extorquens* AM1: Purification and properties. *Eur. J. Biochem.* **270**, 325–333 (2003).

55. Boyington, J. C., Gladyshev, V. N., Khangulov, S. V., Stadtman, T. C. & Sun, P. D. Crystal structure of formate dehydrogenase H: Catalysis involving Mo, molybdopterin, selenocysteine, and an Fe4S4 cluster. *Science (80).* **275**, 1305–1308 (1997).

56. Thomé, R. *et al.* A sulfurtransferase is essential for activity of formate dehydrogenases in *Escherichia coli. J. Biol. Chem.* **287**, 4671–4678 (2012).

57. Arnoux, P. *et al.* Sulphur shuttling across a chaperone during molybdenum cofactor maturation. *Nat. Commun.* **6**, 1–8 (2015).

58. Schrapers, P. *et al.* Sulfido and cysteine ligation changes at the molybdenum cofactor during substrate conversion by formate dehydrogenase (FDH) from *Rhodobacter capsulatus. Inorg. Chem.* **54**, 3260–3271 (2015).

59. Axley, M. J., Bockt, A. & Stadtman, T. C. Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium (selenocysteine/selenoenzymes/TGA codon/kinetics). *Proc. Nati. Acad. Sci. USA* **88**, 8450–8454 (1991).

60. Nielsen, C. F., Lange, L. & Meyer, A. S. Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO_2 utilization. *Biotechnol. Adv.* **37**, 107408 (2019).

61. Grimaldi, S., Schoepp-Cothenet, B., Ceccaldi, P., Guigliarelli, B. & Magalon, A. The prokaryotic Mo/W-bisPGD enzymes family: A catalytic workhorse in bioenergetic. *Biochim. Biophys. Acta - Bioenerg.* **1827**, 1048–1085 (2013).

62. Niks, D. & Hille, R. Molybdenum- and tungsten-containing formate dehydrogenases and formylmethanofuran dehydrogenases: Structure, mechanism, and cofactor insertion. *Protein Sci.* **28**, 111–122 (2019).

63. Axley, M. J., Böck, A. & Stadtman, T. C. Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8450–8454 (1991).

64. Gladyshev, V. N. *et al.* Characterization of crystalline formate dehydrogenase H from *Escherichia coli*: Stabilization, EPR spectroscopy, and preliminary crystallographic analysis. *J. Biol. Chem.* **271**, 8095–8100 (1996).

65. Leopoldini, M., Chiodo, S. G., Toscano, M. & Russo, N. Reaction mechanism of molybdoenzyme formate dehydrogenase. *Chem. - A Eur. J.* **14**, 8674–8681 (2008).

66. Hartmann, T., Schwanhold, N. & Leimkühler, S. Assembly and catalysis of molybdenum or tungsten-containing formate dehydrogenases from bacteria. *Biochim. Biophys. Acta - Proteins Proteomics* **1854**, 1090–1100 (2015).

67. Tiberti, M., Papaleo, E., Russo, N., De Gioia, L. & Zampella, G. Evidence for the formation of a Mo-H intermediate in the catalytic cycle of formate dehydrogenase. *Inorg. Chem.* **51**, 8331–8339 (2012).

68. Mota, C. S. *et al.* The mechanism of formate oxidation by metal-dependent formate dehydrogenases. *J. Biol. Inorg. Chem.* **16**, 1255–1268 (2011).

69. Cerqueira, N. M. F. S. A., Fernandes, P. A., Gonzalez, P. J., Moura, J. J. G. & Ramos, M. J. The

sulfur shift: An activation mechanism for periplasmic nitrate reductase and formate dehydrogenase. *Inorg. Chem.* **52**, 10766–10772 (2013).

70. Niks, D., Duvvuru, J., Escalona, M. & Hille, R. Spectroscopic and kinetic properties of the molybdenum-containing, NAD⁺-dependent formate dehydrogenase from *Ralstonia eutropha*. *J. Biol. Chem.* **291**, 1162–1174 (2016).

71. Dong, G. & Ryde, U. Reaction mechanism of formate dehydrogenase studied by computational methods. *J. Biol. Inorg. Chem.* **23**, 1243–1254 (2018).

72. Zinoni, F., Birkmann, A., Leinfelder, W. & Bock, A. Cotranslational insertion of selenocysteine into formate dehydrogenase from *Escherichia coli* directed by a UGA codon. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 3156–3160 (1987).

73. Heidelberg, J. F. *et al.* The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat. Biotechnol.* **22**, 554–559 (2004).

74. Duarte, A. G. *et al.* Redox loops in anaerobic respiration - The role of the widespread NrfD protein family and associated dimeric redox module. *Biochim. Biophys. Acta - Bioenerg.* **1862**, (2021).

75. da Silva, S. M. *et al.* Function of formate dehydrogenases in *Desulfovibrio vulgaris* Hildenborough energy metabolism. *Microbiol. (United Kingdom)* **159**, 1760–1769 (2013).

76. Lee, S. H., Choi, D. S., Kuk, S. K. & Park, C. B. Photobiocatalysis: Activating Redox Enzymes by Direct or Indirect Transfer of Photoinduced Electrons. *Angew. Chemie - Int. Ed.* **57**, 7958–7985 (2018).

77. Maciá-Agulló, J. A., Corma, A. & Garcia, H. Photobiocatalysis: The Power of Combining Photocatalysis and Enzymes. *Chem. - A Eur. J.* **21**, 10940–10959 (2015).

78. Schmermund, L. *et al.* Photo-Biocatalysis: Biotransformations in the Presence of Light. *ACS Catal.* **9**, 4115–4144 (2019).

79. Singleton, J. *Band Theory and Electronic Properties of Solids Band*. (Oxford University Press, 2001).

80. Ameta, R., Solanki, M. S., Benjamin, S. & Ameta, S. C. Photocatalysis. in *Advanced Oxidation Processes for Wastewater Treatment: Emerging Green Chemical Technology* 135–175 (2018).

81. Voet, D. & Voet, J. G. Photosynthesis. in *Biochemistry* 901–939 (John Wiley & Sons, 2010).

82. Habisreutinger, S. N., Schmidt-Mende, L. & Stolarczyk, J. K. Photocatalytic reduction of CO_2 on TiO2 and other semiconductors. *Angew. Chemie - Int. Ed.* **52**, 7372–7408 (2013).

83. Corma, A. & Garcia, H. Photocatalytic reduction of CO_2 for fuel production: Possibilities and challenges. *J. Catal.* **308**, 168–175 (2013).

84. Lan, Y., Lu, Y. & Ren, Z. Mini review on photocatalysis of titanium dioxide nanoparticles and their solar applications. *Nano Energy* **2**, 1031–1045 (2013).

85. Rochkind, M., Pasternak, S. & Paz, Y. Using dyes for evaluating photocatalytic properties: A critical review. *Molecules* **20**, 88–110 (2015).

86. Chowdhury, P., Gomaa, H. & Ray, A. K. Sacrificial hydrogen generation from aqueous triethanolamine with Eosin Y-sensitized Pt/TiO2 photocatalyst in UV, visible and solar light irradiation. *Chemosphere* **121**, 54–61 (2015).

87. Katoh, R. *et al.* Efficiencies of Electron Injection from Excited N3 Dye into Nanocrystalline

Semiconductor (ZrO$_2$, TiO$_2$, ZnO, Nb$_2$O$_5$, SnO$_2$, In$_2$O$_3$) Films. *J. Phys. Chem. B* **108**, 4818–4822 (2004).

88. Wang, Z. S., Sayama, K. & Sugihara, H. Efficient eosin Y dye-sensitized solar cell containing Br$^-$/Br$_3$$^-$ electrolyte. *J. Phys. Chem. B* **109**, 22449–22455 (2005).

89. Pellegrin, Y. & Odobel, F. Sacrificial electron donor reagents for solar fuel production. *Comptes Rendus Chim.* (2016).

90. Amao, Y. & Watanabe, T. Photochemical and enzymatic methanol synthesis from HCO$_3$$^-$ by dehydrogenases using water-soluble zinc porphyrin in aqueous media. *Appl. Catal. B Environ.* **86**, 109–113 (2009).

91. Secundo, F. & Amao, Y. Visible-light-driven CO$_2$ reduction to formate with a system of water-soluble zinc porphyrin and formate dehydrogenase in ionic liquid/aqueous media. *RSC Adv.* **10**, 42354–42362 (2020).

92. Miller, M. *et al.* Interfacing Formate Dehydrogenase with Metal Oxides for the Reversible Electrocatalysis and Solar-Driven Reduction of Carbon Dioxide. *Angew. Chemie - Int. Ed.* **58**, 4601–4605 (2019).

93. Fleishman, S. J. & Baker, D. Role of the biomolecular energy gap in protein design, structure, and evolution. *Cell* **149**, 262–273 (2012).

94. Goldenzweig, A. & Fleishman, S. J. Principles of Protein Stability and Their Application in Computational Design. *Annu. Rev. Biochem.* **87**, 105–129 (2018).

95. Anfinsen, C. B., Haber, E., Sela & White, F. H. The Kinetics of Formation of Native Ribonuclease During Oxidation of the Reduced Polypeptide Chain. *Proc. Natl. Acad. Sci.* **47**, 1309–1314 (1961).

96. Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys* **65**, 44–45 (1968).

97. Clark, C. A. Protein folding: Are we there yet? *Arch Biochem Biophys.* **469**, 1–3 (2008).

98. Beadle, B. M. & Shoichet, B. K. Structural bases of stability-function tradeoffs in enzymes. *J. Mol. Biol.* **321**, 285–296 (2002).

99. Tokuriki, N., Stricher, F., Serrano, L. & Tawfik, D. S. How protein stability and new functions trade off. *PLoS Comput. Biol.* **4**, 35–37 (2008).

100. Musil, M., Konegger, H., Hon, J., Bednar, D. & Damborsky, J. Computational Design of Stable and Soluble Biocatalysts. *ACS Catal.* **9**, 1033–1054 (2019).

101. Stimple, S. D., Smith, M. D. & Tessier, P. M. Directed evolution methods for overcoming trade-offs between protein activity and stability. *AIChE* **66**, (2020).

102. Mendes, J., Guerois, R. & Serrano, L. Energy estimation in protein design. *Curr. Opin. Struct. Biol.* **12**, 441–446 (2002).

103. Guerois, R., Nielsen, J. E. & Serrano, L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **320**, 369–387 (2002).

104. Chen, Y. *et al.* PremPS: Predicting the impact of missense mutations on protein stability. *PLoS Comput. Biol.* **16**, (2020).

105. Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat. Methods* **17**, 665–680 (2020).

106. Alford, R. F. *et al.* The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).

107. Simons, K. T., Kooperberg, C., Huang, E. & Baker, D. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.* **268**, 209–225 (1997).

108. Whitehead, T. A. *et al.* Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing. *Nat. Biotechnol.* **30**, 543–548 (2012).

109. Goldenzweig, A. *et al.* Automated Structure- and Sequence-Based Design of Proteins for High Bacterial Expression and Stability. *Mol. Cell* **63**, 337–346 (2016).

110. Ganesan, A. *et al.* Structural hot spots for the solubility of globular proteins. *Nat. Commun.* **7**, 1–15 (2016).

111. Dombkowski, A. A., Sultana, K. Z. & Craig, D. B. Protein disulfide engineering. *FEBS Lett.* **588**, 206–212 (2013).

112. Bednar, D. *et al.* FireProt: Energy- and Evolution-Based Computational Design of Thermostable Multiple-Point Mutants. *PLoS Comput. Biol.* **11**, 1–20 (2015).

113. Musil, M. *et al.* FireProt: web server for automated design of thermostable proteins. *Nucleic Acids Res.* **45**, W393–W399 (2017).

114. Campeotto, I. *et al.* One-step design of a stable variant of the malaria invasion protein RH5 for use as a vaccine immunogen. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 998–1002 (2017).

115. Khersonsky, O. *et al.* Automated Design of Efficient and Functionally Diverse Enzyme Repertoires. *Mol. Cell* **72**, 178–186 (2018).

116. Bassegoda, A., Madden, C., Wakerley, D. W., Reisner, E. & Hirst, J. Reversible interconversion of CO_2 and formate by a molybdenum-containing formate dehydrogenase. *J. Am. Chem. Soc.* **136**, 15473–15476 (2014).

117. Caserta, G. *et al.* The large subunit of the regulatory [NiFe]-hydrogenase from *Ralstonia eutropha* - a minimal hydrogenase? *Chem. Sci.* **11**, 5453–5465 (2020).

118. Keller, K. L., Wall, J. D. & Chhabra, S. *Methods for Engineering Sulfate Reducing Bacteria of the Genus Desulfovibrio. Synthetic Biology, Part A* vol. 497 (Elsevier Inc., 2011).

119. Micsonai, A. *et al.* Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3095–E3103 (2015).

120. Micsonai, A. *et al.* BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res.* **46**, W315–W322 (2018).

121. Alford, R. F. *et al.* The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput* **13**, 3031–3048 (2017).

122. Hanwell, M. D., Curtis, D. E., Lonie, D. C. & Vandermeersch, T. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **4**, (2012).

123. Altschul, S. F., Gertz, E. M., Agarwala, R., Schäffer, A. A. & Yu, Y. K. PSI-BLAST pseudocounts and the minimum description length principle. *Nucleic Acids Res.* **37**, 815–824 (2009).

124. Li, W. & Godzik, A. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).

125. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graph.* **14**, 33–98 (1996).

126. O'Meara, M. J. *et al.* Combined covalent-electrostatic model of hydrogen bonding improves structure prediction with Rosetta. *J. Chem. Theory Comput.* **11**, 609–622 (2015).

127. Gasteiger, E. *et al.* ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–3788 (2003).

128. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2- A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

129. Layne, E. Spectrophotometric and turbidimetric methods for measuring proteins. *Methods Enzymol.* **3**, 447–454 (1957).

130. Ugulava, N. B., Gibney, B. R. & Jarret, J. T. Iron-Sulfur Cluster Interconversions in Biotin Synthase: Dissociation and Reassociation of Iron during Conversion of [2Fe-2S] to [4Fe-4S] Clusters. *Biochemistry* **39**, 5206–5214 (2000).

131. Fisher, W. R., Taniuchi, H. & Anfinsen, C. B. On the role of heme in the formation of the structure of cytochrome c. *J. Biol. Chem.* **248**, 3188–3195 (1973).

132. Palmer, T. & Berks, B. C. The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol.* **10**, 483–496 (2012).

133. Jormakka, M., Byrne, B. & Iwata, S. Formate dehydrogenase - A versatile enzyme in changing environments. *Curr. Opin. Struct. Biol.* **13**, 418–423 (2003).

134. DeLisat, M. P., Tullman, D. & Georgiou, G. Folding quality control in the export of proteins by the bacterial twin-arginine translocation pathway. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6115–6120 (2003).

135. Schwanhold, N., Iobbi-Nivol, C., Lehmann, A. & Leimkühler, S. Same but different: Comparison of two system-specific molecular chaperones for the maturation of formate dehydrogenases. *PLoS One* **13**, 1–24 (2018).

136. Li, Y., Guo, M., Peng, S., Lu, G. & Li, S. Formation of multilayer-Eosin Y-sensitized TiO_2 via Fe_3^+ coupling for efficient visible-light photocatalytic hydrogen evolution. *Int. J. Hydrogen Energy* **34**, 5629–5636 (2009).

137. Liu, Z., Xu, X., Fang, J., Zhu, X. & Li, B. Synergistic degradation of eosin y by photocatalysis and electrocatalysis in UV irradiated solution containing hybrid $BiOCl/TiO_2$ particles. *Water. Air. Soil Pollut.* **223**, 2783–2798 (2012).

138. Herculano, L. S. *et al.* Investigation of the photobleaching process of eosin y in aqueous solution by thermal lens spectroscopy. *J. Phys. Chem. B* **117**, 1932–1937 (2013).

139. Alvarez-Martin, A. *et al.* Photodegradation mechanisms and kinetics of Eosin-Y in oxic and anoxic conditions. *Dye. Pigment.* **145**, 376–384 (2017).

140. Zahradník, J. *et al.* Flexible regions govern promiscuous binding of IL-24 to receptors IL-20R1 and IL-22R1. *FEBS J.* **286**, 3858–3873 (2019).

Annex

**Table 12:** Primers designed for the insertion of the *Eco*RI restriction site in the plasmid pRec-FdhAB-Strep.

Primer Forward (5' to 3')	GCC GCT TGG GTG GAG AGG GAA TTC GGC TAT GAC TGG GCA C
Primer Reverse (5' to 3')	GTG CCC AGT CAT AGC CGA ATT CCC TCT CCA CCC AAG CGG C
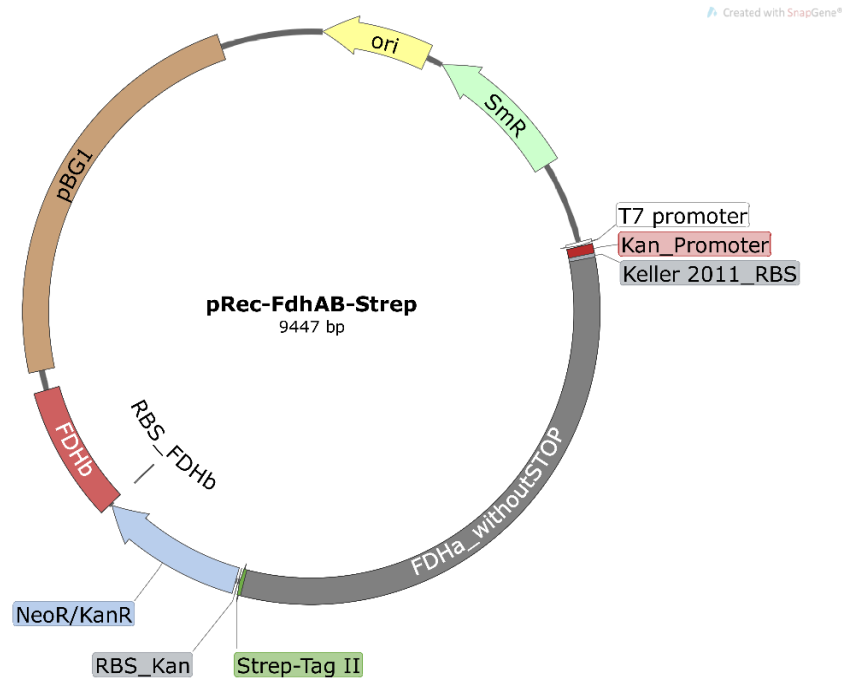
Figure 26: Representation of the pRec-FdhAB-Strep plasmid. Ori and pBG1 are the origins of replication in *E. coli* and *D. vulgaris*, respectively. *SmR* and *NeoR/KanR* are cassettes of resistances to spectinomycin and kanamycin, respectively. The latter confers resistance to gentinomycin in *D. vulgaris*. *FdhA_withoutSTOP* and *FdhB* encode the α- and β-subunits of formate dehydrogenase from *D. vulgaris*. Constructed using SnapGene®
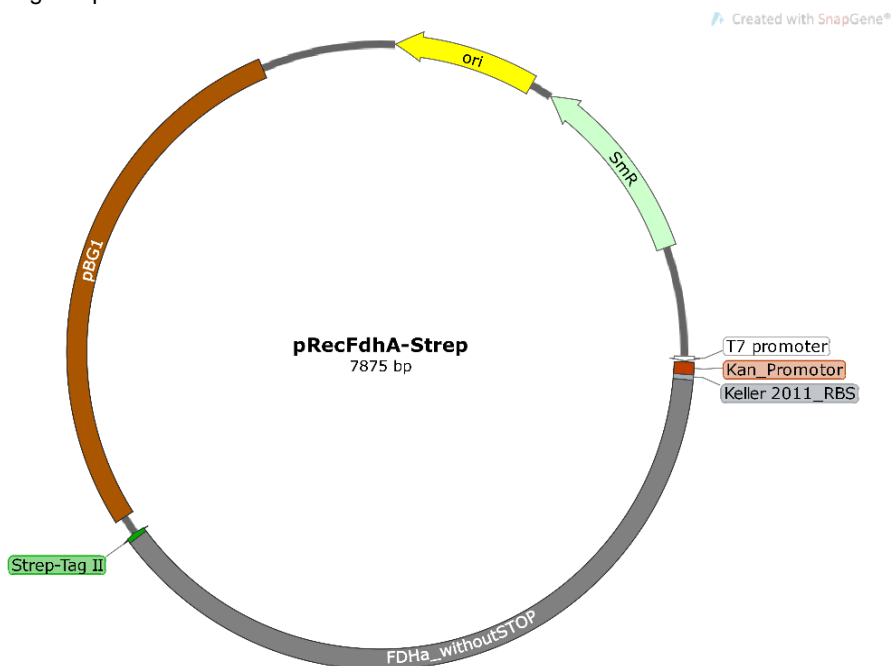

Figure 27: Representation of the pRec-FdhA-Strep plasmid. *Ori* and *pBG1* are the origins of replication in *E. coli* and *D. vulgaris*, respectively. *SmR* is a cassette of resistances to spectinomycin. *FdhA_withoutSTOP* encodes the α-subunit of formate dehydrogenase from *D. vulgaris*. Constructed using SnapGene®
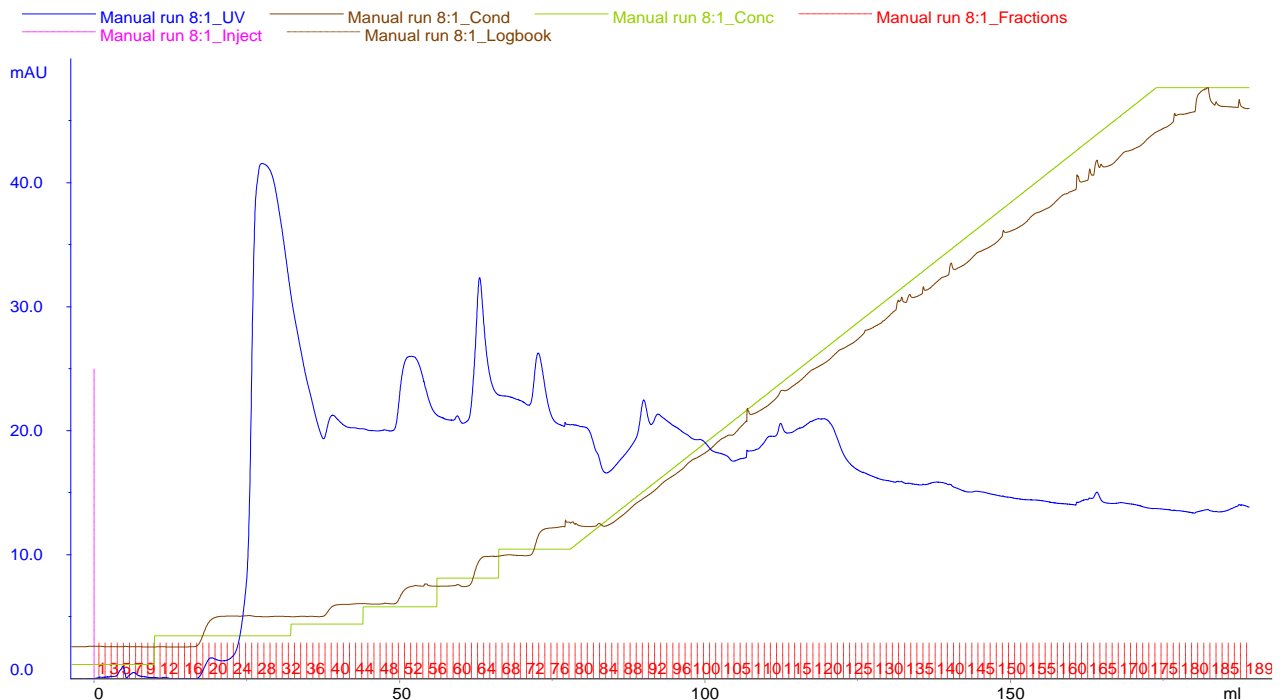
Figure 28: Chromatogram of the purification of Elution A from the aerobic purification of FdhA using a Q-Resource™ column, equilibrated with sample buffer (20 mM Tris-HCl pH 7.6 with 10% (v/v) glycerol and 10 mM NaNO₃). The elution was performed by increasing NaCl concentrations (green line), with a first increment of 50 mM followed by an increment of 25 mM and by successful increments of 50 mM NaCl until a concentration of 250 mM was reached. Afterwards a linear gradient was applied. At each mL of eluted sample, it is possible to follow the absorbance at 280 nm (blue line) and the corresponding tubes that the sample is being eluted at that time (red lines and numbers). Several tubes were pooled together generated fractions P0 (tube 0-22), P1 (tube 23-36), P2 (tube 37-48), P3 (tube 49-60), P4 (tube 61-71), P5 (tube 72-86), P6 (tube 87-106) and P7 (tube 107-189).
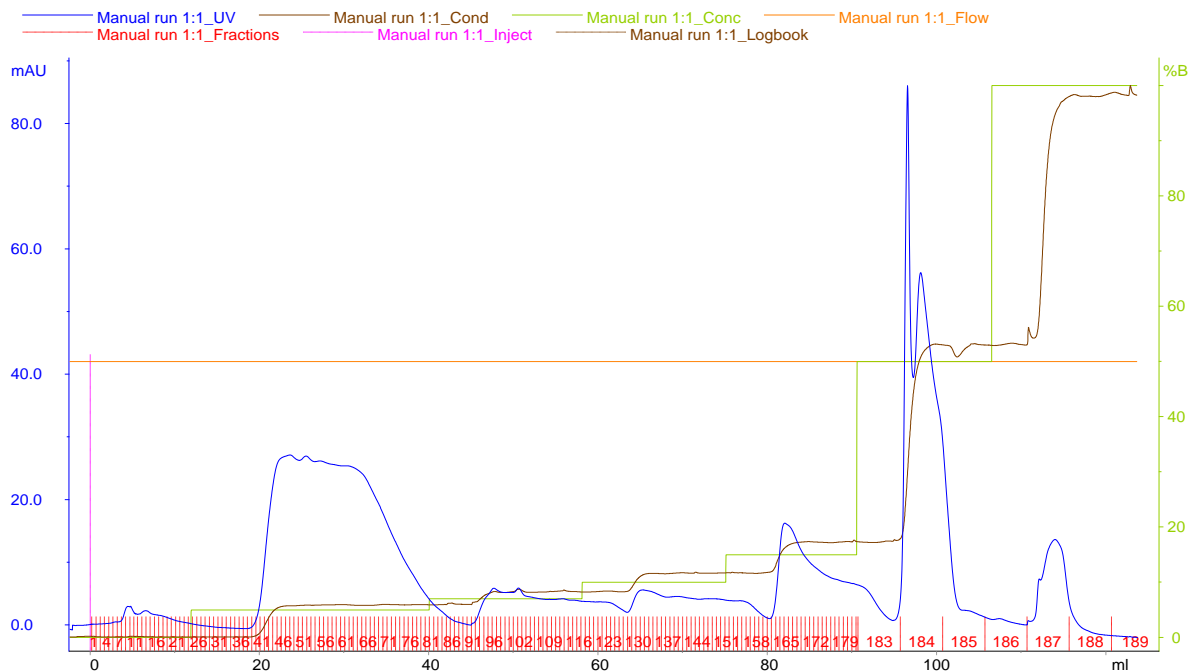

Figure 29: Chromatogram of the purification of Elution B from the aerobic purification of FdhA using a Q-Resource™ column, equilibrated with sample buffer (20 mM Tris-HCl pH 7.6 with 10% (v/v) glycerol and 10 mM NaNO₃). The elution was performed by increasing NaCl concentrations (green line), with a first increment of 50 mM followed by an increment of 25 mM and by successful increments of 50 mM NaCl until a concentration of 250 mM was reached. Afterwards NaCl concentration was increased to 500 mM and to 1 M. At each mL of eluted sample, it is possible to follow the absorbance at 280 nm (blue line) and the corresponding tubes that the sample is being eluted at that time (red lines and numbers). Several tubes were pooled together generated fractions P0 (tube 0-38), P1 (tube 39-90), P2 (tube 91-127), P3 (tube 128-161), P4 (tube 162-182), P5 (tube 183-186), P6 (tube 187-189) and P7 (tube 107-189).
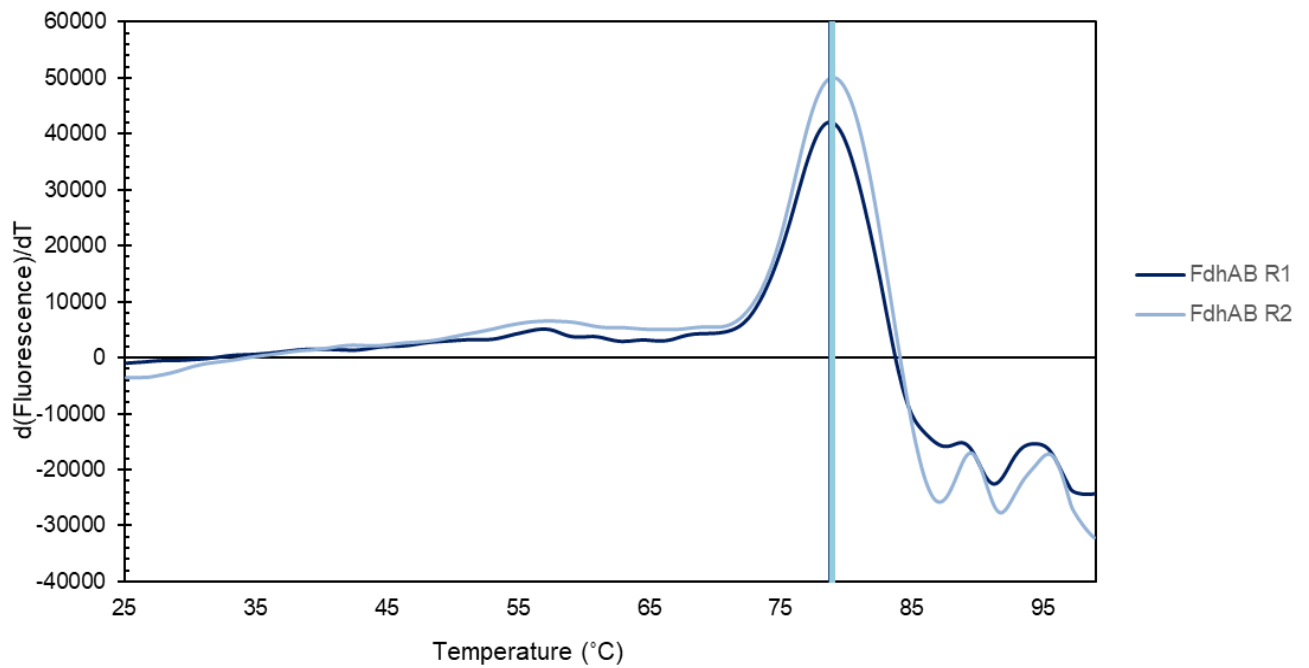
Figure 30: Thermal shift assay of FdhAB, with the derivative of the fluorescence over time being represented in the y axis and the temperature (˚C) in the x axis. The two replicates performed are represented in dark blue (FdhAB R1) and light blue (FdhAB R2), with the maximum of d(Fluorescence)/dT (the melting temperature) being represented for each in their respective colour.
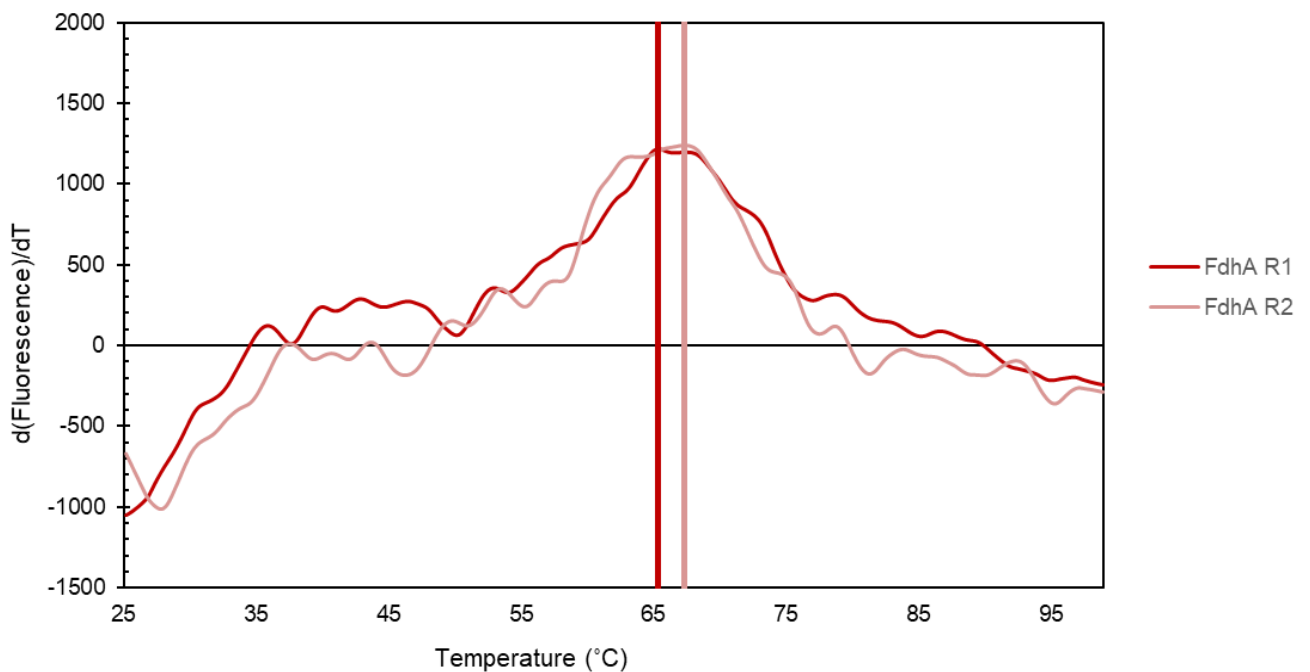
Figure 31: Thermal shift assay of FdhA, with the derivative of the fluorescence over time being represented in the y axis and the temperature (˚C) in the x axis. The two replicates performed are represented in dark red (FdhA R1) and light red (FdhA R2), with the maximum of d(Fluorescence)/dT (the melting temperature) being represented for each in their respective colour.

Table 13: Total number of amino acids being subjected to design for each protein (FdhA and FdhAB). Total number of amino acids fixed and restricted to the algorithm and the list of the residues (number and chain). In blue, residues that are restricted for the design of both proteins are represented.

Protein	Total Number of Amino acids	Number of Amino acids Fixed and Restricted	List of Amino acids Fixed and Restricted
FdhA	970	75	15A, 16A, 17A,18A, 20A, 21A, 22A, 52A, 53A, 55A, 56A, 127A, 129A, 153A,156A, 157A, 158A, 190A, 191A, 192A, 193A, 194A, 197A, 198A, 199A, 200A, 217A, 218A, 219A, 220A, 221A, 223A, 235A, 237A, 238A, 239A, 240A, 369A, 370A, 371A, 372A, 407A, 408A, 411A, 517A, 518A, 519A, 522A, 523A, 544A, 545A, 573A, 574A, 575A, 576A, 579A, 810A, 837A 847A, 848A, 849A, 850A, 851A, 853A, 854A, 855A, 856A, 858A, 859A, 929A, 930A, 931A, 942A, 944A, 961A
FdhAB	1184 (970 α-subunit and 214 β-subunit)	111 (70 α-subunit and 41 β-subunit)	15A, 16A, 17A, 18A, 20A, 21A, 22A, 52A, 54A, 55A, 56A, 127A, 129A, 153A, 156A, 158A, 191A, 192A, 193A,197A,198A, 199A, 200A, 217A, 219A, 220A, 221A, 223A, 237A, 238A, 240A, 369A, 370A, 371A, 372A, 407A, 408A, 411A, 517A, 518A, 519A, 522A, 523A, 544A, 545A, 573A, 574A, 575A, 576A, 810A, 837A, 846A, 847A, 848A, 849A, 850A, 851A, 852A, 853A, 854A, 855A, 856A, 858A, 859A, 929A, 930A, 942A, 944A,961A, 962A, 974B, 981B, 982B, 983B, 984B, 985B, 986B, 987B, 991B, 995B, 1019B, 1020B, 1022B, 1040B, 1043B, 1044B, 1045B, 1046B, 1049B, 1050B, 1051B, 1072B, 1074B, 1090B, 1091B, 1092B, 1094B, 1095B, 1106B, 1107B, 1108B, 1109B, 1110B, 1120B, 1121B, 1122B, 1126B, 1127B, 1128B, 1130B, 1131B

,

,

FdhA and Design Sequences

```
                      10         20         30         40
FdhA/1-970     1  ELQKLQWAKQTTSICCYCAVGCGLIVHTAKDGQGRAVNVEGDPDHPINEG  50
dFdhA1/1-970   1  ELQKLRRAKQTTSICCYCAVGCGLIVYVDKDGQGRAINVEGDPDHPINEG  50
dFdhA2/1-970   1  ELQKLRRAKQTTSICCYCAVGCGLIVYVDKDGKGRAINVEGDPDHPINEG  50
dFdhA3/1-970   1  QLQKLKRAKQTTSICCYCAVGCGLIVYVDKDGKGRAIHVEGDPDHPINEG  50
dFdhA4/1-970   1  QLQKLKRAKQTTSICCYCAVGCGLIVYVDKDGKGRAINVEGDPDHPINEG  50
dFdhA5/1-970   1  QLQKLKRAKQTTSICCYCAVGCGLIVYVDKDGKGRAINVEGDPDHPINEG  50
dFdhA6/1-970   1  ELPKLKRAKQTTSICCYCAVGCGLIVYVDKDGKGRAINVEGDPDHPINEG  50
dFdhA7/1-970   1  QLQKLKDAKQTTSICCYCAVGCGLIVYVDKDGKGRAINVEGDPDHPINEG  50
dFdhA8/1-970   1  QLLKLKDAKETTSICCYCAVGCGLIVYVDKDGKGRVINVEGDPDHPINEG  50

Consensus         QLQKLKRAKQTTSICCYCAVGCGLIVYVDKDGKGRAINVEGDPDHPINEG


                      60         70         80         90
FdhA/1-970    51  SLCPKGASIFQLGENDQRGTQPLYRAPFSDTWKPVTWDFALTEIAKRIKK  100
dFdhA1/1-970  51  SLCPKGASIWQLVNNPQRITQPLYRAPYSDKWEPVSWDFALDEIAKRIKK  100
dFdhA2/1-970  51  SLCPKGASIWQLVNNPQRLTQPLYRAPYSDKWEPVSWDFALDEIAKRIKK  100
dFdhA3/1-970  51  SLCPKGASIWQLVENPQRVTKPLYRAPYSDKWEPVSWDFALDRIAKRIKK  100
dFdhA4/1-970  51  SLCPKGASIWQLVENPQRITKPLYRAPYSDKWEEVSWDFALDRIAKRIKK  100
dFdhA5/1-970  51  SLCPKGASIWQLVENPQRITKPLYRAPYSDKWEEVSWDFALDRIAKRIKK  100
dFdhA6/1-970  51  SLCPKGASIWQLVENPQRVTKPMYRAPYSDKWEEVSWDFALDRIARRIKK  100
dFdhA7/1-970  51  SLCPKGASIWQLVENPQRVTKPMYRAPYSDKWEEVSWDFALDRIARRIKK  100
dFdhA8/1-970  51  SLCPKGASILQLTENPNRVQKPMYRAPYSDKWEEVSWDFALDRIAKLIKE  100

Consensus         SLCPKGASIWQLVENPQRVTKPLYRAPYSDKWEEVSWDFALDRIAKRIKK


                     110        120        130        140
FdhA/1-970   101  TRDASFTEKNAAGDLVNRTEAIASFGSAAMDNEECWAYGNILRSLGLVYI  150
dFdhA1/1-970 101  TRDETFMEKNAAGQTVNRVEAIASLGSAAMSNEECWVYTAMLRALGIVYI  150
dFdhA2/1-970 101  TRDETFIEKNAAGQTVNRVEAIASLGSAAMSNEECWVYTAMLRALGIVYI  150
dFdhA3/1-970 101  TRDETFVEKNAAGQTVNRVEAIASLGSAAMSNEECWVWTAMLRALGIVYI  150
dFdhA4/1-970 101  TRDETFVEKNAAGQTVNRVEAIASLGSAAMSNEECWVWTAMLRALGIVYI  150
dFdhA5/1-970 101  TRDETFIEKNDAGQTVNRVEAIASLGSAAMSNEECWVYTAMMRALGLVYI  150
dFdhA6/1-970 101  TRDKTFIEKNDKGQTVNRVEAIASLGSAAMSNEECWVYTAMMRALGLVYI  150
dFdhA7/1-970 101  TRDKTFVEKNDKGQTVNRVEAIASLGSAAMSNEECWVYTAMMRALGLVYI  150
dFdhA8/1-970 101  TRDKTFVEKNDKGQTVNRVEAIASLGSAAMSNEECWVYTAMMRALGLVYI  150

Consensus         TRDETFVEKNAAGQTVNRVEAIASLGSAAMSNEECWVYTAMLRALGLVYI


                     160        170        180        190
FdhA/1-970   151  EHQARICHSPTVPALAESFGRGAMTNHWNDLANSDCILIMGSNAAENHPI  200
dFdhA1/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhA2/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhA3/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhA4/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhA5/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200
dFdhA6/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200
dFdhA7/1-970 151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200
dFdhA8/1-970 151  EHQARICHSATVAALAETFGRGAMTNHWIDLKNADCILIMGSNAAENHPI  200

Consensus         EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI


                     210        220        230        240
FdhA/1-970   201  AFKWVLRAKDKGATLIHVDPRFTRTSARCDVYAPIRSGADIPFLGGLIKY  250
dFdhA1/1-970 201  AFKWVLRAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA2/1-970 201  AFKWVLRAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA3/1-970 201  AFKWVLRAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA4/1-970 201  AFKWVLEAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA5/1-970 201  AFKWVLEAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA6/1-970 201  AFKWVLEAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA7/1-970 201  AFKWVLEAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhA8/1-970 201  AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250

Consensus         AFKWVL+AKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY
```

74

Multiple sequence alignment (residues 251–500)

```
                         260       270       280       290
FdhA/1-970     251 ILDNKLYFTDYVREYTNASLIVGEKFSFKDGLFSGYDAANKKYDKSMWAF 300
dFdhA1/1-970   251 ILDNNLYHKDYVREYTNASLIVDEKFDFKDGLFSGYDPANKKYDKSTWAF 300
dFdhA2/1-970   251 ILDNNLYFKDYVREYTNASLIVDEKFDFKDGLFSGYDPANKKYDKSTWAF 300
dFdhA3/1-970   251 ILDNNLYHKDYVREYTNASFIVDEKFDFKDGLFSGYDPKNKKYDKSTWAF 300
dFdhA4/1-970   251 ILDNNLYFKDYVREYTNASFIVDPKFDFKDGLFSGYDPKNKKYDKSTWAF 300
dFdhA5/1-970   251 ILDNNLYFKDYVREYTNASLIVDPKFDFKDGLFSGYDPKNKKYDKSTWAF 300
dFdhA6/1-970   251 ILDNNLYFKDYVREYTNASFIVDPKFDFKDGLFSGYDPKNKKYDKSTWAF 300
dFdhA7/1-970   251 ILDNNLYHKDYVREYTNASFIVRPKFDFKDGLFSGYDPKNKKYDKSTWAF 300
dFdhA8/1-970   251 ILDNNLYFHDVVEYTNATFIVRPDFDFKDGLFSGYDPKNKKYDKSKWAF  300

Consensus          ILDNNLYFKDYVREYTNASFIVDPKFDFKDGLFSGYDPKNKKYDKSTWAF
```

```
                         310       320       330       340
FdhA/1-970     301 ELDANGVPKRDPALKHPRCVINLLKKHYERYNLDKVAAITGTSKEQLQQV 350
dFdhA1/1-970   301 ELDENGVPKRDPTLKHPRCVFNLLKKHYSRYNLDKVAKITGTSKEQLQV  350
dFdhA2/1-970   301 ELDENGVPKRDPTLKHPRCVFNLLKKHYSRYNLDKVAKITGTSKEQLLQV 350
dFdhA3/1-970   301 ELDENGVPKRDPTLKHPRCVFNLLKKHYSRYNLDKVASITGTSKEQLLQV 350
dFdhA4/1-970   301 ELDENGVPKRDPTLKHPRCVFNLLKKHYSRYDLDKVAKITGTSKEQLLQV 350
dFdhA5/1-970   301 ELDENGVPKRDPTLKHPRCVFNLLKKHYSRYDLDKVAKITGTSKEQLLKV 350
dFdhA6/1-970   301 ELDENGVPKRDPTLQHPRCVFNLLKKHYSRYDLDKVAKITGTSKEQLLKV 350
dFdhA7/1-970   301 ELDENGVPKRDPTLQHPRCVFNLLKKHYSRYDLDKVAKITGTSKEQLLKV 350
dFdhA8/1-970   301 ELDENGVPKRDPTLQHPRCVFNLLKKHYSRYDLDKVSKITGTSKEQLLKV 350

Consensus          ELDENGVPKRDPTLKHPRCVFNLLKKHYSRYDLDKVAKITGTSKEQLLQV
```

```
                         360       370       380       390
FdhA/1-970     351 YKAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLLGNIGVAGG 400
dFdhA1/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLLGNIGVAGG 400
dFdhA2/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHTVGVQNIRAMAMIQLLLGNIGVAGG 400
dFdhA3/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLLGNIGIAGG 400
dFdhA4/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHTVGVQNIRAMAMIQLLLGNIGIAGG 400
dFdhA5/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHTVGVQNIRAMAMIQLLLGNIGIAGG 400
dFdhA6/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHTVGVQNIRAMAMIQLLLGNIGMAGG 400
dFdhA7/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHTVGVQNIRAMAMIQLLLGNIGMAGG 400
dFdhA8/1-970   351 YKAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRTMAMIQLLLGNIGMAGG 400

Consensus          YKAYAATGKPDKAGTIMYAMGWTQHTVGVQNIRAMAMIQLLLGNIG+AGG
```

```
                         410       420       430       440
FdhA/1-970     401 GVNALRGESNVQGSTDQGLLAHIWPGYNPVPNSKAATLELYNAATPQSKD 450
dFdhA1/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNASTPKTKD 450
dFdhA2/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNASTPKTKD 450
dFdhA3/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNASTPKTKD 450
dFdhA4/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPKSKHATLEDYNASTPKTKD 450
dFdhA5/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPKSKHATLEDYNASTPKTKD 450
dFdhA6/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPKSKHATLEDYNASTPKTKD 450
dFdhA7/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPKSKHATLEDYNAATPKTKD 450
dFdhA8/1-970   401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPKSKHQTLEDYNATTPKTKD 450

Consensus          GVNALRGESNVQGSTDQGLLAHILPGYLPVPKSKHATLEDYNASTPKTKD
```

```
                         460       470       480       490
FdhA/1-970     451 PMSVNWWQNRPKYVASYLKALYPDEEPAAAYDYLPRIDAGRKLTDYFWLN 500
dFdhA1/1-970   451 PMSVNWWQNTPKYVVSFLKAMYPDEEPAAYQYLPKWDADRKLADYFWLF  500
dFdhA2/1-970   451 PMSVNWWQNKPKYVVSFLKAMYPDEEPAAYQYLPKWDADKKLTDYFWLF  500
dFdhA3/1-970   451 PMSVNWWQNKPKYVVSFLKAMYPDEEPAAYQYLPKWDADKPLADYFWLF  500
dFdhA4/1-970   451 PMSVNWWQNKPKYVVSFLKAMYPDEEPKAAYQYLPKMDAGKPDADYFWLF 500
dFdhA5/1-970   451 PMSVNWWQNKPKYVVSFLKAMYPDEEPAAYQYLPKWDADGKKDADYFWLF 500
dFdhA6/1-970   451 PKSVNWWQNRPKYIVSFLKAMYPDEEPAAYQYLPKWDADKKDADYFWLF  500
dFdhA7/1-970   451 PKSANWWQNRPKYIVSFLKAMYPDEEPKAAYQYLPKMDADKPDADYFWLF 500
dFdhA8/1-970   451 PKSANWWQNRPKFIVSMLKAMYPDEEPEFAYQYLPKWDADKPDADYFWMF 500

Consensus          PMSVNWWQN+PKYVVSFLKAMYPDEEPAAYQYLPKWDADKKDADYFWLF
```

Multiple sequence alignment, residues 501–750.

Name	501	510	520	530	540	550

```
FdhA/1-970     501  IFEKMDKGEFKGLFAWGMNPACGGANANKNRKAMGKLEWLVNVNLFENET  550
dFdhA1/1-970   501  IFDMMYKGKFKGMFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET  550
dFdhA2/1-970   501  IFDKMYKGKFKGLFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET  550
dFdhA3/1-970   501  IFDKMYKGKFKGLFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET  550
dFdhA4/1-970   501  IFDKMYKGKFKGMFCWGMNPACGGPNANKIRKAMSKLDWMVVVNLFENET  550
dFdhA5/1-970   501  IFDMMYKGKFKGMFCWGMNPACGGPNANKIRKAMSKLDWMVVVNLFENET  550
dFdhA6/1-970   501  IFDKMYKGKFKGLFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET  550
dFdhA7/1-970   501  IFDKMYKGKFKGMFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET  550
dFdhA8/1-970   501  IFDKMYKGKFKGLFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET  550

Consensus           IFDKMYKGKFKGLFCWGMNPACGGPNANKIRKAMAKLDWMVVVNLFENET


FdhA/1-970     551  SSFWKGPGMNPAEIGTEVFFLPCCVSIEKEGSVANSGRWMQWRYRGPKPY  600
dFdhA1/1-970   551  SAFWRGPGMDPAEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYQGPKPY  600
dFdhA2/1-970   551  SAFWRGPGMDPAEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYQGPKPY  600
dFdhA3/1-970   551  SAFWRGPGMDPAEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYKGPEPY  600
dFdhA4/1-970   551  ASFWKGPGMDPSEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYKGPEPY  600
dFdhA5/1-970   551  ASFWKGPGMDPSEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYKGPEPY  600
dFdhA6/1-970   551  AAFWRGPGMDPSEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYKGPEPY  600
dFdhA7/1-970   551  AAFWRGPGMDPSKIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYKGPEPY  600
dFdhA8/1-970   551  AAFWRGPGMDPSKIKTEVFFLPCCVSIEKEGSITNSGRWMQWRWKGPEPY  600

Consensus           AAFWRGPGMDPSEIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRYKGPEPY


FdhA/1-970     601  AETKPDGDIMLDMFKKVRELYAKEGGAYPAPIAKLNIADWEEHNEFSPTK  650
dFdhA1/1-970   601  GEAKPDGDIMLELFKRVRELYAKEGGAFPAPILKLNIADWEEHNEFSPHK  650
dFdhA2/1-970   601  GEAKPDGDIMLELFLRIRELYAKEGGAFPAPILKLNWADWAEHNEFSPHK  650
dFdhA3/1-970   601  GETKPDGDIMIELFNRIRELYAKEGGAFPAPILKLNWADWEEHNEFSPHK  650
dFdhA4/1-970   601  GEAKPDGDIMIELFKRIRELYAKEGGAFPAPILKLNWADWAEHNEFSPHK  650
dFdhA5/1-970   601  GETKPDGDIMLELFLRIRELYAKEGGAFPAPILKLNWEDWAEHNEFSPHK  650
dFdhA6/1-970   601  GEAKPDGDIMLELFKRIRELYAKEGGAFPAPILKLNMDDWEEHNHFSPHK  650
dFdhA7/1-970   601  GEAKPDGDIMLELFKRIRELYAKEGGAFPAPILKLNMDDWEEHNHFSPHK  650
dFdhA8/1-970   601  GEAKPDGDIMLELFKRIRELYKKEGGAFPAPILKLNMDDWEEHDHFSPHK  650

Consensus           GEAKPDGDIMLELFKRIRELYAKEGGAFPAPILKLNWADWEEHNEFSPHK


FdhA/1-970     651  VAKLMNGYFLKDTEVGGKQFKKGQQVPSFAFLTADGSTCSGNWLHAGSFT  700
dFdhA1/1-970   651  VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTADGSTCCGNWLHAGSFT  700
dFdhA2/1-970   651  VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTADGSTCCGNWLHAGSFT  700
dFdhA3/1-970   651  VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTADGSTCCGNWLHAGSFT  700
dFdhA4/1-970   651  VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCCGNWLHCGSFT  700
dFdhA5/1-970   651  VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCCGNWLHCGSFT  700
dFdhA6/1-970   651  VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCCGNWLHCGSFT  700
dFdhA7/1-970   651  VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCCGNWIFCGSYT  700
dFdhA8/1-970   651  VAKLINGYFLKDTTVGGKTFKKGQQVPSFAYLTDDGSTCCGNWIFCGSYT  700

Consensus           VAKLINGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCCGNWLHCGSFT


FdhA/1-970     701  DAGNLMARRDKTQTPEQARIGLFPNWSFCWPVNRRILYNRASVDKTGKPW  750
dFdhA1/1-970   701  DAGNLMARRDLTQTPEQARIGLFPNWAWCWPANRRILYNRASVDLNGKPW  750
dFdhA2/1-970   701  DAGNLMARRDLTQTPDQARIGLFPNWAWCWPANRRILYNRASVDLNGKPW  750
dFdhA3/1-970   701  DEGNLMARRDLTQTPMQAAIGLFPNWAWAWPANRRILYNRASVDLNGKPW  750
dFdhA4/1-970   701  DEGNLMARRDLTQTPMQANIGLFPNWAWCWPANRRILYNRASVDLNGKPW  750
dFdhA5/1-970   701  DEGNLMARRDLTQTPMQANIGLFPNWAWCWPANRRILYNRASVDLNGKPW  750
dFdhA6/1-970   701  DEGNLMARRDLTQTPMQANIGLFPNWAWCWPANRRILYNRASVDLTGKPW  750
dFdhA7/1-970   701  DEGNLMARRDLTQTPMQANIGLFPNWAWCWPANRRILYNRASVDLTGKPW  750
dFdhA8/1-970   701  DEGNMMARRDLTQTPMQANIGLFPNWAWCWPANRRILYNRASVDLNGKPW  750

Consensus           DEGNLMARRDLTQTPMQANIGLFPNWAWCWPANRRILYNRASVDLNGKPW
```
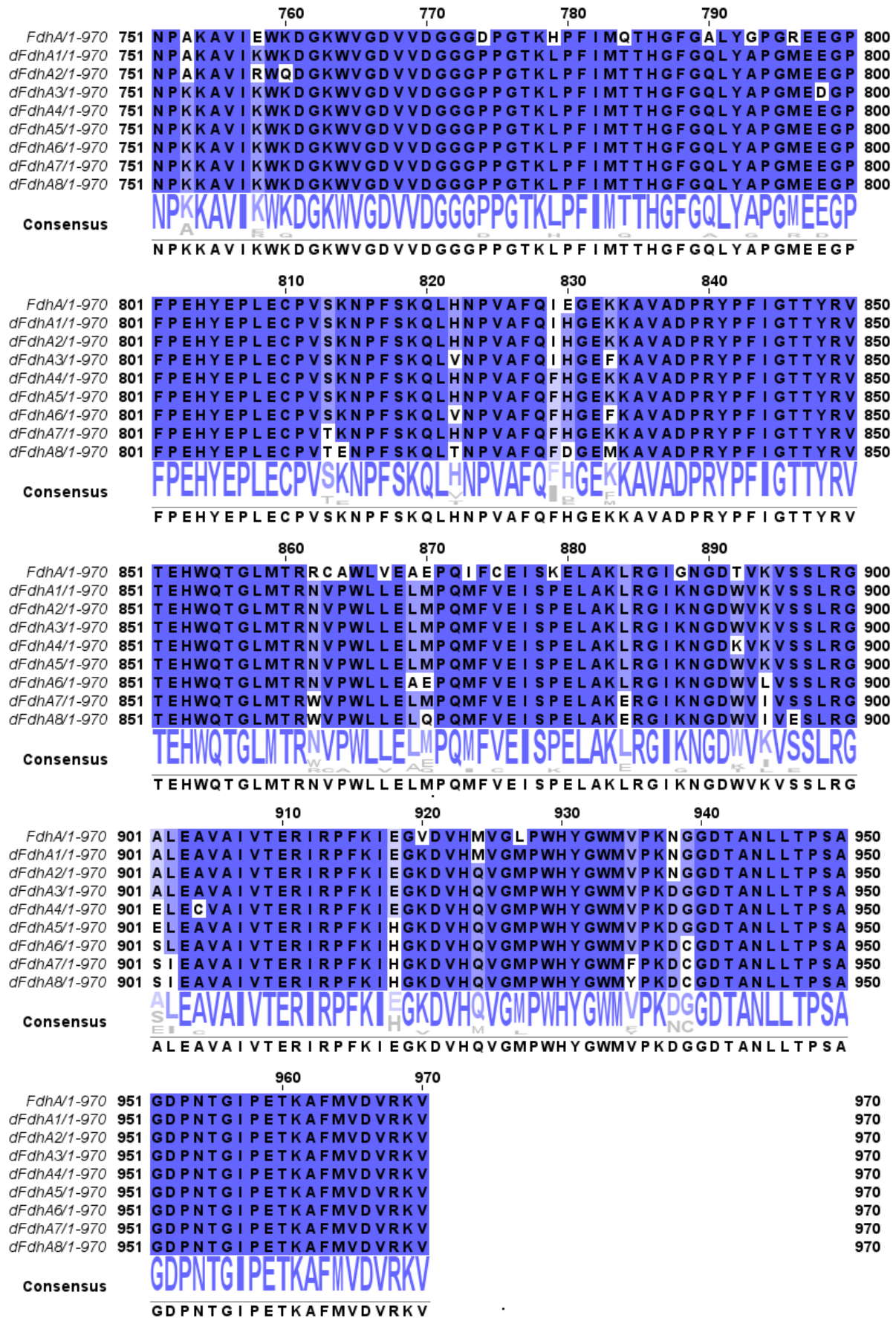
Figure 32: Sequence alignment of the simplified FdhA and of the designs. The shades of blue represent the degree of conservation of each position in the sequences (darkest shade: > 80%; medium shade: > 60%; light shade: > 40%; white colour < 40%). The alignment was conducted using JALVIEW[128].

FdhAB and Design Sequences


```
                      10        20        30        40
   FdhAB/1-1184    1  ELQKLQWAKQTTSICCYCAVGCGLIVHTAKDGQGRAVNVEGDPDHPINEG   50
dFdhAB1/1-1184    1  ELDKLKWAKQTTSICCYCAVGCGLIVYVAKDGQGRAVNVEGDPDHPINEG   50
dFdhAB2/1-1184    1  ELLKLKWAKQTTSICCYCAVGCGLIVYVAKDGQGRAVNVEGDPDHPINEG   50
dFdhAB3/1-1184    1  ELLKLKWAKQTTSICCYCAVGCGLIVYVAKDGQGRAVNVEGDPDHPINEG   50
dFdhAB4/1-1184    1  ELLKLKWAKQTTSICCYCAVGCGLIVYVAKDGQGRAVNVEGDPDHPINEG   50
dFdhAB5/1-1184    1  ELLKLKWAKQTTSICCYCAVGCGLIVHVAKDGQGRAVNVEGDPDHPINEG   50
dFdhAB6/1-1184    1  ELLKLKWAKQTTSICCYCAVGCGLIVYVDKDGQGRAVNVEGDPDHPINEG   50
dFdhAB7/1-1184    1  ELLKLKWAKQTTSICCYCAVGCGLIVHVDKDGQGRAVNVEGDPDHPINEG   50
dFdhAB8/1-1184    1  ELLKLKWAKETTSICCYCAVGCGLIVYVDKDGKGRAVNIEGDPDHPINEG   50

    Consensus        ELLKLKWAKQTTSICCYCAVGCGLIVYVAKDGQGRAVNVEGDPDHPINEG

                     ELLKLKWAKQTTSICCYCAVGCGLIVYVAKDGQGRAVNVEGDPDHPINEG


                     60        70        80        90
   FdhAB/1-1184   51  SLCPKGASIFQLGENDQRGTQPLYRAPFSDTWKPVTWDFALTEIAKRIKK  100
dFdhAB1/1-1184   51  SLCPKGASIWQLVENDQRGTKPLYRAPFSDKWKPVSWDFALTEIAKRIKK  100
dFdhAB2/1-1184   51  SLCPKGASIWQLVENDQRGTKPLYRAPFSDKWKPVSWDFALTEIAKRIKK  100
dFdhAB3/1-1184   51  ALCPKGASIWQLVENDQRGTKPLYRAPFSDKWKPVSWDFALTEIAKRIKK  100
dFdhAB4/1-1184   51  ALCPKGASIWQLVENDQRGTKPLYRAPFSDKWKPVSWDFALTEIAKRIKK  100
dFdhAB5/1-1184   51  ALCPKGASIWQLVENDQRGTKPLYRAPYSDKWEPVSWDFALTEIAKRIKK  100
dFdhAB6/1-1184   51  ALCPKGASIWQLVENPQRGTKPLYRAPYSDKWEPVSWDFALTEIAKRIKK  100
dFdhAB7/1-1184   51  ALCPKGASIWQLVENPQRGTKPLYRAPYSDKWEPVSWDFALTEIAKRIKK  100
dFdhAB8/1-1184   51  ALCPKGASIWQLVENPQRGTKPLYRAPYSDKWEPVSWDWALTEIAKRIKK  100

    Consensus        ALCPKGASIWQLVENDQRGTKPLYRAPFSDKWPVSWDFALTEIAKRIKK

                     ALCPKGASIWQLVENDQRGTKPLYRAPFSDKWKPVSWDFALTEIAKRIKK


                     110       120       130       140
   FdhAB/1-1184  101  TRDASFTEKNAAGDLVNRTEAIASFGSAAMDNEECWAYGNILRSLGLVYI  150
dFdhAB1/1-1184  101  TRDESFTEKNANGQLVNRTEAIASFGSAAMDNEECWAYSAILRSLGLVYI  150
dFdhAB2/1-1184  101  TRDESFTEKNANGQLVNRTEAIASFGSAAMDNEECWAYSAILRSLGLVYI  150
dFdhAB3/1-1184  101  TRDESFTEKNANGQLVNRTEAIASFGSAAMDNEECWAYSALLRSLGLVYI  150
dFdhAB4/1-1184  101  TRDESFIEKNANGQLVNRTEAIASFGSAAMSNEECWAYSALLRSLGLVYI  150
dFdhAB5/1-1184  101  TRDESFIEKNDQGQTVNRTEAIASFGSAAMSNEECWAYSALMRALGLVYI  150
dFdhAB6/1-1184  101  TRDESFIEKNDKGQTVNRTEAIASFGSAAMSNEECWAYTALMRALGLVYI  150
dFdhAB7/1-1184  101  TRDESFIEKNDKGQTVNRTEAIASFGSAAMSNEECWVYTAMMRALGLVYI  150
dFdhAB8/1-1184  101  TRDESFIEKNDKGQTVNRVEAIASFGSAAMSNEECWVYTALMRALGLVYI  150

    Consensus        TRDESFIEKNNGQLVNRTEAIASFGSAAMSNEECWAYSALLRLGLVYI

                     TRDESFIEKNANGQLVNRTEAIASFGSAAMSNEECWAYSALLRSLGLVYI


                     160       170       180       190
   FdhAB/1-1184  151  EHQARICHSPTVPALAESFGRGAMTNHWNDLANSDCILIMGSNAAENHPI  200
dFdhAB1/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhAB2/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhAB3/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhAB4/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI  200
dFdhAB5/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200
dFdhAB6/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200
dFdhAB7/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200
dFdhAB8/1-1184  151  EHQARICHSATVAALAESFGRGAMTNHWIDLANADCILIMGSNAAENHPI  200

    Consensus        EHQARICHSATVAALAESFGRGAMTNHWIDLANDCILIMGSNAAENHPI

                     EHQARICHSATVAALAESFGRGAMTNHWIDLANSDCILIMGSNAAENHPI


                     210       220       230       240
   FdhAB/1-1184  201  AFKWVLRAKDKGATLIHVDPRFTRTSARCDVYAPIRSGADIPFLGGLIKY  250
dFdhAB1/1-1184  201  AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB2/1-1184  201  AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB3/1-1184  201  AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB4/1-1184  201  AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB5/1-1184  201  AFKWVLKARDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB6/1-1184  201  AFKWVLKARDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB7/1-1184  201  AFKWVLKARDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250
dFdhAB8/1-1184  201  AFKWVLKARDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY  250

    Consensus        AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY

                     AFKWVLKAKDKGATLIHVDPRFTRTAARCDIYAPIRSGADIAFLGGMIKY
```

```
                         260        270        280        290
FdhAB/1-1184    251 ILDNKLYFTDYVREYTNASLIVGEKFSFKDGLFSGYDAANKKYDKSMWAF 300
dFdhAB1/1-1184  251 ILDNKLYHKDYVREYTNASLIVHEDFSFKDGLFSGYDEENKKYDKSTWAF 300
dFdhAB2/1-1184  251 ILDNNLYHKDYVREYTNASLIVDEKFSFKDGLFSGYDEENKKYDKSTWAF 300
dFdhAB3/1-1184  251 ILDNNLYHKDYVREYTNASLIVHEDFSFKDGLFSGYDEKNKKYDKSKWAF 300
dFdhAB4/1-1184  251 ILDNNLYHKDYVREYTNASLIVHEDFSFKDGLFSGYDEKNKKYDKSKWAF 300
dFdhAB5/1-1184  251 ILDNNLYHKDYVREYTNASLIVHEDFSFKDGLFSGYDPKNRKYDKSKWAF 300
dFdhAB6/1-1184  251 ILDNNLYHKDYVREYTNASLIVHEDFSFKDGLFSGYDPKNRKYDKSKWAF 300
dFdhAB7/1-1184  251 ILDNNLYHKDYVREYTNASLIVRPDFDFKDGLFSGYDPKNRKYDKSKWAF 300
dFdhAB8/1-1184  251 ILDNNLYHKDYVREYTNASLIVHEDFDFKDGLFSGYDPKNRKYDKSKWAF 300

Consensus          ILDNNLYHKDYVREYTNASLIVHEDFSFKDGLFSGYD+KNKKYDKSKWAF


                         310        320        330        340
FdhAB/1-1184    301 ELDANGVPKRDPALKHPRCVINLLKKHYERYNLDKVAAITGTSKEQLQQV 350
dFdhAB1/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYNLDKVASITGTSKEQLLKV 350
dFdhAB2/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVASITGTSKEQLLKV 350
dFdhAB3/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVASITGTSKEQLLKV 350
dFdhAB4/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVAKITGTSKEQLLKV 350
dFdhAB5/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVASITGTSKEQLLKV 350
dFdhAB6/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVASITGTSKEQLLKV 350
dFdhAB7/1-1184  301 ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVSKITGTSKEQLLKV 350
dFdhAB8/1-1184  301 ELDENGVPKRDPTLQHPRCVLQLLKKHYSRYDLDKVSKITGMSKEQLLKV 350

Consensus          ELDENGVPKRDPTLQHPRCVIQLLKKHYSRYDLDKVASITGTSKEQLLKV


                         360        370        380        390
FdhAB/1-1184    351 YKAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG 400
dFdhAB1/1-1184  351 YKAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG 400
dFdhAB2/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG 400
dFdhAB3/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG 400
dFdhAB4/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG 400
dFdhAB5/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG 400
dFdhAB6/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGIAGG 400
dFdhAB7/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQIIRAMAMIQLLGNIGIAGG 400
dFdhAB8/1-1184  351 YEAYAATGKPDKAGTIMYAMGWTQHSVGVQIIRAMAMIQLLGNIGIAGG 400

Consensus          YEAYAATGKPDKAGTIMYAMGWTQHSVGVQNIRAMAMIQLLGNIGVAGG


                         410        420        430        440
FdhAB/1-1184    401 GVNALRGESNVQGSTDQGLLAHIWPGYNPVPNSKAATLELYNAATPQSKD 450
dFdhAB1/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYNPVPNAKHATLEDYNAATPQSKD 450
dFdhAB2/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYNPVPNSKHATLEDYNAATPQSKD 450
dFdhAB3/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNAATPQSKD 450
dFdhAB4/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNAATPQSKD 450
dFdhAB5/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNAKHATLEDYNAATPQSKD 450
dFdhAB6/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNAATPKSKD 450
dFdhAB7/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNASTPKSKD 450
dFdhAB8/1-1184  401 GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNATTPKTKD 450

Consensus          GVNALRGESNVQGSTDQGLLAHILPGYLPVPNSKHATLEDYNAATPQSKD


                         460        470        480        490
FdhAB/1-1184    451 PMSVNWWQNRPKYVASYLKALYPDEEPAAAYDYLPRIDAGRKLTDYFWLN 500
dFdhAB1/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDETDYSWLA 500
dFdhAB2/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDATDYFWLM 500
dFdhAB3/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDETDYSWLA 500
dFdhAB4/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDATDYFWLM 500
dFdhAB5/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDATDYFWLN 500
dFdhAB6/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDETDYSWLN 500
dFdhAB7/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDEEPDFAYNYLPRLDADKDETDYSWLK 500
dFdhAB8/1-1184  451 PKSVNWWQNRPKYVASYLKALYPDQEPDFAYHYLPRLDADKDETDYSWLK 500

Consensus          PKSVNWWQNRPKYVASYLKALYPDEEPEEAYDYLPRLDAGKDETDYSWLN
```

Block 1 (positions 501-550):

FdhAB/1-1184	501	IFEKMDKGEFKGLFAWGMNPACGGANANKNRKAMGKLEWLVNVNLFENET	550
dFdhAB1/1-1184	501	IFEKMYKGEFKGLFAWGMNPACGGANANKIRKALAKLEWLVVVNLFENET	550
dFdhAB2/1-1184	501	IFEKMYKGEFKGLFAWGMNPACGGANANKIRKALSKLEWLVVVNLFENET	550
dFdhAB3/1-1184	501	IFEKMYKGEFKGLFAWGMNPACGGANANKIRKALAKLEWLVVVNLFENET	550
dFdhAB4/1-1184	501	IFEKMYKGEFKGLFAWGMNPACGGANANKIRKALAKLEWLVVVNLFENET	550
dFdhAB5/1-1184	501	IFDRMYKGEFKGLFCWGMNPACGGANANKIRKALAKLEWLVVVNLFENET	550
dFdhAB6/1-1184	501	IFDKMYKGEFKGLFCWGMNPACGGANANKIRKALAKLEWLVVVNLFENET	550
dFdhAB7/1-1184	501	IFDRMYKGEFKGLFCWGMNPACGGANANKIRKALSKLEWLVVVNLFENET	550
dFdhAB8/1-1184	501	IFDRMYKGEFKGLFCWGMNPACGGANANKIRKALAKLEWLVVVNLFENET	550

Consensus

IFEKMYKGEFKGLFAWGMNPACGGANANKIRKALAKLEWLVVVNLFENET

Block 2 (positions 551-600):

FdhAB/1-1184	551	SSFWKGPGMNPAEIGTEVFFLPCCVSIEKEGSVANSGRWMQWRYRGPKPY	600
dFdhAB1/1-1184	551	SSFWRGPGMDPAKIGTEVFFLPCCVSIEKEGSVANSGRWMQWRYRGAKPY	600
dFdhAB2/1-1184	551	SSFWRGPGMDPAKIGTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY	600
dFdhAB3/1-1184	551	ASFWKGPGMDPAKIKTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY	600
dFdhAB4/1-1184	551	ASFWRGPGMDPAKIKTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY	600
dFdhAB5/1-1184	551	ASFWKGPGMDPKKIKTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY	600
dFdhAB6/1-1184	551	ASFWKGPGMDPKKIKTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY	600
dFdhAB7/1-1184	551	ASFWKGPGMDPSKIKTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY	600
dFdhAB8/1-1184	551	ASFWRGPGMDPKKIKTEVFFLPCCVSIEKEGSVTNSGRWMQWRWQGAKPY	600

Consensus

ASFWRGPGMDPAKIKTEVFFLPCCVSIEKEGSVANSGRWMQWRYQGAKPY

Block 3 (positions 601-650):

FdhAB/1-1184	601	AETKPDGDIMLDMFKKVRELYAKEGGAYPAPIAKLNIADWEEHNEFSPTK	650
dFdhAB1/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIADWEEHNHFSPHK	650
dFdhAB2/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIADWEEHNHFSPHK	650
dFdhAB3/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIADWEEHNHFSPHK	650
dFdhAB4/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIDDWEEHNHFSPHK	650
dFdhAB5/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIDDWEEHNHFSPHK	650
dFdhAB6/1-1184	601	GETKPDGDIMLDLFKKVKELYAKEGGAFPAPILKLNTDDWEEHNHFSPHK	650
dFdhAB7/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIDDWEEHNHFSPHK	650
dFdhAB8/1-1184	601	GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNTDDWEEHDHFSPHK	650

Consensus

GETKPDGDIMLDLFKKVRELYAKEGGAFPAPILKLNIDDWEEHNHFSPHK

Block 4 (positions 651-700):

FdhAB/1-1184	651	VAKLMNGYFLKDTEVGGKQFKKGQQVPSFAFLTADGSTCSGNWLHAGSFT	700
dFdhAB1/1-1184	651	VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTADGSTCSGNWLHAGSYT	700
dFdhAB2/1-1184	651	VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTADGSTCSGNWLHAGSYT	700
dFdhAB3/1-1184	651	VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTADGSTCSGNWLHAGSYT	700
dFdhAB4/1-1184	651	VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCSGNWLHAGSYT	700
dFdhAB5/1-1184	651	VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCCGNWLHAGSYT	700
dFdhAB6/1-1184	651	VAKLMNGYFLKDVEVGGKLFKKGQQVPSFAFLTDDGSTCCGNWLHAGSYT	700
dFdhAB7/1-1184	651	VAKLINGYFLKDVTVGGKEYKKGQQVPSFAFLTDDGSTCCGNWLHCGSYT	700
dFdhAB8/1-1184	651	VAKLINGYFLRDVTVGGKEYKKGQQVPSFAYLQDDGSTACGNWLYCGSYT	700

Consensus

VAKLMNGYFLKDVEVGGKEFKKGQQVPSFAFLTDDGSTCSGNWLHAGSYT

Block 5 (positions 701-750):

FdhAB/1-1184	701	DAGNLMARRDKTQTPEQARIGLFPNWSFCWPVNRRILYNRASVDKTGKPW	750
dFdhAB1/1-1184	701	DAGNLMARRDNTQTPDQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW	750
dFdhAB2/1-1184	701	DAGNLMARRDNTQTPMQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW	750
dFdhAB3/1-1184	701	DAGNLMARRDNTQTPMQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW	750
dFdhAB4/1-1184	701	DEGNLMARRDNTQTPMQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW	750
dFdhAB5/1-1184	701	DEGNLMARRDNTQTPMQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW	750
dFdhAB6/1-1184	701	EEGNLMARRDNTQTPDQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW	750
dFdhAB7/1-1184	701	EEGNLMARRDNTQTPMQAGIGLFPNWAFCWPANRRILYNRASVDKTGKPW	750
dFdhAB8/1-1184	701	EEGNMMARRDNTQTPMQAKIGLFPNWAFCWPANRRILYNRASVDLTGKPW	750

Consensus

DEGNLMARRDNTQTPMQAKIGLFPNWSFCWPANRRILYNRASVDKTGKPW

FdhAB/1-1184	751	N P A K A V I E W K D G K W V G D V V D G G G D P G T K H P F I M Q T H G F G A L Y G P G R E E G P	800
dFdhAB1/1-1184	751	N P D K A V I E W K D G K W V G D V V D G G G D P G T K Y P F I M Q T H G F G A L Y A P G M E E G P	800
dFdhAB2/1-1184	751	N P D K A V I E W K D G K W V G D V V D G G G D P G T K Y P F I M Q T H G F G A L Y A P G M E E G P	800
dFdhAB3/1-1184	751	N P D K A V I E W K D G K W V G D V V D G G G D P G T K Y P F I M Q T H G F G A L Y A P G M E E G P	800
dFdhAB4/1-1184	751	N P E K A V I E W K D G K W V G D V V D G G G D P G T K Y P F I M Q T H G F G A L Y A P G M E E G P	800
dFdhAB5/1-1184	751	N P K K A V I E W K D G K W V G D V V D G G G D P G T K Y P F I M Q T H G F G A L Y A P G M E E G P	800
dFdhAB6/1-1184	751	N P K K A V I E W K D G K W V G D V V D G G G P P G T K Y P F I M T T H G F G Q L Y A P G M E E G P	800
dFdhAB7/1-1184	751	N P K K A V I E W K D G K W V G D V V D G G G P P G T K Y P F I M T T H G F G Q L F A P G M E E G P	800
dFdhAB8/1-1184	751	N P D K A V I E W K D G K W V G D V V D G G G P P G T K Y P F I M T T D G F G Q L F A P G M E E G P	800

Consensus

N P D K A V I E W K D G K W V G D V V D G G G D P G T K Y P F I M Q T H G F G A L Y A P G M E E G P

FdhAB/1-1184	801	F P E H Y E P L E C P V S K N P F S K Q L H N P V A F Q I E G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB1/1-1184	801	F P E H Y E P L E C P V K K N P F S K Q L H N P V A Y K F E G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB2/1-1184	801	F P E H Y E P L E C P V K K N P F S K Q L H N P V A Y K F D G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB3/1-1184	801	F P E H Y E P L E C P V K K N P F S K Q L H N P V A Y K F E G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB4/1-1184	801	F P E H Y E P L E C P V K K N P F S K Q L H N P T A Y K F E G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB5/1-1184	801	F P E H Y E P L E C P V K K N P F S K Q L H N P T A Y K F E G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB6/1-1184	801	F P E H Y E P L E C P V K K N P F S K Q L H N P V A Y K F E G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB7/1-1184	801	F P E H Y E P M E C P V K K N P F S K Q L H N P T A Y K F D G E K K A V A D P R Y P F I G T T Y R V	850
dFdhAB8/1-1184	801	F P E H Y E P M E C P V T K N P F S K Q L H N P T A Y K F E G E K K A V A D P R Y P F V A T T Y R V	850

Consensus

F P E H Y E P L E C P V K K N P F S K Q L H N P V A Y K F E G E K K A V A D P R Y P F I G T T Y R V

FdhAB/1-1184	851	T E H W Q T G L M T R R C A W L V E A E P Q I F C E I S K E L A K L R G I G N G D T V K V S S L R G	900
dFdhAB1/1-1184	851	T E H W Q T G L M T R H C A W L L E A Q P Q I F C E I S K E L A K K R G I K N G D T V K V E S L R G	900
dFdhAB2/1-1184	851	T E H W Q T G L M T R H C A W L L E A Q P Q I F C E I S K E L A K K R G I K N G D T V K V E S L R G	900
dFdhAB3/1-1184	851	T E H W Q T G L M T R H C E W L L E A Q P Q I F C E I S K E L A K E R G I K N G D T V K V E S L R G	900
dFdhAB4/1-1184	851	T E H W Q T G L M T R N C P W L L E A Q P Q I F C E I S K E L A K E R G I K N G D T V K V E S L R G	900
dFdhAB5/1-1184	851	T E H W Q T G L M T R N C P W L L E A Q P Q I F C E I S K E L A K E R G I K N G D K V K V E S L R G	900
dFdhAB6/1-1184	851	T E H W Q T G L M T R N C P W L L E A Q P Q I F C E I S K E L A K E R G I K N G D T V K V E S L R G	900
dFdhAB7/1-1184	851	T E H W Q T G L M T R N C P W L L E A Q P Q I F C E I S K E L A K E R G I K N G D T V K V E S L R G	900
dFdhAB8/1-1184	851	T E H W Q T G L M T R N C P W L L E A Q P Q I F C E I S K E L A K E R G I K N G D W V V V E S L R G	900

Consensus

T E H W Q T G L M T R N C P W L L E A Q P Q I F C E I S K E L A K E R G I K N G D T V K V E S L R G

FdhAB/1-1184	901	A L E A V A I V T E R I R P F K I E G V D V H M V G L P W H Y G W M V P K N G G D T A N L L T P S A	950
dFdhAB1/1-1184	901	A L E A V A I V T E R I R P F K I E G K D V H M V G L P W H Y G W M T P K D G G D T A N L L T P S V	950
dFdhAB2/1-1184	901	A L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W M T P K D G G D T A N L L T P S V	950
dFdhAB3/1-1184	901	A L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W L T P K D G G D T A N L L T P S V	950
dFdhAB4/1-1184	901	A L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W M A P K D G G D T A N L L T P S V	950
dFdhAB5/1-1184	901	A L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W M A P K D G G D T A N L L T P S V	950
dFdhAB6/1-1184	901	S L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W M M P K D G G D T A N L L T P S V	950
dFdhAB7/1-1184	901	S L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W L M P K D G G D T A N L L T P S V	950
dFdhAB8/1-1184	901	S L E A V A M V T E R I Q P F K I Q G K D V H Q V G L P W H F G W M A P K D G G D T A N L L T P S V	950

Consensus

A L E A V A M V T E R I R P F K I E G K D V H M V G L P W H Y G W M + P K D G G D T A N L L T P S V

FdhAB/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB1/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB2/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB3/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB4/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB5/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB6/1-1184	951	G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB7/1-1184	951	G D P N T G I P E T K A F M V N V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000
dFdhAB8/1-1184	951	G D P N T G I P E T K A F M V N V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E	1000

Consensus

G D P N T G I P E T K A F M V D V R K V G K M F F V D L S R C T A C R G C Q I A C K Q W K N L P A E
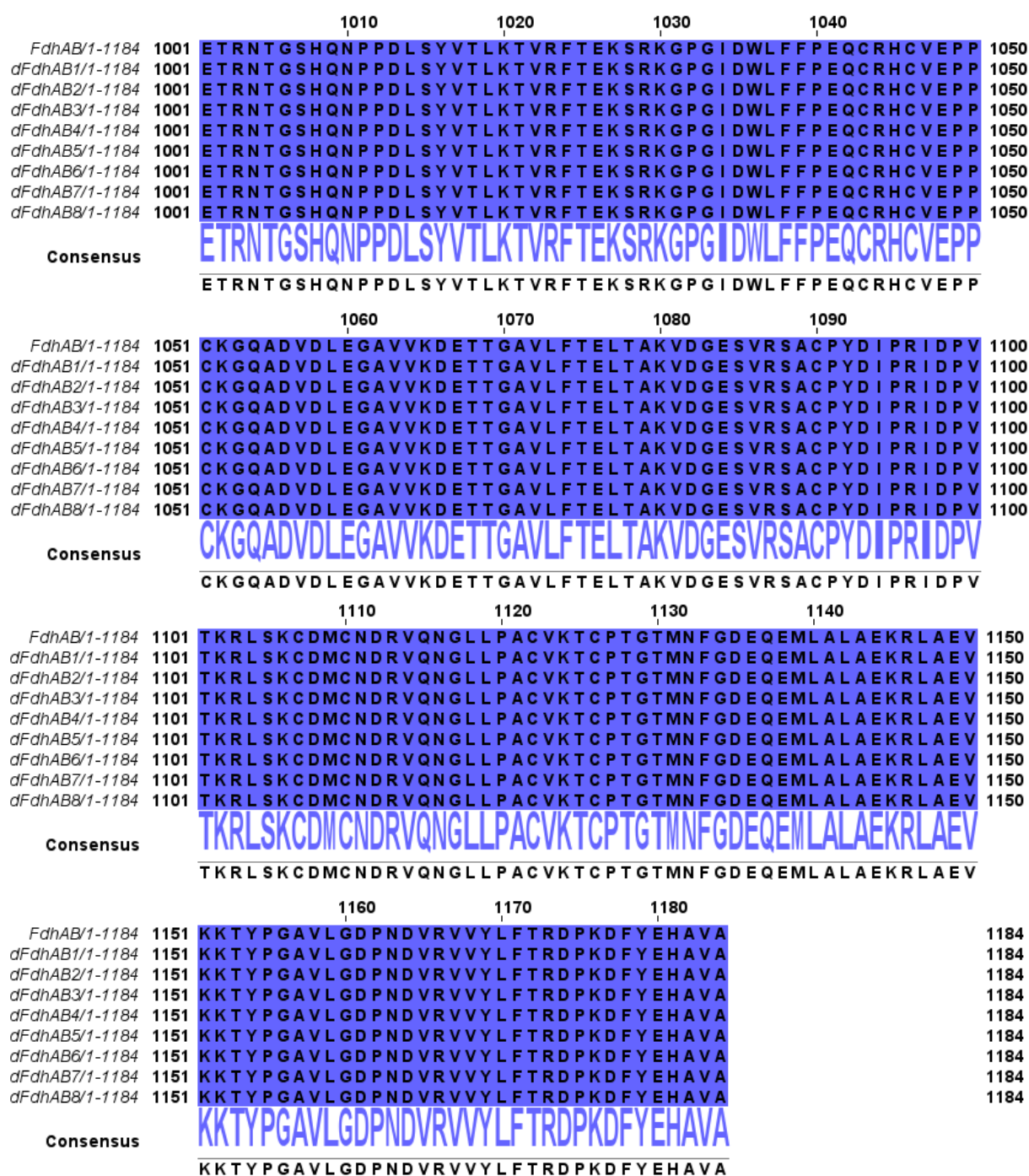
Figure 33: Sequence alignment of the simplified FdhAB and of the designs. The shades of blue represent the degree of conservation of each position in the sequences (darkest shade: > 80%; medium shade: > 60%; light shade: > 40%; white colour < 40%). The alignment was conducted using JALVIEW[128].

Table 14: List of mutations for the designed models dFdhA7 and dFdhAB7. Mutations that are present in both designed structures are coloured in dark blue, mutations that only exist in one design are coloured in black and mutations from a given position that differ between the two models are coloured in red.

Model	Mutations
dFdhA7	E1Q, Q6K, W7D, H27Y, T28V, A29D, Q33K, V37I, F60W, G63V, D66P, G69V, Q71K, L73M, F78Y, T81K, K83E, P84E, T86S, T92D, E93R, K96R, A104K, S105T, T107V, A111D, A112K, D114Q, L115T, T119V, F125L, D131S, A137V, G139T, N140A, I141M, L142M, S144A, P160A, P163A, N179I, S184A, R207E, S226A, V231I, P242A, L247M, K255N, F258H, T259K, L270F, G273R, E274P, S277D, A288P, A289K, M297T, A304E, A313T, K315Q, I321F, E329S, N332D, A338K, Q348L, Q349K, S376T, V397M, W424L, N428L, N432K, A435H, L440D, Q447K, S448T, M452K, V454A, V464I, A465V, Y467F, L471M, A478K, D482Q, R486K, I487M, G490D, R491K, K492P, L493D, T494A, N500F, E503D, D506Y, E509K, L513M, A515C, A525P, N530I, G535A, E538D, L540M, N542V, S551A, S552A, K555R, N560D, A562S, E563K, G565K, A584T, R595K, K598E, A601G, T603A, D612E, M613L, K616R, V617I, Y628F, A633L, I637M, A638D, E645H, T649H, M655I, T663V, Q669E, A684D, S690C, L694I, H695F, A696C, F699Y, A702E, K711L, E716M, R719N, S727A, F728W, V732A, K745L, A753K, E758K, D774P, H779L, Q784T, A790Q, G793A, R796M, S813T, I829F, E830H, R862W, C863V, A864P, V867L, A869L, E870M, I873M, C875V, K879P, L884E, G888K, T892W, K894I, A901S, L902I, E918H , V920K, M924Q, L927M, V935F, N938D, G939C
dFdhAB7	Q3L, Q6K T28V, A29D, S51A, F60W, G63V, D66P, Q71K F78Y, T81K, K83E, T86S, A104E, T107I, A111D, A112K, D114Q, L115T, D131S, A137V, G139T, N140A, I141M, L142M, S144A, P160A, P163A, N179I, S184A, R207K, K209R, S226A, V231I, P242A, L247M, K255N, F258H, T259K, G273R, E274P, K275D, S277D, A288P, A289K, K291R, M297K, A304E, A313T, K315Q, N322Q, E329S, N332D, A337S, A338K, Q348L, Q349K, K352E, N381I, V397I, W424L, N428L, A435H, L440D, A444S, Q447K, M452K, A478D, A479F, D482N, I487L, G490D, R491K, K492D, L493E, F497S, N500K, E503D, K504R, D506Y, A515C, N530I, M534L, G535S, N542V, S551A, N560D, A562S, E563K, G565K, R595Q, P597A, A601G, M613L, Y628F, A633L, A638D, E645H, T649H, M655I, T663V, Q669E, A684D, S690C, A696C, F699Y, D701E, A702E, K711N, E716M, R719G, S727A, V732A, A753K, D774P, H779Y, Q784T, A790Q, Y792F, G793A, R796M, L808M, S813K, V825T, F827Y, Q828K, I829F, E830D, R862N, A864P, V867L, E870Q, L884E, G888K, S896E, A901S, I907M, V920K, M934L, V935M, N938D, A950V, D966N