

# **Cryogenic electron microscopy studies on hydrogenases**

**Cristiana Furlan**

**Doctor of Philosophy**

**University of York  
Department of Chemistry**

**September 2023**

# Abstract

Hydrogenases are enzymes that produce or consume hydrogen ( $H_2$ ). They have central roles in microbial metabolism and have been of biotechnological interest. In this thesis two are studied: the putatively bifurcating hydrogenase from *Thermotoga maritima* (*TmHydABC*) and the oxygen-tolerant hydrogenase from *E. coli* (*Hyd-1*).

Structural studies of the apo-*TmHydABC* enzyme with cryogenic electron microscopy (cryo-EM) reveal its multimetric structure and how the active sites are connected. A deeper analysis of the single-particle data with symmetry expansion shows a bridge domain that changes conformation, which is most likely mechanistically relevant. Further studies to find the ferredoxin-bound and the holo-enzyme structure were attempted but were of limited success. These structural studies in combination with theoretical thermodynamic considerations are used to propose that *TmHydABC* only couples reactions kinetically and not thermodynamically.

A new purification strategy of *Hyd-1* was developed. A his-tag is swapped for a strep-tag and placed in a less destabilizing position, using recently developed CRISPR-cas9 genetic methods for homologous recombination. The resulting preparation is homogenous and contains all three subunits: HyaA, B, and C. This preparation was then subjected to cryo-EM analysis but a severe preferred orientation problem precluded a high-resolution 3D structure. Attempts to get around this problem are presented.

Finally, the hydrogenase field as a whole is critically evaluated in terms of real-world relevance and found to be of limited commercial potential, at least in terms of the hydrogen economy.

# **Declaration**

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for a degree or other qualification at this University or elsewhere. All sources are acknowledged as references.

# **Acknowledgements**

I would like to thank my supervisors Jamie Blaza and Alison Parkin for their support over the past four years. I am also grateful to Sam Hart and Johan Turkenburg for their cryo-EM technical support. I would like also to thank all the technicians who made our labs run smoothly. Finally but not least I would like to thank all my lab mates, in particular a special thank to Pooja Gupta and Morwan Osman for helping me at the very start with the biochemical techniques.

## Associated publications

1. Chris Furlan, Nipa Chongdar, Pooja Gupta, Wolfgang Lubitz, Hideaki Ogata, James N Blaza, James A Birrell (2022) Structural insight on the mechanism of an electron-bifurcating [FeFe] hydrogenase. *eLife* 11:e79361.
2. Solomon L. D. Wrathall, Barbara Procacci, Marius Horch, Emily Saxton, Chris Furlan, Julia Walton, Yvonne Rippers, James N. Blaza, Gregory M. Greetham, Michael Towrie, Anthony W. Parker, Jason Lynam, Alison Parkin, Neil T. Hunt (2022) Ultrafast 2D-IR spectroscopy of [NiFe] hydrogenase from *E. coli* reveals the role of the protein scaffold in controlling the active site environment. *Phys. Chem. Chem. Phys.*, 2022,24, 24767-24783

# Table of contents

<b>1. Introduction.....</b>	<b>1</b>
1.1 Bioenergetics .....	1
1.1.1 Bioenergetic principles.....	1
1.1.2 The importance of bioenergetics.....	2
1.2 Redox proteins .....	3
1.2.1 Biological function of redox proteins.....	3
1.2.2 Electron transfer principles.....	4
1.2.3 Redox centres types.....	8
1.3 Hydrogen and hydrogenases .....	10
1.3.1 The biological role of hydrogen and hydrogenases.....	10
1.3.2 Hydrogenases.....	11
1.4 Hydrogenases in the context of the hydrogen economy.....	12
1.4.1 Hydrogen economy.....	12
1.4.2 Biological hydrogen.....	13
1.5 Biological hydrogen production.....	15
1.5.1 Biohydrogen production: practical requirements .....	15
1.5.2 Biohydrogen production: methods.....	16
1.5.3 <i>In vitro</i> biohydrogen production: Cell-free enzymatic pathways .....	17
1.5.4 <i>In vitro</i> biohydrogen production: Enzymatic electrosynthesis .....	20
1.5.5 <i>In vivo</i> biohydrogen production: Direct microbial production .....	21
1.5.6 <i>In vivo</i> biohydrogen production: Microbial electrolysis cells.....	24
1.5.7 Biohydrogen production: prospects and conclusions.....	26
1.6 Hydrogen to energy: Fuel cells.....	27
1.6.1 Fuel cells and the drawbacks of platinum-based catalysts.....	27
1.6.2 Enzymatic fuel cells: can hydrogenases substitute platinum-based catalysts?.....	28
1.6.3 Hydrogenases-inspired catalysts.....	30
1.7 Why should we study hydrogenases?.....	30
1.7.1 Potential avenues.....	30
1.7.2 Hydrogenases in synthetic biology .....	31
1.7.3 Hydrogenases in pathogenic bacteria.....	32
1.8 Cryogenic electron microscopy.....	34
1.8.1 Structural biology techniques and Cryo-EM .....	34
1.8.2 Using electrons to visualize proteins: theoretical resolution .....	36

1.8.3	Cryo-EM sample preparation and vitrification .....	39
1.8.4	The electron microscope.....	41
1.8.5	Image formation in cryoEM: Electron interactions.....	43
1.8.6	Image formation in cryoEM: CTF function (pre-concepts) .....	46
1.8.7	Image formation in cryoEM: CTF function .....	52
1.8.8	Image processing: Single particle reconstruction .....	54
1.8.9	Image processing: Single particle reconstruction pipeline .....	55
1.8.10	Resolution assessment.....	60
1.8.11	Building a 3D atomic model.....	61
<b>2</b>	<b>Materials and methods.....</b>	<b>63</b>
2.1	<i>Thermotoga maritima</i> hydrogenase methods.....	63
2.1.1	Protein expression and purification.....	63
2.1.2	Inductively coupled plasma mass spectrometry.....	64
2.1.3	Grid preparation and imaging.....	64
2.1.4	Image processing .....	67
2.1.5	Model building and validation .....	70
2.2	Molecular biology: strep-tagging hydrogenase-1.....	71
2.2.1	Bacteria strains, plasmids, and media used.....	71
2.2.2	Glycerol stocks preparation from purchased plasmid-harboring strains and K12 MG1665 ...	72
2.2.3	Plasmids extraction .....	73
2.2.4	Making electrocompetent <i>E. coli</i> K12 MG1665 cells .....	74
2.2.5	Transformation of <i>E. coli</i> K12 MG1665 with pEcCas using electroporation .....	75
2.2.6	Making electrocompetent <i>E. coli</i> K12 MG1665 cells harboring pEcCas plasmid.....	76
2.2.7	Restriction digestion (optional) .....	76
2.2.8	Customization of the gRNA sequence in pEcgRNA plasmid.....	77
2.2.9	Transformation of competent <i>E. coli</i> DH5 $\alpha$ cells with customized pEcgRNA plasmid .	79
2.2.10	Sequencing of customized pEcgRNA plasmid .....	80
2.2.11	Transformation of electrocompetent <i>E. coli</i> K12 MG1665 cells harboring pEcCas with pEcgRNA(NTHyB) and template dsDNA.....	81
2.2.12	Colony PCR and sequencing of edited <i>E. coli</i> K12 MG1665 cells .....	82
2.2.13	Plasmid curing.....	83
2.3	Purification of His-tagged hydrogenase-1 from <i>E. coli</i> .....	84
2.3.1	Protein preparation and purification (initial method) .....	84
2.3.2	Protein purification without detergent .....	86
2.4	Purification of strep-tagged hydrogenase-1 from <i>E. coli</i> .....	87
2.4.1	Protein preparation and purification with DDM.....	87

2.4.2	LMNG buffer exchange .....	91
2.5	Analytical methods .....	92
2.5.1	Cell concentration .....	92
2.5.2	Determination of protein concentration .....	92
2.5.3	Determination of DNA concentration .....	93
2.5.4	Gel electrophoresis .....	94
2.5.5	UV-Vis spectroscopy .....	95
2.5.6	Activity assay .....	95
2.6	Cryo-EM methods (hydrogenase-1) .....	96
2.6.1	Grid preparation and imaging .....	96
2.6.2	Image processing .....	96
2.6.3	Preparation of PEGylated gold grids .....	97
<b>3</b>	<b><i>Thermotoga maritima</i> hydrogenase: Experimental structural studies and thermodynamic modelling .....</b>	<b>98</b>
3.1	Introduction .....	98
3.1.1	Iron-sulfur cluster redox reactions .....	98
3.1.2	[FeFe]-hydrogenases: active site reactions .....	99
3.1.3	Hydrogenases in <i>Thermotoga maritima</i> metabolism .....	101
3.1.4	Discovery of the bifurcating <i>T. maritima</i> [FeFe]-hydrogenase and its properties .....	103
3.1.5	Historical overview of proposed mechanisms for <i>T. maritima</i> hydrogenase .....	106
3.1.6	Homologous proteins to <i>T. maritima</i> HydABC .....	108
3.1.7	Definition of electron bifurcation/confurcation .....	109
3.1.8	Potential inversion in electron bifurcation/confurcation .....	110
3.1.9	Classical models of electron bifurcation/confurcation .....	112
3.1.10	Useful general thermodynamic concepts .....	114
3.1.11	Thermodynamics in the classical model of electron bifurcation .....	116
3.1.12	Electron bifurcation/confurcation in characterized enzymes .....	128
3.1.13	Examples of electron bifurcation in characterized enzymes .....	131
3.1.14	<i>T. maritima</i> hydrogenase does not fit the classical FBEB model .....	136
3.2	Aims of the chapter .....	137
3.3	Results .....	138
3.3.1	Screening the holoenzyme and apoenzyme preparation of <i>T. maritima</i> hydrogenase .....	138
3.3.2	Data collection and data processing pipeline: apoenzyme .....	139
3.3.3	2D classification results .....	140
3.3.4	3D map (D2 symmetry) and atomic model .....	141
3.3.5	Local resolution of the 3D map .....	142

3.3.6	The controversy over the zinc site .....	143
3.3.7	Structural comparison of HydABC with homologous proteins .....	144
3.3.8	Investigating mobile domains: ‘The bridges’ domains .....	146
3.3.9	The mobile domains revealed.....	148
3.3.10	The overall cofactor arrangement .....	149
3.3.11	Exploring the ferredoxin binding site.....	150
3.3.12	The function of the mobile C-terminus of HydA.....	153
3.3.13	Holoenzyme structure.....	154
3.4	Critical Discussion .....	156
3.4.1	Mechanism premises .....	156
3.4.2	Some observations on NiFe-HydABC SL hydrogenase from <i>A. mobile</i> .....	157
3.4.3	Generalized electron pathway model of the minimal functional unit.....	160
3.4.4	The mechanism of HydABC: Can it be bifurcating? .....	162
3.4.5	Lack of evidence that <i>TmHydABC</i> is bifurcating (and potential evidence against).....	165
3.4.6	Lack of evidence that HydABC from <i>A. woodii</i> , <i>T. kivui</i> , and <i>R. albus</i> are bifurcating.....	169
3.4.7	Potential evidence that HydABC from <i>M. thermoacetica</i> is bifurcating.....	171
3.4.8	Lack of evidence that NiFe-HydABC SL hydrogenase from <i>A. mobile</i> is bifurcating....	173
3.4.9	HydABC-type enzymes may not be bifurcating but only perform kinetic coupling....	174
3.4.10	Proposed kinetic coupling mechanism of HydABC .....	175
3.5	Conclusions and future work .....	177
<b>4</b>	<b>Tagging Hyd-1 using CRISPR-Cas9 genome editing .....</b>	<b>178</b>
4.1	Introduction .....	178
4.1.1	Genome editing techniques in <i>E. coli</i> .....	178
4.1.2	CRISPR/Cas-assisted bacterial genome editing.....	179
4.1.3	Lambda-Red homologous recombination .....	180
4.1.4	Templates used for homologous recombination.....	181
4.1.5	gRNA and Cas9 complex .....	182
4.1.6	pEcCas and pEcgRNA plasmids.....	183
4.1.7	Editing with the pEcCas and pEcgRNA plasmids .....	186
4.1.8	Natively expressing proteins from the genome.....	187
4.1.9	Affinity tag types .....	188
4.2	Aims of the chapter.....	189
4.3	Results.....	190
4.3.1	Analysis of the structure of hydrogenase-1 (Hyd-1) from <i>E. coli</i> .....	190
4.3.2	Choosing the affinity tag type and location .....	191
4.3.3	Analysis of the <i>Hya</i> operon .....	193

4.3.4	N-terminus Strep-tagged HyaB sequence.....	194
4.3.5	Comparison with the C-terminus His-tagged <i>HyaA</i> sequence.....	195
4.3.6	Design of the gRNA .....	195
4.3.7	Design of the template DNA .....	197
4.3.8	Transformation of <i>E. coli</i> DH5 $\alpha$ with pEcgRNA(NTHyaB).....	199
4.3.9	Sequencing the pEcgRNA plasmids.....	201
4.3.10	Transformation of <i>E. coli</i> K12 MG1665 with pEcgRNA(NTHyaB) and template dsDNA.....	203
4.3.11	Sequencing the edited <i>E. coli</i> K12 MG1665 cells .....	205
4.3.12	Plasmid curing.....	206
4.3.13	Estimating the probability of plasmid loss.....	208
4.4	Conclusions .....	213
<b>5</b>	<b>Cryo-EM studies on Hydrogenase-1 from <i>E. coli</i>.....</b>	<b>214</b>
5.1	Introduction .....	214
5.1.1	Hydrogenases from <i>E. coli</i> .....	214
5.1.2	Anaerobic regulation of the <i>Hya</i> operon .....	215
5.1.3	Structures of Hydrogenase-1 .....	216
5.1.4	The function of the cytochrome subunit .....	218
5.1.5	Mechanism of oxygen tolerance in Hyd-1 .....	219
5.1.6	Homologous hydrogenases to Hyd-1.....	220
5.1.7	Purification Strategies and limitations.....	222
5.1.8	Cryo-EM to study membrane proteins .....	223
5.2	Aims of the chapter.....	224
5.3	Results: His-Tagged preparation.....	225
5.3.1	Overview of the purification strategy .....	225
5.3.2	Affinity purification .....	226
5.3.3	Size exclusion chromatography (separation of oligomeric states).....	228
5.3.4	Cryo-EM analysis.....	230
5.3.5	Preferred orientation: use of the CHAPSO detergent .....	234
5.3.6	Discussion and summary: His-Tagged preparation.....	236
5.4	Results: Strep-tagged preparation.....	238
5.4.1	Overview of the purification strategy.....	238
5.4.2	Affinity purification .....	239
5.4.3	Protein yields and the effect of medium composition .....	241
5.4.4	Size exclusion chromatography .....	242
5.4.5	Activity assay.....	243
5.4.6	Heme spectroscopy.....	244

5.4.7	Cryo-EM: Hyd-1 after affinity purification with DDM .....	245
5.4.8	Cryo-EM: PEGylated gold grids .....	246
5.4.9	Cryo-EM: concentrated Hyd-1 after affinity purification with DDM .....	248
5.4.10	Cryo-EM: Hyd-1 with detergent removal and addition of LMNG .....	249
5.4.11	Cryo-EM: Anaerobic Hyd-1 grids with LMNG and GDN .....	251
5.5	Conclusions .....	252
<b>6.</b>	<b>Discussion and future perspectives .....</b>	<b>253</b>
<b>7.</b>	<b>Appendix .....</b>	<b>256</b>
<b>8.</b>	<b>References .....</b>	<b>271</b>

# 1. Introduction

## 1.1 Bioenergetics

### 1.1.1 Bioenergetic principles

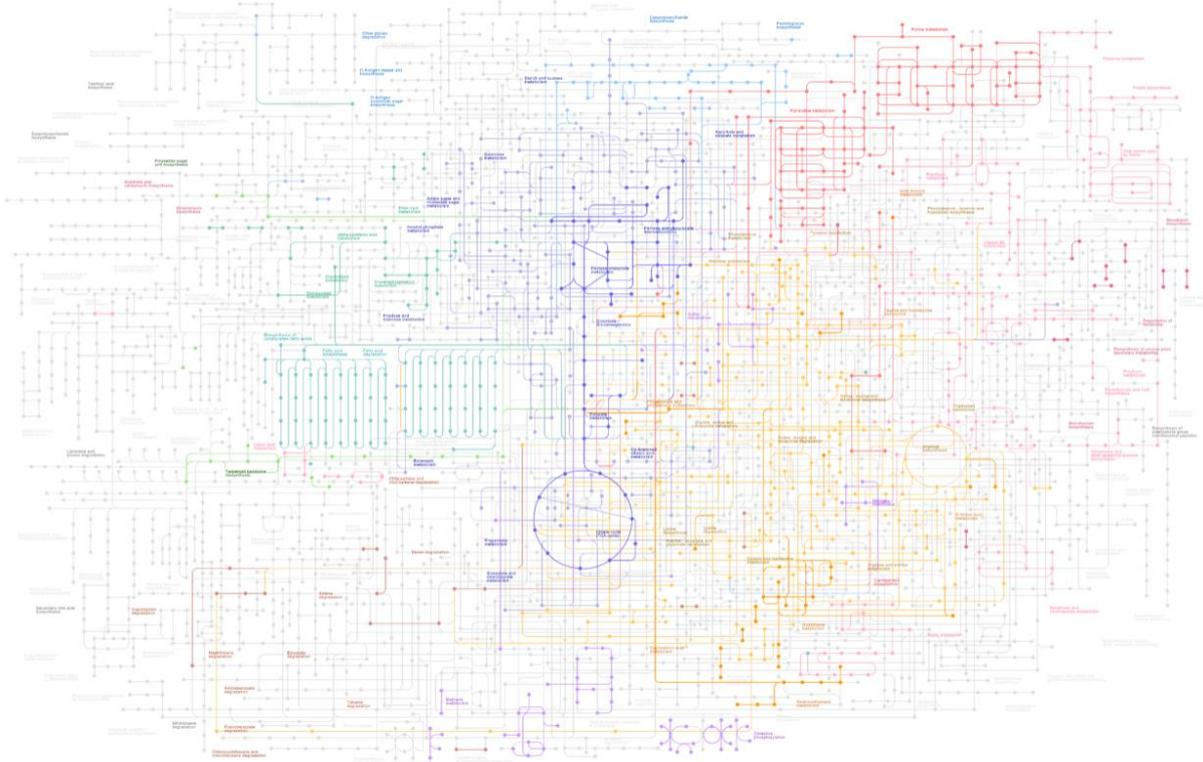
Living cells require a constant supply of energy in order to maintain their highly organized structure. Energy is required in every aspect of a living cell, including synthesizing complex molecules from simple precursors (e.g., to make proteins from amino acids), generating and maintaining differences in the concentrations of various substances inside and outside the cell, and movement (e.g., bacteria motility). All these processes have a common characteristic: a decrease in entropy (S) or “disorder”. This seems to contradict the second principle of thermodynamics which states that the entropy of an adiabatically isolated system never decreases, however, living cells exchange heat and matter with the environment and are therefore not isolated systems. As a consequence, every process and reaction happening in a cell must contribute to an overall increase in entropy when considered together with its environment therefore explaining why a constant supply of energy is required. This can be translated mathematically as:

$$\Delta S_{univ} = \Delta S_{cell} + \Delta S_{env} > 0 \quad (1.1)$$

Where  $\Delta S_{univ}$ ,  $\Delta S_{cell}$ ,  $\Delta S_{env}$ , are the entropy changes in the universe (by definition an adiabatically isolated system), in the cell, and in the environment surrounding the cell, respectively.

Thermodynamic modelling of cells is complex as they are open systems (they exchange heat and matter with the environment), however, one can imagine the cell as a combination of thousands of connected chemical reactions (Fig.1.1) and processes at constant pressure (as pressure is constant in most settings), each of them contributing to an increase in entropy in the universe while maintaining an overall “low-entropy” state within the cell, far from being in thermodynamic equilibrium with its environment. The energy supply needed to maintain this state is externally provided either by light (e.g. photosynthesis), the chemical potential of

surrounding “energy-rich” molecules (e.g. glucose, molecular hydrogen), or both. Bioenergetics is a branch of biochemistry and biophysics that studies these processes to understand how energy is transformed and used to support the functions of life, such as growth, reproduction, and metabolism.



**Figure 1.1.** Metabolic pathways diagram in *Escherichia coli* K-12 MG1655. Each node represents a chemical compound and two connected nodes represent a chemical reaction. Image taken from the KEGG (Kyoto Encyclopedia of Genes and Genomes) PATHWAY Database (1).

### 1.1.2 The importance of bioenergetics

Bioenergetics studies the energy flow and metabolism in cells and more complex multicellular organisms. As such, it has a fundamental role in multiple fields including:

- **Medicine and physiology:** bioenergetics is used to understand the energy metabolism of cells and how it relates to diseases, such as cancer and diabetes.
- **Biomedical research:** bioenergetics is used to study the fundamental processes of life, such as cell growth, differentiation, and death.
- **Biotechnology:** bioenergetics is used to improve the efficiency of industrial processes, such as fermentation and bioprocessing.

- Environmental science: bioenergetics is used to study the energy flow in ecosystems and how it relates to the environment.
- Agriculture: bioenergetics is used to improve crop yields and animal production.
- Food science: bioenergetics is used to understand how food is metabolized and how it affects human health.

In particular, this thesis focuses on bacteria bioenergetics connected to hydrogen metabolism.

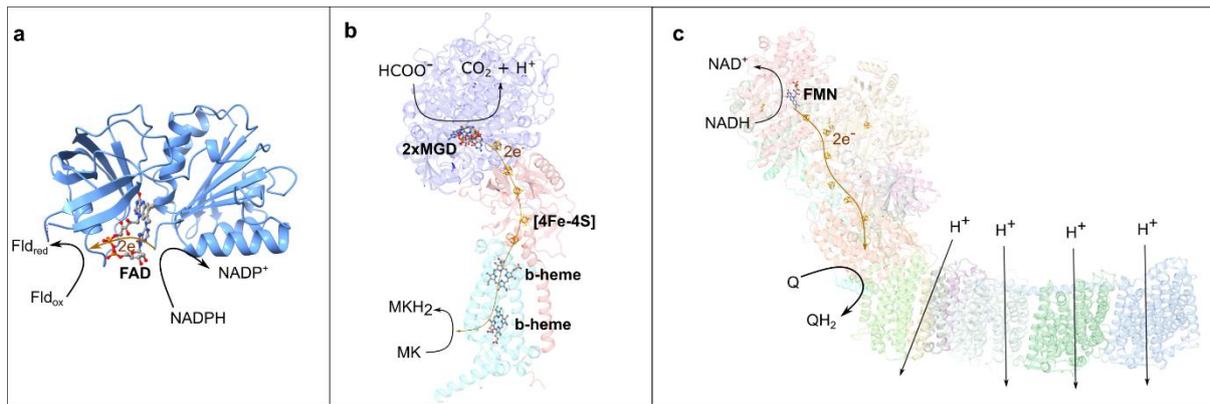
## 1.2 Redox proteins

### 1.2.1 Biological function of redox proteins

Bioenergetic redox reactions involve the movement of electrons from low-potential donors to high-potential acceptors. These electron transfer reactions normally involve proteins that contain redox-active cofactors. Redox proteins can be enzymes (e.g. catalases, hydrogenases, cytochrome *bc*<sub>1</sub> complex) or act as electron transfer carriers (e.g. ferredoxins, rubredoxins, cupredoxins, flavodoxins, cytochrome *c*). In addition, specific kinds of redox proteins containing stacked heme filaments are used in electrically conductive appendages, called bacterial nanowires, for long-range extracellular electron transfer (LET) (2).

Redox proteins play a crucial role in cells, from the energy transduction of photosynthesis and respiration to many fundamental metabolic and biosynthetic reactions. In particular, the processes of photosynthesis, aerobic respiration, and anaerobic fermentation, all dependent on redox enzymes and proteins, are at the heart of the energy supply of cells.

Redox enzymes can vary from simple with one single redox cofactor (e.g. flavodoxin reductase, Fig.1.2.a), to significantly complicated with more than one active site and a network of redox centres that connect them in analogy to a conductive wire (e.g. formate dehydrogenase, Fig.1.2.b, complex I, Fig.1.2.c).



**Figure 1.2.** **a.** 3D structure of flavodoxin reductase from *E. coli*, a FAD-containing oxidoreductase that transports electrons from NADPH to flavodoxin or ferredoxin. PDBID: 1FDR. (3) **b.** Formate dehydrogenase from *E. coli*, a membrane-bound oxidoreductase containing hemes and iron-sulfur clusters catalysing the oxidation of formate into  $\text{CO}_2$  and the reduction of quinones, PDBID: 1KQF (4). **c.** Complex I from *Thermus thermophilus*, an enzyme part of the respiratory electron transport chain that transfers electrons from the NADH to reduce ubiquinone to ubiquinol. The energy from the redox reaction is used to pump four protons. PDBID: 4HEA (5).

Redox enzymes can also be arranged in large multienzyme complexes, where several copies of one or several enzymes are packed into one assembly (e.g. pyruvate dehydrogenase complex) or be part of electron transfer chains (ETC) in which, according to the chemiosmotic hypothesis (Mitchell, 1961 (6)), light or a chemical potential is converted into an electrochemical gradient by translocating chemical charges across a membrane that in turns drives the production of ATP, an energy-rich molecule that provides energy to drive many processes in living cells.

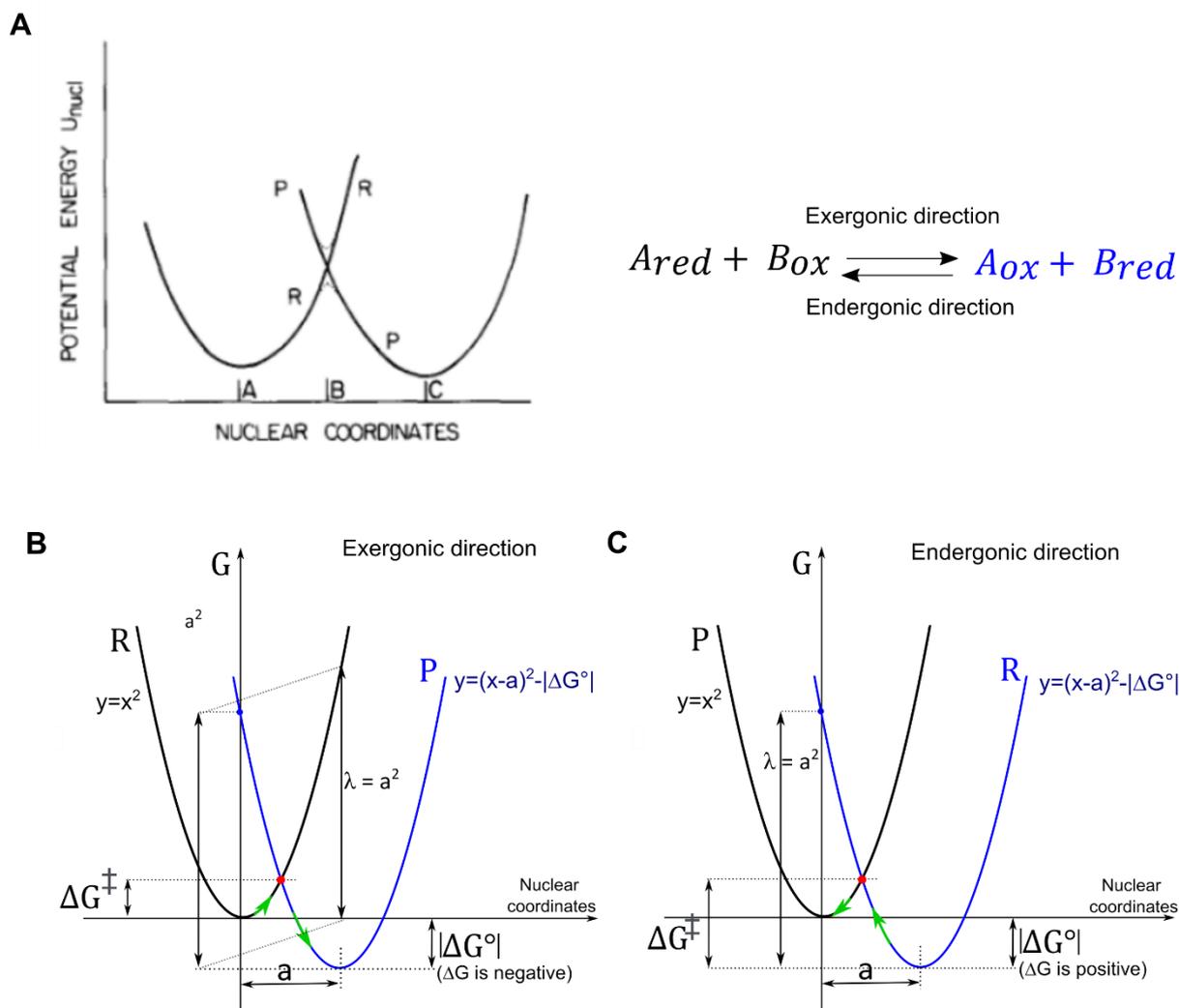
## 1.2.2 Electron transfer principles

Redox proteins contain one or more cofactors capable of accepting or donating electrons, each acting as a redox centre. When redox centres are kept separated from each other by the protein scaffold (imagine redox centres as raisins in a cake) the electrons must be transferred without direct contact; in this case, electrons have to ‘hop’ between redox centres, and therefore no chemical bond is made or broken. This biologically very common electron transfer process is substantially different from that of a substrate molecule transferring electrons by direct contact with a cofactor in the active site (e.g. NADH transfers two electrons to the cofactor FMN through transferring a hydride ion,  $\text{H}^-$ ) and requires specific kinetic modelling.

To describe and model the transfer of electrons between redox centres that are kept separated by the protein scaffold, an adapted experimental version of Marcus' theory is typically used. The classical Marcus theory models outer sphere one-electron transfer reactions between two redox species (an electron jumps from a reduced donor to an oxidized acceptor,  $A_{red} + B_{ox} \rightleftharpoons A_{ox} + B_{red}$ ) according to solvent polarization. Marcus' approach was to use a parabolic model as a simplified representation of reactant and product energy surface (Fig.1.3) (7). The values for reorganization energy,  $\lambda$ , (i.e. energy required to reorganize nuclei without electron transfer occurring) and change in Gibbs free energy between reactants and products ( $\Delta G^\circ$ ) represented in the graphical model are used to calculate the Gibbs free energy of activation ( $\Delta G^\ddagger$ ) (Fig.1.3.b,c):

$$\Delta G^\ddagger = \frac{(\Delta G^\circ + \lambda)^2}{4\lambda}, \quad \Delta G^\circ \text{ can be positive (endergonic direction) or negative (exergonic direction)} \quad (1.2)$$

The value of  $\Delta G^\circ$  can be calculated from the standard Gibbs energies of formation of reactant and products ( $\Delta G^\circ = G_P - G_R = (\Delta G_f^\circ(A_{ox}) + \Delta G_f^\circ(B_{red})) - (\Delta G_f^\circ(A_{red}) + \Delta G_f^\circ(B_{ox}))$ ) or equivalently from the standard reduction potentials ( $\Delta G^\circ = -nF \cdot \Delta E^\circ$ ). It is important to note that  $\Delta G^\circ$  is used instead of the general  $\Delta G$  of the reaction which depends on concentrations through the reaction quotient  $Q_r$  ( $\Delta G = \Delta G^\circ + RT \ln Q_r$ ) and therefore varies as the reaction reaches equilibrium.



**Figure 1.3 a.** Original image (from Marcus' work) of potential energy surfaces of the reactant (reduced donor and oxidized acceptor) and product (oxidized donor and reduced acceptor) represented as intersecting, parabolic, simple harmonic oscillator wells, with the reaction coordinate being a generalized one of nuclear vibration/reorganization (8). Image taken from: (7). **b.** and **c.** Representation of Gibbs free energies of reactants (*R*) and products (*P*) according to Marcus's model in the exergonic and endergonic direction, respectively. The intersection point of the parabolas (red dot) can be calculated and from its y-axis position and the following formula can be obtained:  $\Delta G^\ddagger = \frac{(\Delta G^\circ + a^2)^2}{4a^2} = \frac{(\Delta G^\circ + \lambda)^2}{4\lambda}$ . Axes are conveniently positioned.

From the combination of classical transition theory (note similarity with Arrhenius and Eyring's equations) with the obtained  $\Delta G^\ddagger$  (Formula 1.2) it is possible to calculate a rate constant for an electron-transfer reaction as:

$$k_{ET} = A \cdot e^{\frac{-\Delta G^\ddagger}{RT}} = A \cdot e^{\frac{-(\Delta G^\circ + \lambda)^2}{4\lambda RT}} \quad (1.3)$$

Where  $k_{ET}$  is the electron-transfer reaction rate, *R* is the universal gas constant, *T* is the temperature, and *A* is a constant (note: this is a simplified derivation in which I left the constant *A* generic). At equilibrium ( $\Delta G = 0$ ) the concentration of reactants and products is such that

the number of electrons travelling in the endergonic direction matches that of the faster electrons (higher  $k_{ET}$ ) travelling in the exergonic direction (Fig.1.3).

Using Fermi's golden rule in combination with the exponential decay of the wave function with distance, one can express the pre-exponential factor ( $A$ ) as a function of the donor-acceptor distance ( $r$ ) (9). The electron transfer reaction rate can therefore be expressed also as:

$$k_{ET} = C e^{-\beta r} e^{-\frac{(\Delta G^\circ + \lambda)^2}{4\lambda RT}} \quad (1.4)$$

Where  $\beta$  is a constant that describes the contribution made by the intervening medium in propagating the wavefunction and  $C$  is a constant (*kept generic to simplify the discussion*).

Further combining Equation 1.4 with the relation between the electrochemical potential difference ( $\Delta E$ ) and Gibbs free Energy ( $\Delta G = -nF\Delta E$ ):

$$k_{ET} = C e^{-\beta r} e^{-\frac{(-nF\Delta E^\circ + \lambda)^2}{4\lambda RT}} \quad (1.5)$$

From Formulas 1.4 and 1.5, three main observations can be drawn:

- 1) The electron transfer rate depends on the distance between the donor and acceptor, the nature of the intervening space, the energy required to reorganize the atomic nuclei after the transfer has occurred, the  $\Delta E^\circ$  between donor and acceptor, and temperature.
- 2) The electron transfer rate decays exponentially with the distance between the donor and the acceptor.
- 3) As the driving force increases ( $\Delta G^\circ$  more negative), the reaction rate rises to a maximum when  $\lambda = -\Delta G^\circ$ , but then unexpectedly falls off again (the inverted region).

Formula 1.4 can also be expressed in an equivalent way (standard logarithmic manipulations):

$$\log_{10}(k_{ET}) = \log_{10}C - \frac{\beta}{2.303}r - \gamma \frac{(\Delta G^\circ + \lambda)^2}{\lambda} \quad (1.6)$$

$$\text{With } \gamma = \frac{1}{2.303 \cdot 4RT} = 4.38 \cdot 10^{-5} \frac{\text{mol}}{\text{J}} = 4.22 \frac{\text{molecule}}{\text{eV}}$$

The above equation was used by Dutton and co-workers to fit the experimental data of different physiological and non-physiological reactions in native or modified proteins (10–12). From this empirical approach, it was found that for the majority of intraprotein electron transfer, a simplified Marcus-type formula could be used (Dutton's ruler):

$$\log_{10}(k_{ET}) = 15 - 0.6r - 3.1 \frac{(\Delta G^\circ + \lambda)^2}{\lambda} \quad (1.7)$$

Where  $r$ , the edge-to-edge distance between redox centres in the protein, is expressed in Å, and  $\Delta G^\circ$  and  $\lambda$  in eV. The resulting  $k_{ET}$  is in  $s^{-1}$ . This formula, contrary to the original Dutton's recommendation and later papers (8, 11), can be used for both exergonic and endergonic electron transfer (paying attention to the sign of  $\Delta G^\circ$ ) as the physically unjustified asymmetry of the original Dutton's formula (11) dealing with endergonic electron transfer was the result of a trivial mathematical mistake (13).

In Formula 1.7, the factor  $\gamma = 3.1$  was found to better fit the experimental data compared to the factor  $\gamma = 4.22$  obtained through the classical Marcus' theory (8). The value of  $\gamma$  was also theoretically obtained using the Hopfield approximation (13, 14) in which electron transfer by tunnelling was modeled. In fact, the observation of temperature-independent reactions at low temperatures in protein redox centres indicated that a tunnelling pathway was introduced by quantum mechanical contributions (15).

In conclusion, for biological electron transfer, which happens through electron tunnelling, the electron transfer rate is mainly dependent on the distance between redox centres as the packing density of proteins (influences  $\beta$ ) and the reorganization energy of most electron transfer reactions are quite constant (8, 11). The value of  $\lambda \sim 0.7$  eV (equivalent to  $0.7 \cdot 96.5 = 67.6$  kJ/mol) seemed to be an adequate generic value (8, 11) and therefore for  $\Delta G^\circ$  significantly smaller than  $\lambda$  (most biological electron transfer reactions have little driving force) an even more simplified formula can be used (derived from Formula 1.7 using  $\Delta G^\circ = 0$  and  $\lambda \sim 0.7$ ):

$$\log_{10}(k_{ET}) = 12.8 - 0.6r \quad (1.8)$$

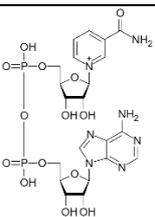
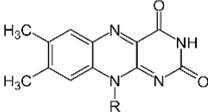
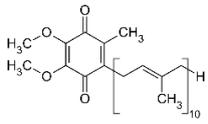
Dutton's Ruler and analysis of existing protein structures show that almost all productive reactions have redox centres at less than 14 Å (11). As a consequence, redox centres that fall outside of  $\sim 14$  Å limit are not considered "electrically connected" as their electron transfer tunnelling rate will not be able to support the overall turnover of the enzyme.

### 1.2.3 Redox centres types

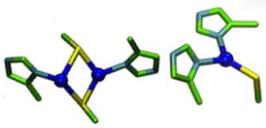
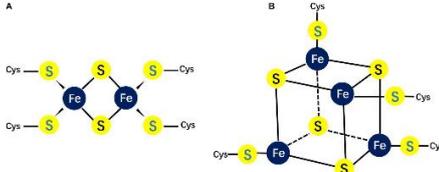
There are many types of organic, inorganic, and hybrid redox centres in biochemistry ranging from simple, such as the single metal copper ion in cupredoxins, to complex, such as the iron-sulfur clusters assemblies (H-clusters) found in the active site of [Fe-Fe] hydrogenases.

Redox centres can be substrates (e.g. NADH, quinones), or cofactors tightly bound to redox proteins as active sites (e.g. FMN in complex I) or as connectors between active sites (e.g. iron-sulfur clusters and hemes in Hydrogenase-1 from *E. coli*). An overview of the most common organic, inorganic, and hybrid redox centres is given in Tables 1.1, 1.2, and 1.3, respectively.

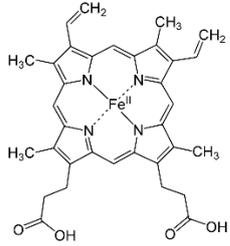
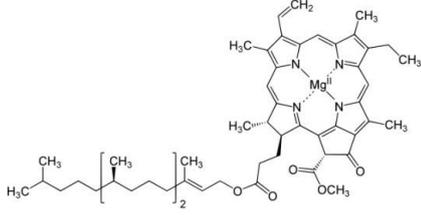
**Table 1.1** Common organic redox centres found in cells with their structure and examples of proteins containing them as cofactors or using them as substrates.

Organic redox centres		
Name	Molecular structure	Examples
Nicotinamide adenine dinucleotide (NADH)		<ul style="list-style-type: none"> <li>- Substrate in complex I</li> <li>- Substrate in <i>Thermotoga maritima</i> hydrogenase</li> <li>- Substrate in malate dehydrogenase (citric acid cycle)</li> </ul>
Flavins (FAD, FMN)		<ul style="list-style-type: none"> <li>- Cofactor in complex I and II.</li> <li>- Cofactor in <i>Thermotoga maritima</i> hydrogenase</li> <li>- Cofactor in glucose oxidase (GOX)</li> </ul>
Quinones (ubiquinone, menaquinone, plastoquinone, etc.)	 A class of multiple molecules derived from aromatic compounds. Here ubiquinone is shown.	<ul style="list-style-type: none"> <li>- Substrate in complex I and II</li> <li>- Substrate in hydrogenase-1 (<i>E. coli</i>)</li> <li>- Substrate in NAD(P)H quinone oxidoreductases (NQO1)</li> </ul>

**Table 1.2** Common inorganic redox centres found in cells with their structure and examples of proteins containing them as cofactors or using them as substrates. Images adapted from: (16, 17).

Inorganic redox centres		
Name	Molecular structure	Examples
Copper ions (mononuclear or dinuclear)		<ul style="list-style-type: none"> <li>- Dinuclear copper centre of cytochrome c oxidase.</li> <li>- Copper-containing nitrite reductases</li> <li>- Mononuclear copper in cupredoxins</li> </ul>
Iron-sulfur clusters		<ul style="list-style-type: none"> <li>- Cofactors in complex I and II.</li> <li>- Cofactors in <i>Thermotoga maritima</i> hydrogenase</li> <li>- Cofactors in hydrogenase-1 (<i>E. coli</i>)</li> </ul>

**Table 1.3** Common hybrid redox centres found in cells with their structure and examples of proteins containing them as cofactors or using them as substrates.

Hybrid (organic+inorganic) redox centres		
Name	Molecular structure	Examples
Hemes (a-heme, b-heme, c-heme)	 <p>A class of multiple molecules with an iron ion coordinated to a porphyrin. Here b-heme is shown.</p>	<ul style="list-style-type: none"> <li>- Cofactor in hydrogenase-1 cytochrome subunit (<i>E. coli</i>)</li> <li>- Cofactor in formate dehydrogenase cytochrome subunit</li> <li>- Stacks of hemes in microbial nanowire proteins</li> </ul>
Chlorophylls and Bacteriochlorophylls	 <p>A class of multiple molecules with a magnesium ion coordinated to a porphyrin. Here chlorophyll a is shown.</p>	<ul style="list-style-type: none"> <li>- Cofactor in photosystem I and II</li> </ul>

## 1.3 Hydrogen and hydrogenases

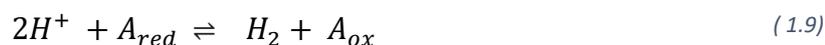
### 1.3.1 The biological role of hydrogen and hydrogenases

Molecular dihydrogen is a simple molecule of two protons and two electrons ( $H_2$ ) which has a high energy density of combustion of 142 kJ/g (Higher Heating Value, HHV) (18). It is thought that initially, the earth's atmosphere was rich in hydrogen and that bacteria evolved to harness that energy using enzymes capable of catalyzing its oxidation (hydrogenases). Over time the atmospheric molecular hydrogen concentration decreased to the current very low levels (0.00005% v/v). The traditional view is that this low concentration of  $H_2$  caused  $H_2$ -utilising microorganisms to remain confined in areas with higher  $H_2$  concentrations, such as hydrothermal vents, or to evolve in utilizing other sources of energy while repurposing hydrogenases for other aspects of their metabolism, such as to dispose of excess reducing

equivalents during fermentation (19). However, recently, it has become clear that some organisms can use specialised hydrogenases to capture H<sub>2</sub> from the low levels found in the atmosphere which is used to sustain electron input into the respiratory chain during energy starvation (20, 21). These specialized hydrogenases are unusual for their high affinity, oxygen insensitivity, and thermostability (22) and were therefore classified as a separate group (group 5 [NiFe]-hydrogenase in table A.1 in appendix). Recently, the structural basis of these unusual properties has been investigated for the high-affinity hydrogenase Huc from *Mycobacterium smegmatis* indicating a potentially major role of gas channels in excluding oxygen while selectively capturing H<sub>2</sub>, however at the expense of catalytical rate (low turnover number ~ 7 s<sup>-1</sup>) (21).

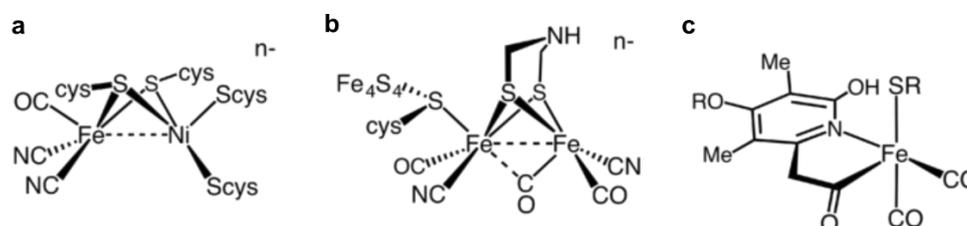
### 1.3.2 Hydrogenases

Hydrogenases catalyse the reversible oxidation of molecular hydrogen (H<sub>2</sub>) using another substrate as an electron acceptor or donor (here A):



Hydrogenases have active sites (which are also redox centres) with different electrochemical potentials depending on the active site's chemical nature and interaction with the surrounding amino acids. Under physiological conditions, some hydrogenases operate as hydrogen uptakers while others as hydrogen producers.

Depending on the active site's chemical nature, hydrogenases, are classified into three classes: [NiFe], [FeFe] and [Fe]-hydrogenases (Fig.1.3).



**Figure 1.3.** a. [NiFe] active site b. [FeFe] active site c. [Fe] active site. Image from (23).

[NiFe]-hydrogenases, which contain a nickel-iron centre, are found in both facultative and obligated anaerobic organisms. In physiological conditions, most [NiFe]-hydrogenases are H<sub>2</sub>

uptake enzymes and oxygen-sensitive. Remarkably, few [NiFe]-hydrogenases are oxygen tolerant (e.g. Hyd-1 from *E. coli*) making them easier to study and work with.

[FeFe]-hydrogenases, which contain an iron-iron centre, are found in green algae and a few prokaryotes (24). These enzymes are highly sensitive to oxygen which irreversibly inactivates them. In physiological conditions, [FeFe]-hydrogenases only catalyze H<sub>2</sub> production.

[Fe]-hydrogenases contain a single iron centre and are only found in hydrogenotrophic methanogenic archaea that use hydrogen for CO<sub>2</sub> reduction to methane (25).

These three main classes are further divided into subgroups (Table A.1 in appendix) (22, 26, 27) based on both gene phylogeny and enzyme function, with a key consideration being whether the enzyme is an H<sub>2</sub>-uptaker or H<sub>2</sub>-producer (26).

Hydrogenases are thought to be a promising system in contributing to the hydrogen economy with some demonstrated to be capable of converting hydrogen at rates approaching those reported for synthetic Pt-containing catalysis in hydrogen fuel cells (electrochemical cell that converts the chemical energy of hydrogen and oxygen into electricity) (28). Finding alternatives to platinum-based catalysts in hydrogen production or utilization can be very valuable as platinum is expensive and scarce, however, all other practical and economical aspects must be taken into consideration when proposing alternative solutions. Whether or not biological systems can or cannot contribute to this endeavour will be soon discussed.

## 1.4 Hydrogenases in the context of the hydrogen economy

### 1.4.1 Hydrogen economy

Molecular hydrogen (H<sub>2</sub>) has an energy density that is two to three times that of fossil fuels and does not produce CO<sub>2</sub> when combusted, making it a promising green alternative. However, so far, its application has been limited due to a multitude of technical and economic problems. Hydrogen gas is very scarce on earth and therefore it needs to be extracted from other substances, such as water, oil, coal, or methane by utilizing energy; **this makes H<sub>2</sub> an energy carrier and not a fuel**. Extraction from methane (steam methane reforming) is the cheapest and most common but it produces large amounts of CO<sub>2</sub>. It is, however, fundamental for many industrial processes (~70 Mt/year of H<sub>2</sub> consumed globally), especially as a reagent for

ammonia-based fertilizers (29). Only hydrogen produced from water electrolysis using electricity from renewable energy sources results in zero to minimal net CO<sub>2</sub> emissions. This “green” hydrogen, however, is still too expensive to produce and potentially wasteful as we are converting one readily usable form of energy (electricity) into another (chemical potential) losing a significant portion in the process ( $\geq 30\%$  (30)). Nevertheless, with increased production of energy from renewable sources (wind turbines, solar panels, etc.) and an improvement in technologies, “green” hydrogen may become more competitive and realistically be used to aid in the decarbonisation of the chemical industry by substituting the “dirty” hydrogen produced from methane. Other sectors that require high energy-dense combustible materials, such as aviation transport or steel and cement production, are potential new avenues, however, substantial development and changes in engines and production lines are needed before they can switch to hydrogen. Instead, the use of hydrogen in residential settings seems unlikely as constructions can be made carbon neutral at a much lower cost and with existing technologies through electrification (e.g. electric heat pump coupled with the energy-efficient underfloor heating) and similarly also its use in road transportation seems unlikely given the fast development of electric cars and the improvements in battery capacities.

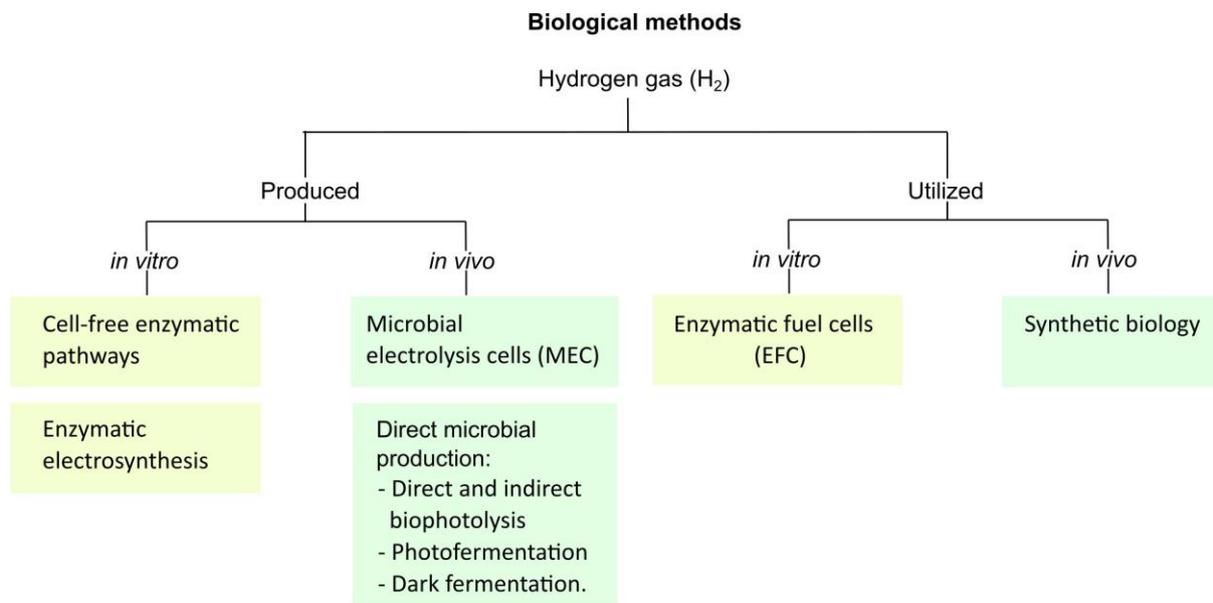
Generally, to improve hydrogen competitiveness, it is fundamental to reduce its production costs, especially that of the “green” hydrogen, and resolve the technical difficulties regarding storage and safety (H<sub>2</sub> is explosive!). Given all these challenges and the fast-paced improvement and development of competitive electric-based technologies, it seems that the outlook of the hydrogen economy may have been hyperinflated, nevertheless, hydrogen is and will remain an important commodity. In all this, it is worth asking whether biological organisms and enzymes may end up playing a role.

### **1.4.2 Biological hydrogen**

Biological systems such as enzymes and cells can be used to produce or utilize hydrogen gas (Fig.1.4). For the production of H<sub>2</sub> there are both *in vitro* and *in vivo* methods. *In vitro* methods include cell-free enzymatic pathways and enzymatic electrosynthesis. *In vivo* methods include microbial electrolysis cells and direct production through direct and indirect biophotolysis, photofermentation, and dark fermentation.

Reported biological methods of H<sub>2</sub> utilization include enzymatic fuel cells (EFCs) which are capable of generating electricity from H<sub>2</sub> oxidation. The corresponding *in vivo* technology, known as microbial fuel cells (MFCs), has not found application in the oxidation of hydrogen to produce electricity; this is because living cells require biomass as a fuel and hydrogen on its own is only compatible with the use of pure enzymes. However, H<sub>2</sub> may be supplied to cell cultures to produce useful chemicals; for example, *Raistonia eutropha* was shown to fix CO<sub>2</sub> into alcohols by using H<sub>2</sub> (31, 32).

All these methods will be reviewed in the following subchapters which will provide an insight into their strength and limitations with a special focus on practical and economical considerations. It will be also clear that hydrogenases play a key role in all these processes which can benefit from improvements in hydrogenase stability and production methods. However, the majority of research seems to have overlooked practical and economic aspects, which, when taken into consideration, reveal the current and future unattractiveness of most biological methods for hydrogen production/utilization.



**Figure 1.4.** Biohydrogen production (left) and hydrogen utilization methods using biological systems (right).

## 1.5 Biological hydrogen production

### 1.5.1 Biohydrogen production: practical requirements

Plenty of organisms are capable of producing hydrogen, however, biological systems have not yet found any practical use in hydrogen production that is thought to be compatible with economic, regulatory, and scalability requirements. Any technology or method used to make biologically derived hydrogen (or biohydrogen) needs to satisfy the following requirements in order to be considered viable and investible:

- 1) The overall **production cost** of bio-hydrogen needs to be comparable to or lower than that of current commercial methods or at least be expected to be in the future. Currently, steam methane reforming (SMR) is the most economically convenient; hydrogen produced with SMR has an average selling price of ~1.5 USD/kg H<sub>2</sub> with production costs that vary from 0.5 to 2.2 USD/kg H<sub>2</sub> depending on location and the presence of carbon capture technologies (33, 34). Hydrogen produced with water hydrolysis using grid electricity is considered the most promising alternative to SMR, however, currently, its production costs remain too high resulting in an average selling price that is 3 to 10 times higher than that of SMR (35) with this difference mainly driven by the electricity costs. If in the future fossil fuels are banned, highly taxed, or exhausted, then the production cost of biohydrogen needs to be comparable to or lower than that of hydrolysis with electricity obtained from renewable energies. Failure in satisfying this requirement implies an unsellable product.
- 2) The technology needs to be **compatible with the current infrastructure**. For example, if bacteria are used to process wastewater in order to produce biohydrogen, then the characteristics of the wastewater infrastructure must be taken into consideration in the economic and feasibility analysis.
- 3) The technology needs to be **scalable**.
- 4) The method of production needs to **comply with regulations**. For example, if genetically modified bacteria are used, then methods of containment are needed which may be unfeasible or increase excessively the production costs.

An overview of the biohydrogen production methods is followed with a particular focus on these practical considerations. This will allow us to draw more realistic expectations on the future of biohydrogen and the contribution of hydrogenases. In particular, being aware of these limitations is fundamental for contextualising the real-world applicability of hydrogenase research.

### 1.5.2 Biohydrogen production: methods

Hydrogen gas is an energy carrier and therefore a source of economically viable energy is needed for its production. Biological systems such as whole organisms or isolated enzymes require a biologically compatible source of energy and biologically compatible substrates. This could be in the form of:

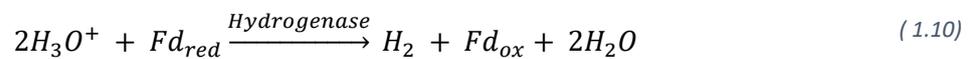
1. **chemical potential** directly provided by biologically compatible substrates to purified hydrogen-producing enzymes, or in more complex designs, provided to an enzyme cocktail that performs an in-vitro synthetic biological pathway aimed at hydrogen production (36, 37).
2. **electrical energy** supplied to hydrogen-producing enzymes that are absorbed in an electrode, for example, by directly connecting an electrical power source to an electrode immersed in a buffer solution at appropriate pH where hydrogenases are immobilized. This process is called **enzymatic electrosynthesis** (38, 39).
3. **biomass**, typically obtained from cultivated crops, food industry byproducts, food waste, domestic sewage, or livestock waste. These can be “fed” to bacteria or algae optimized for hydrogen production, or, as an alternative, that are capable of diverting electrons to an anode resulting in electrons transferred in the cathode where protons are reduced (**microbial electrolysis cells**). These methods can be coupled with light in photofermentation.
4. **light**, photosynthetic bacteria and algae can convert light into hydrogen gas through direct and indirect photolysis.

Options 1 and 2 are both cell-free and therefore can be considered *in-vitro* techniques, while options 3 and 4 are *in-vivo* methods that require the maintenance of living cells. There are

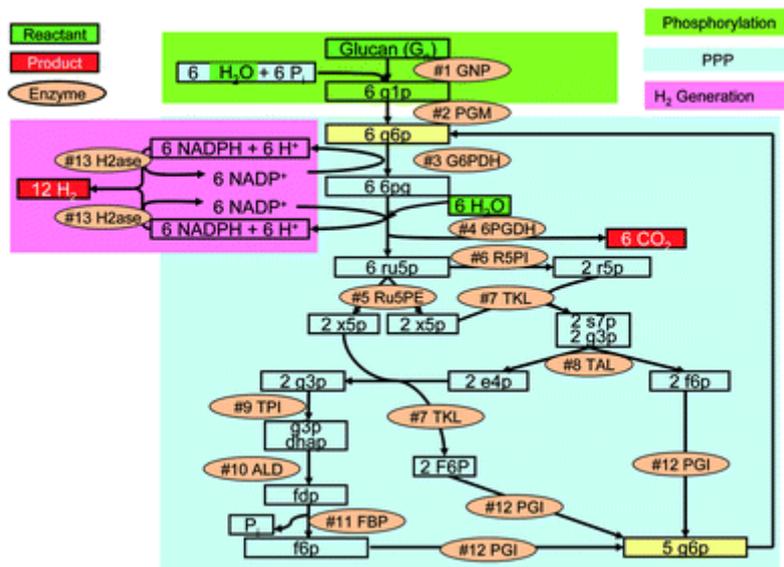
advantages and disadvantages to both *in vitro* and *in vivo* approaches, which will be analysed in the following subchapters.

### 1.5.3 *In vitro* biohydrogen production: Cell-free enzymatic pathways

Using chemical potential to sustain cell-free hydrogen production, requires the use of *in vitro* synthetic biology platforms. In their conceptually simplest design, hydrogenase enzymes are provided with a substrate that can be directly converted into hydrogen gas, for example by having a purified hydrogenase enzyme in a solution containing reduced ferredoxin (Fd, a biological substrate with high reducing power) at appropriate pH:

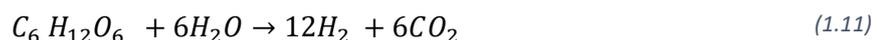


Simple one-step reactions are extremely inefficient and expensive as there are very few energy-carrier substrates that can be directly “fed” to hydrogen-producing enzymes and their production would have very high costs. Instead, a more sensible approach would be to use a readily available cheap feedstock containing substrates that can be utilized in a multi-step enzymatic pathway. In this case, the pre-treated cheap feedstock is added to an enzyme cocktail which catalyzes a cascade of intermediate reactions culminating in hydrogen production. Proof of principle studies conducted by Zhang *et al.* have shown that *in vitro* enzymatic pathways of 10 or more enzymes can be used to produce H<sub>2</sub> from various carbohydrates, including starch (Fig.1.5) (36, 40), cellulosic materials (41), xylose (42), sucrose (43), a mixture of biomass monosaccharides (37), and xylooligosaccharides (44).



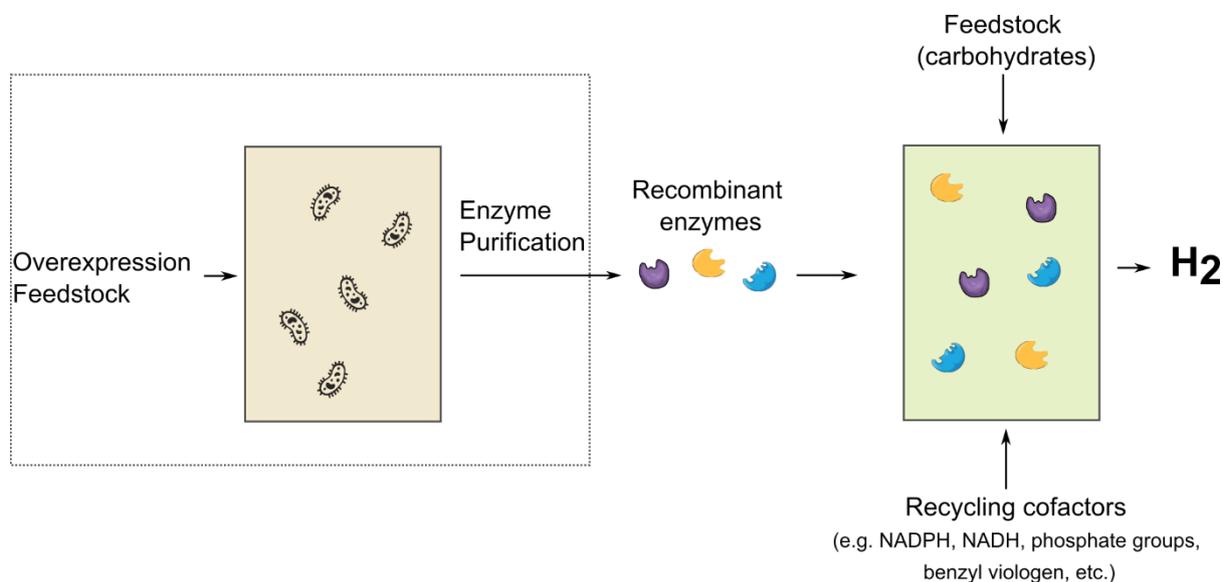
**Figure 1.5.** Scheme of an *in-vitro* synthetic biology pathway utilizing 13 thermophilic enzymes for the production of H<sub>2</sub> through complete utilization of glucan. The final step corresponds to the production of H<sub>2</sub> catalyzed by a hydrogenase enzyme which is unique in its ability to reduce hydrogen using NADPH instead of ferredoxin. Phosphate groups and NADPH are utilized as recycling cofactors (45). Image from Zhang *et al.* (45).

The above studies have shown that artificial ATP-free (ATP is expensive) and cofactor-balanced cell-free enzymatic pathways allow for exceeding the theoretical *in vivo* limit, the so-called “Thauer limit” of 4 mol H<sub>2</sub> produced per 1 mol of glucose (46). Some of these pathways can even approach the stoichiometric potential of 12 H<sub>2</sub>/glucose (Equation 1.11) resulting in a yield that is also much higher than the one from chemical catalysis of biomasses (typically ~50-57%) (47).



The high yields in cell-free enzymatic pathways compared to *in vivo* productions are not too surprising given that there is no formation of cell mass and the only reactions utilized are specific and optimized for the desired product. There are also several additional advantages, namely, modest reaction conditions as compared to chemical catalysis, no toxic chemicals required or produced, broad reaction conditions (e.g., high temperature and low pH) as compared with microorganisms, and easy operation and control (47). However, unlike living organisms, cell-free enzymatic pathways do not possess the ability to produce and replace their enzymes and recycling cofactors (substrates that are continuously recycled in the enzymatic pathway such as NADH, NADPH, methyl viologen etc.), therefore these must be produced beforehand using overexpression systems (enzymes) or chemical and enzymatical synthesis (cofactors) as shown in Fig.1.6. The limited lifespan of enzymes and recycling cofactors

(enzymes degrade and cofactors form irreversibly enzymatically inactive byproducts) is a major contribution to production costs. To reduce production costs, enzymatic pathways utilizing immobilized (hyper)thermostable enzymes and less expensive recycling cofactors can be used (41), however, so far all attempts remained commercially unsuccessful.



**Figure 1.6.** Schematic of H<sub>2</sub> production from a cell-free enzymatic pathway. The enzymes used to convert the carbohydrate feedstock into H<sub>2</sub> need to be produced beforehand using overexpression systems. The purified enzymes are then mixed with recycling cofactors and added to the reactor containing the carbohydrate feedstock. Recombinant enzymes and recycling cofactors have a limited lifespan and must be replaced.

In this regard, only one patent (EP2018394A2) was found on artificial enzymatic pathways for the production of hydrogen which was based on the work of Zhang (among the inventors). This patent was filed in 2007 but was then abandoned and currently is not active in any country, indicating that commercially it remains unsuccessful.

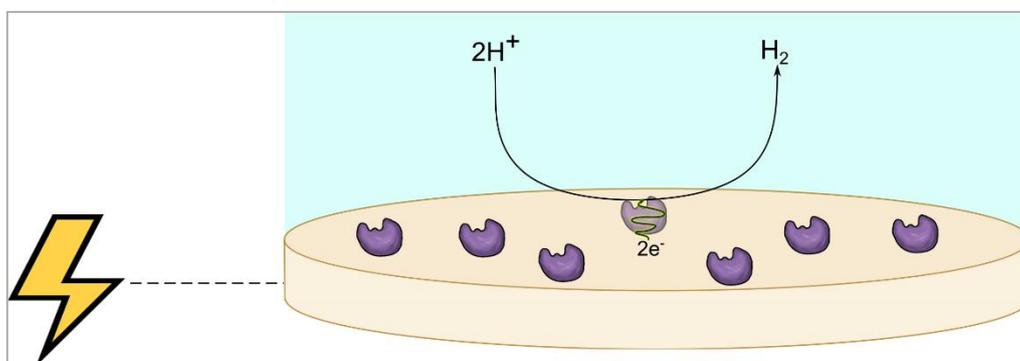
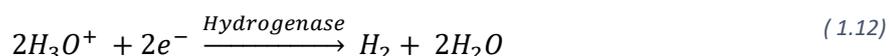
In 2010, Zhang *et al.* published a theoretical economic model of hydrogen-producing enzymatic pathways (48), which predicted a future cost of 1.5 USD/kg H<sub>2</sub> dependent on improvements in the stability of cofactors and enzymes. However, in a following technical report (2015) it was declared that production cost was estimated at 10,000 USD/kg H<sub>2</sub> at that time (49), which, in light of more realistic expectations, may explain the abandonment of the patent.

More recent proposals in enzymatic pathway designs claim to make *in vitro* biohydrogen production more attractive, although, this remains speculative (40). I also argue that publications seem to have been overly optimistic and unrealistic; with a simple analysis (Supplementary text A.1 in Appendix), one can show that utilizing cell-free enzymatic

pathways for the production of hydrogen is by far not economically viable without major advancements in enzyme stabilization techniques in combination with a significant reduction in protein production costs. These advancements are likely to require many decades and may even be too far from what could be ever realistically achieved. In particular, this analysis also shows the important contribution of the hydrogenase enzyme in the production cost and provides an estimate of its minimum catalytical requirements compatible with a commercial process.

### 1.5.4 *In vitro* biohydrogen production: Enzymatic electrosynthesis

Enzymatic electrosynthesis of hydrogen gas combines enzymatic catalysis and electrochemical techniques. Typically hydrogenase enzymes are immobilized on an electrode surface immersed in a buffer solution at appropriate pH. The electrode is connected to an electricity source and provides the electrons necessary for the reduction of protons (Fig.1.7) according to Equation 1.12.



**Figure 1.7** Enzymatic electrosynthesis of  $\text{H}_2$  schematic. A source of electrical current is connected to an electrode where hydrogenase enzymes are absorbed and stabilized in a buffer solution. Electrons transfer to the catalytic site enables the reduction of hydronium ions into hydrogen.

The technique has gained interest for the potential for the use of electrical energy from renewable resources and for the use of enzymes which being highly specific can reach high production yields (50). Furthermore, like all systems utilizing biocatalysts, it requires modest reaction conditions.

However, enzymatic electrosynthesis suffers from low production rates which are often limited by the rate of electron transfer from the electrode surface to the enzyme. However, electrode functionalization techniques with redox polymers have shown promising results (39, 51). With this strategy, it was demonstrated that the  $\text{H}_2$  production rate can reach  $50 \mu\text{mol}\cdot\text{min}^{-1}$  per 1 mg of the enzyme with a high faradaic efficiency of 80-90% (ratio between the experimentally obtained moles of  $\text{H}_2$  and the moles of  $\text{H}_2$  theoretically expected based on current flow) (39). Despite these advancements, the hydrogen production rate still remains too slow to meet the needs of practical applications, something that has been recognized also in other electro-enzymatic syntheses (50).

Most importantly there is the major bottleneck of enzyme stability which was encountered also in the cell-free enzymatic pathways discussed in Chapter 1.5.3. In particular, all enzymes that have been reported in studies involving the electrochemical production of hydrogen are oxygen-sensitive (hydrogenases, nitrogenases (52), HDCR (51)). Requiring an anaerobic environment or oxygen protection strategies (53, 54) can significantly increase production costs. In principle, one can choose an oxygen-tolerant or insensitive hydrogenase (group 1d and group 5 respectively, see table A.1 in the appendix), however, these [Ni-Fe]-hydrogenases, are less suitable for hydrogen production due to the higher electrochemical potential of the [NiFe] cluster compared to the oxygen-sensitive [FeFe]-hydrogenases.

We can better understand the feasibility and limitations by using a similar economic evaluation to the one that was presented for the cell-free enzymatic pathways. From this simple analysis (Supplementary text A.2 in Appendix), it is clear that the enzymatic electrosynthesis of hydrogen gas, like all electricity-based hydrogen production methods, is currently not commercially competitive with steam methane reforming (SMR). However, contrary to standard electrolysis, its future seems to be grimmer and highly dependent on major breakthroughs in enzyme stabilization and production technologies regardless of how advanced and affordable electrochemical devices get.

### **1.5.5 In vivo biohydrogen production: Direct microbial production**

Direct hydrogen production requires organisms capable of producing hydrogen through fermentative or light-dependent processes. There are multiple metabolic pathways that microorganisms utilise for the production of hydrogen (Table 1.4). These are divided into direct biophotolysis, indirect biophotolysis, photofermentation, and dark fermentation (57).

- **Direct photolysis:** Some algae and cyanobacteria split water directly into oxygen and hydrogen to vent excess electrons during photosynthesis. The photosynthetic electrons are transported by ferredoxin and other intermediates to a hydrogenase enzyme, which catalyses the production of H<sub>2</sub>. However, the hydrogenase is rapidly inactivated by photosynthetically evolved O<sub>2</sub>, and therefore direct photolytic H<sub>2</sub> production is only possible when the cultures are incubated under special conditions such as sulfur deprivation that allow for the maintenance of an anaerobic environment (58). It has been reported that this method allows the production of hydrogen gas with a purity of up to 98%, however, it suffers excessively low production rates. Research work has been carried out to engineer algae and bacteria so that the majority of the solar energy is diverted to hydrogen production and to overcome the limitation of oxygen sensitivity (59).
- **Indirect biophotolysis:** Cyanobacteria and microalgae utilize photosynthesis to capture solar energy and convert it to some form of carbohydrate, which is stored and can be later used to produce hydrogen by oxygen-sensitive nitrogenase enzymes (cyanobacteria) or hydrogenases (microalgae and cyanobacteria). The advantage of indirect biophotolysis is that the stage of oxygen generation can be temporally or spatially separated from the stage of hydrogen evolution which can prevent the inactivation of oxygen-sensitive enzymes (60, 61). However, significant challenges remain with still too low hydrogen production rate and yields.
- **Photofermentation:** Purple non-sulfur bacteria produce H<sub>2</sub> under nitrogen-deficient conditions using light energy and organic compounds such as acetate, butyrate, and lactate. However, the practical applicability of photofermentation is limited by a low light conversion efficacy and the requirement for suitable sterile and environmental conditions of bioreactors (62).
- **Dark fermentation:** A type of anaerobic digestion in which hydrogen can be produced by anaerobic bacteria on carbohydrate-rich substrates. These fermentation processes predominantly give rise to acetic and butyric acids together with hydrogen gas. According to reaction stoichiometry, bioconversion of 1 mol of glucose into acetate yields 4 mol H<sub>2</sub>/mol glucose (Thauer limit), but only 2 mol H<sub>2</sub>/mol glucose is formed when butyrate is the end product. Currently, fermentative processes produce 2.4 to 3.2 moles of hydrogen per mole of glucose (63). Dark fermentation does not depend on light and therefore can continuously produce energy and, as oxygen is not produced, oxygen-sensitive hydrogenases are not inhibited. However, contrary to biophotolysis, it requires organic

biomass as a feedstock and H<sub>2</sub> is produced as mixed biogas. Many scale-up challenges in the production of biohydrogen through dark fermentation have been identified (64).

**Table 1.4.** Different metabolic processes for H<sub>2</sub> production in different organisms. Advantages and disadvantages are presented. Adapted from Engin Gurtekin's paper (65).

Process	Organisms	Advantages	Disadvantages
Direct biophotolysis	- Green algae - Cyanobacteria	- Abundant inexpensive substrate (H <sub>2</sub> O) and energy source (light) - High purity H <sub>2</sub> (98%)	- Incompatibility between photosynthetically evolved O <sub>2</sub> and H <sub>2</sub> evolving enzymes - The resulting H <sub>2</sub> and O <sub>2</sub> mixtures are highly explosive - Extremely low production rate (0.07 mmol/L/hr) - Requires a high-intensity light source and large surface areas
Indirect biophotolysis	- Microalgae - Cyanobacteria	- Abundant inexpensive substrate (H <sub>2</sub> O) and energy source (light)	- Uptake hydrogenase enzymes are to be removed to stop the degradation of H <sub>2</sub> - Very low production rate (0.36 mmol/L/hr) - Requires a high-intensity light source and large surface areas
Photofermentation	- Purple non-sulfur bacteria	- A wide spectral light energy can be used - Can use different organic wastes	- O <sub>2</sub> has an inhibitory effect on nitrogenases - Light conversion efficiency is very low (1–5%)
Dark Fermentation	- Anaerobic bacteria	- It can produce H <sub>2</sub> continuously - A variety of carbon sources can be used - It is an anaerobic process, so there is no O <sub>2</sub> inhibition problem	- H <sub>2</sub> is produced as a mixture of CO <sub>2</sub> and other gases - Requires a continuous supply of biomass

We should remind ourselves that even though these processes have been studied for decades, none has found practical application in the production of H<sub>2</sub>. However, anaerobic digestion has found widespread real-world application in the production of biogas which includes a small percentage of H<sub>2</sub> (< 1%) but is mostly methane (50–75%). While H<sub>2</sub> only constitutes a very small fraction of biogas, it has an important role in its production as the H<sub>2</sub> produced by syntrophic bacteria is consumed by methanogens to catalyze the reduction of CO<sub>2</sub> and/or other simple organic molecules into methane (CH<sub>4</sub>) (66).

This biogas is currently being produced from waste materials (agricultural waste, manure, sewage, food waste etc.) using anaerobic digesters and has found use as a substitute for natural gas and to produce electricity. Increasing the percentage of H<sub>2</sub> produced in anaerobic digestion is particularly challenging as it requires changing the natural microbiota present in the organic

waste feedstock, meaning that further processing is required such as special treatments to kill the hydrogen-consuming methane-generating bacteria (US20040050778A1) or large-scale sterilization of feedstock and inoculation with the desired bacteria strains which can be natural or genetically engineered. If genetically engineered bacteria are used, a significantly increased H<sub>2</sub> production can be achieved, however, any environmental wild-type bacteria will most likely out-compete the mutated strains and tight sterile conditions must be maintained. Furthermore, the use of genetically engineered bacteria (OGM) is strictly regulated by governments and any environmental contamination must be avoided thus also limiting the uses of the digestate material e.g. as a fertilizer. All these issues increase H<sub>2</sub> production costs and limit the economical feasibility and attractiveness of the production of higher-percentage H<sub>2</sub> biogas. Other limitations are related to safety concerns (H<sub>2</sub> is explosive), limited hydrogen yield (Thauer limit), and low purity as a mixture of gasses would be produced. Furthermore, we are already extracting valuable products utilizing standard anaerobic digestion (i.e. digestate and methane-rich biogas) and therefore any attempt to produce higher-percentage H<sub>2</sub> biogas must be justified with an expected overall higher revenue.

Another strategy is to utilize the methane-rich biogas produced using anaerobic fermentors as a reagent in steam reforming. This has recently been tested by a spinout company in Austria claiming high-purity hydrogen production for 5 €/kg utilizing manure biogas (67). It remains to see whether this will emerge in the future as an additional source of economically convenient hydrogen.

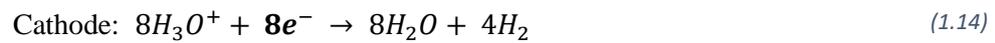
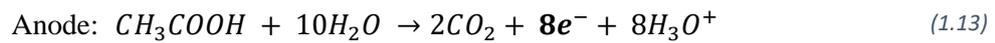
### **1.5.6 *In vivo* biohydrogen production: Microbial electrolysis cells**

The generation of ATP in cells is driven by exergonic redox reactions in which an “energy-rich” substrate (e.g. glucose) is oxidised. The electrons from this oxidation are ultimately received by a terminal electron acceptor; in aerobic respiration, this is molecular oxygen (O<sub>2</sub>) while in anaerobic fermentation various organic and inorganic molecules are used such as pyruvate (e.g. acid lactic fermentation), fumarate, nitrates (NO<sub>3</sub><sup>-</sup>), sulfate (SO<sub>4</sub><sup>2-</sup>), and many others.

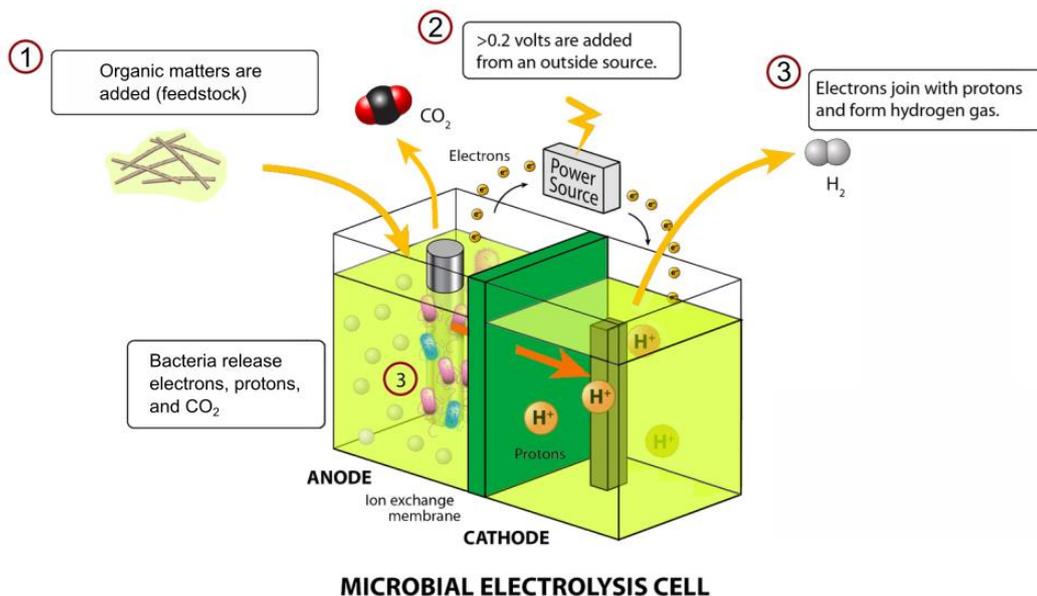
Some microorganisms, such as *S. aureus*, *E. faecalis*, and *Shewanella oneidensis*, are able to transfer electrons extracellularly to a solid conductor or strong extracellular oxidizing agents by contact with outer membrane cytochromes or through conductive appendages or pili

(nanowires) (68). This property allows these microbes (called exoelectrogens) to utilize extracellular terminal electron acceptors to sustain their metabolism.

Exoelectrogen microorganisms have been extensively used in Microbial Electrolysis Cells (MEC) where they convert biodegradable matter into electric current through electron transfer to an electrode (anode) which acts as a terminal electron acceptor in their metabolism (Reaction 1.13 in case acetate is the substrate). The current produced can be used for multiple purposes including reducing protons resulting in H<sub>2</sub> evolution (Reaction 1.14).



However, the electrical potential produced by these devices is not sufficient on its own for proton reduction, and therefore additional low-energy input (0.2–0.8 V) is supplied from an external source (Fig.1.8) (69).



**Figure 1.8.** General schematic of a microbial electrolysis cell. Organic matter is added as a feedstock in the anode compartment where exoelectrogen microbes are present. These microbes use the electrode (anode) as a terminal electron acceptor in their metabolism resulting in the production of an electric current that can be used, together with an external electric contribution, to reduce protons in the cathode generating H<sub>2</sub>. The external energy supply is necessary as the electric potential generated by the bacteria activity is not sufficient on its own to drive proton reduction. Image adapted from Zina Deretsky, National Science Foundation (Wikipedia).

Even though a small energy input is required, the MEC technology remains energetically advantageous with a theoretical electrical energy yield of up to 10 times that of abiotic water

electrolysis (70). This means that at least in theory, in MEC devices 10 times less electric energy could be used to produce the same amount of hydrogen gas with the rest coming from “inexpensive” chemical potential stored in waste biomasses and with the additional safety advantage of not producing any oxygen in the process (explosion hazard). Furthermore, a higher hydrogen yield can be achieved compared to dark fermentation which is limited to a maximum of 4 mol H<sub>2</sub>/mol glucose (Thauer limit).

However, in pilot-scale experimental realizations, the electric energy yields are significantly reduced compared to the theoretical ones, and most importantly, current densities remain dramatically low (less than 1 A·m<sup>-2</sup>) primarily due to the relatively low activity and concentration of the microorganisms involved in the process. High current densities are imperative for scaling up industrial electrolyzers as many chemical engineer issues are exacerbated when the electrode surface becomes larger (greater likelihood of clogging or leakages, greater difficulty in maintaining flatness, higher number of connexions, more delicate and longer disassembly/reassembly operations, higher risk of stray current, etc.) (70).

A further issue to overcome is to make industrial electrochemical processes compatible with MEC (70); all industrial-scale electrochemical processes developed so far require electrolytes with the simplest possible chemical composition while sustaining microbial growth requires chemically rich media.

Overall, there is theoretical potential for the use of MEC especially when “inexpensive” waste biomasses are used, however, performances are still unsatisfactory especially when considered in the context of scaling up.

### **1.5.7 Biohydrogen production: prospects and conclusions**

I showed that *in vitro* technologies for the production of hydrogen suffer major challenges mostly (but not only) determined by their requirement for a continuous supply of enzymes and cofactors combined with the prohibitive cost associated with them. *In vivo* approaches do not have such an issue as cells are complex machines capable of continuously producing the enzymes and cofactors needed. However, cells spend a significant amount of the energy provided for growth and maintenance making them less efficient “energy converters” with typically low hydrogen evolution rates.

All taken into consideration, challenges such as low yields, prohibitive costs and scaling-up challenges make large-scale production of hydrogen from biological systems highly unlikely

to emerge in the future as a competitor technology to steam reforming and standard water electrolysis. On the other hand, biogas production from anaerobic digestion is likely to increasingly emerge as a valuable relatively low-cost technology to extract energy from waste biomasses in the form of methane which may be used in steam reforming for the production of hydrogen.

## **1.6 Hydrogen to energy: Fuel cells**

### **1.6.1 Fuel cells and the drawbacks of platinum-based catalysts**

Fuel cells are electrochemical devices that convert the energy released by the oxidation of a fuel, such as hydrogen, into electricity. In particular, alkaline fuel cells, which oxidize hydrogen with oxygen to produce electricity and water, are considered the cheapest and most electrically efficient (71). However, the adoption of fuel cells is limited by prohibitive costs, safety concerns, the absence of widespread consumer hydrogen providers, and limited market demands.

In particular, the high cost of fuel cells is a major limitation and is attributed to multiple factors (72) with platinum catalysts (platinum is expensive and scarce) claimed by some academics to be a major contributor by more than 40% of the production cost (73). However, according to the US Department of Energy, the cost of manufacturing the fuel cell stack is the main contributor and not the materials (i.e. platinum) used to produce it (74). Undoubtedly the use of an expensive metal such as platinum has a contribution, however, the extent of it remains uncertain with multiple sources contradicting each other.

We also know that some fuel cells are based on cheaper and more common metals such as nickel, however, platinum-based catalysts remain superior due to their higher activity which allows operation at a lower temperature (~60-80 °C). Efforts have been made to reduce the amount of platinum used and to find cheaper inorganic catalysts with equivalent performances (75, 76), which resulted in significant advances.

In addition, some academics support the idea that biological systems can eventually substitute expensive platinum catalysts, claiming that hydrogenases emerge as good candidates (77). These claims appear to be superficial and out of touch with reality. In this regard, the following

Section will explain more in detail why the use of enzymes as alternative catalysts in fuel cells is unreasonable.

### **1.6.2 Enzymatic fuel cells: can hydrogenases substitute platinum-based catalysts?**

Before jumping to conclusions claiming that hydrogenases will make fuel cells economical, we should better analyze the practical problem to solve and see whether hydrogenases are realistically a solution.

Typically, a fuel cell requires 0.1 and 0.5 grams of platinum per kilowatt (kW) of power output (78, 79). At current market prices, platinum costs around 30-40 USD per gram. So, the cost of platinum in a fuel cell could range from \$3-\$20 per kW of power output. At the largest annual production volume (50,000 units per year), the overall fuel cell system cost per kilowatt was found to be between 1,215 and 1,875 USD/kW (80), which means that the cost of platinum as a raw material contributes roughly 0.2-1.3% of the total production cost. Based on these data, the cost of platinum seems to be less critical than what is claimed in the academic literature. However, platinum catalysts suffer degradation issues and can be poisoned by gasses such as CO and H<sub>2</sub>S (81–83). The durability of the catalyst is an important factor in the adoption of fuel cells and an increase from the current ~10,000 operational hours is needed (84–86).

Now, it remains to be answered whether hydrogenases as substitutes for platinum catalysts have the potential to reduce production costs and/or improve the durability of fuel cells.

Contrary to H<sub>2</sub> production processes, here [NiFe]-hydrogenases can be chosen as the most suitable candidates for H<sub>2</sub> oxidation (physiologically hydrogen uptakers). All known oxygen-tolerant and insensitive hydrogenases belong to this category, meaning that there is potential for developing a hydrogenase-based catalyst without the issue of oxygen sensitivity, for example, by carefully choosing and engineering an oxygen-insensitive candidate. In the literature, there are a few such examples where wild-type hydrogenases with various degrees of oxygen tolerance (they retain part of their catalytical activity in the presence of oxygen) are tested as fuel cell catalysts (87, 88). Furthermore, hydrogenases do not suffer irreversible carbon-monoxide poisoning (89) and are very selective catalysts.

Having said that, there are major prohibitive factors that make the use of hydrogenases as fuel cell catalysts very unrealistic even with substantial development (Table 1.5).

The most prohibitive factors are related to limited lifespan and excessive costs. Imagining an enzyme surviving on an electrode for more than 10 days at room temperature or higher while maintaining most of the catalytical activity would be met with scepticism by most biochemists even for stable enzymes. This is likely an intrinsic limit of using enzymes regardless of how advanced the stabilization techniques get.

Furthermore, when considering the costs of enzymes we must also consider the cumulative effect of replacement. Not only are hydrogenases more expensive per amount of energy produced than platinum, but they also require an unsustainable frequent replacement, therefore, increasing labour costs.

**Table 1.5** Advantages and disadvantages of platinum-based catalysts vs. hydrogenases for hydrogen oxidation in fuel cells. The text in red represents the most critical disadvantages.

Property	Platinum-based catalyst	Hydrogenase-based catalyst
Lifespan	Common Pt-based fuel cells can last at least ~10000 operational hours	Replacement every few days
Cost of raw material (platinum vs. enzymes) per kWh produced	<0.002 USD/kWh Pt metal contributes ~\$3-\$20 per kW of power output and operational time is 10000 hours	>1.5 USD/kWh \$1000 per kg of hydrogenase, 33.3kWh/kg H <sub>2</sub> , overestimated TTN <sub>hyd</sub> = 1,000,000
Availability	Scarce resource	Can be made as needed
Catalytic rate	Can sustain requirements	It seems hydrogenases can only compete with Pt-based catalysts at low hydrogen concentrations which results in low current outputs (28).
A degree of oxygen inactivation	No	Typically yes, but this may be overcome by selecting and engineering appropriate hydrogenases.
CO irreversible poisoning	Yes	No

All attempts to make enzymatic fuel cells with hydrogenases have commercially failed. In this regard, Fraser Armstrong's academic work on using hydrogenases including oxygen-tolerant ones in fuel cells resulted in two patents (EP1421638A2 and EP1787350B1) both then abandoned. Another related patent (EP1939961A1) derived from academic research and protecting a fuel cell technology utilizing covalently bound hydrogenases was abandoned as well.

### **1.6.3 Hydrogenases-inspired catalysts**

Having shown the prohibitive limitation of using hydrogenases as substitutes for platinum-based catalysts, one may wonder whether hydrogenases can at least be used to inspire new inorganic catalysts.

Biomimetic catalysts are chemical catalysts that mimic certain key features of enzymatic systems, for example by imitating aspects of binding cavities (e.g. cyclodextrins and cucurbiturils-based catalysts, resorcinarene-derived cavitands, molecularly imprinted polymers), self-assembling properties (metal-directed macrocycles), and catalytical interactions (oligopeptides) (90).

However, all these examples are only very marginally related to their biological counterpart and a very detailed understanding of specific enzyme mechanisms is unlikely to be beneficial. In simpler terms, it is like extensively studying birds in order to design aeroplanes (Blaza, personal communication), two very different things that have only a modest conceptual correlation.

Nevertheless, many academics embarked on a journey of understanding in great detail the mechanisms of enzymes, implicating that these studies can be potentially relevant for designing novel biomimetic catalysts. In particular, numerous efforts have been made to characterize the active sites of hydrogenases to aid in the development of biomimetic alternatives to platinum catalysts. Given the profound difference in nature between inorganic catalysts and enzymes, it is unlikely that these extremely detailed studies will provide any benefit to the biomimetic field, however, they may be more interesting as a matter of general but unpractical curiosity.

## **1.7 Why should we study hydrogenases?**

### **1.7.1 Potential avenues**

As discussed in the above Sections, I am highly sceptical of the use of hydrogenases in the context of the hydrogen economy. In this regard, I identified two possible applied avenues that may speculatively benefit from a deeper understanding of hydrogenase enzymes. One is the engineering of bacterial metabolism for the production of chemicals that currently have

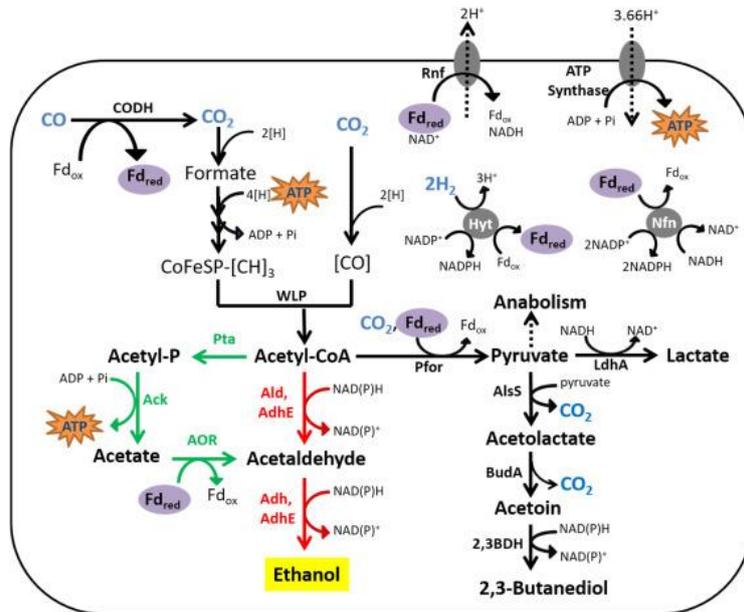
significantly higher production costs than hydrogen, or that have regioselectivity challenges, and the other is the role of hydrogenases in pathogenic bacteria which may be used as a drug target.

Overall my work can be considered classical academic research of doing things because they are deemed “interesting” which I argue delays resolving pressing real-world problems. Nevertheless, by academic standards, this work is a relevant contribution to the hydrogenase field.

### **1.7.2 Hydrogenases in synthetic biology**

Bioengineering the metabolism of organisms requires adding or deleting genes encoding various metabolically relevant enzymes. When engineering cells, the influxes and outfluxes of intermediate products and electrons must be taken into consideration and hydrogenases play an important role in this balance.

For example, gas fermentation has emerged as a promising technology that converts industrial waste gases or syngas containing CO, CO<sub>2</sub> and H<sub>2</sub> into fuels without impacting food production (91). In these processes, bifurcating hydrogenases oxidase hydrogen producing reduced NADH and reduced ferredoxin, contributing to balancing the cofactors in the synthetic pathway (Fig.1.9). Understanding the precise mechanisms of bifurcating hydrogenases, such as the one presented in this thesis, may help with the development of engineered enzymes that can be integrated into the genome of bacteria, however, this remains highly speculative.



**Figure 1.9** Autotrophic product formation in *C. autoethanogenum*. The ATP-efficient, indirect ethanol route employs a NADP-dependent electron bifurcating hydrogenase (Hyt). Image taken from (91).

### 1.7.3 Hydrogenases in pathogenic bacteria

Pathogenic microorganisms use various mechanisms to conserve energy in host tissues and environmental reservoirs. One of these mechanisms is through the consumption or production of H<sub>2</sub> (92).

Several major human pathogens use the H<sub>2</sub> produced by colonic microbiota as an energy source for aerobic or anaerobic respiration. This process is critical for the growth and virulence of gastrointestinal bacteria such as *Salmonella enterica* and *Helicobacter pylori* (92). Other pathogens produce H<sub>2</sub> as an end product of fermentation processes. This is the case for some pathogenic facultative anaerobes (e.g. *E. coli*, *Giardia intestinalis*) and for obligate anaerobes (e.g. *Trichomonas vaginalis*, *Clostridioides difficile*) (92).

Even though H<sub>2</sub> metabolism is important for many pathogens, this has often been overlooked and we currently lack a detailed understanding of it. In principle, H<sub>2</sub> metabolism could be used as a possible target for drug development, especially in the current context of increasing antibiotic resistance. One possible approach is to develop small-molecule inhibitors targeting hydrogenases or their maturation proteins. However, hydrogenase inhibition is likely to not exert sufficiently severe effects as bacteria metabolism is often very flexible. Nevertheless,

some pathogens such as *Helicobacter pylori* and obligate fermenters such as *Trichomonas vaginalis* have more limited metabolic flexibility and therefore can be more susceptible to hydrogenase inhibition (92). To maximise the efficacies of treatments, therapies that inhibit hydrogenases together with other targets can be used.

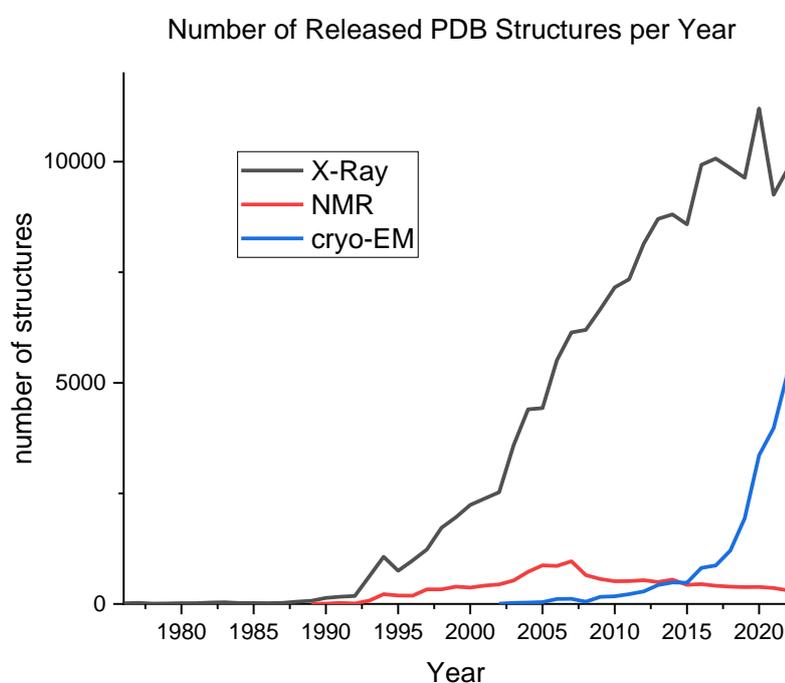
Another challenge in using hydrogenase inhibitors is the possibility of off-target effects which may disrupt the healthy gut microbiota (common also with antibiotics), or more seriously, cause dysfunction in other enzymes in human cells; for example, if quinone analogues are used to inhibit the activity of membrane-bound hydrogenases this can also lead to dysfunction of other quinone-dependent enzymes such as complex I. However, many hydrogenases and their maturation proteins have no sequence similarities with any mammalian protein, for example, this is the case for Hydrogenase-I, or the [Ni-Fe]-hydrogenase and the maturation proteins hypE or hypF from *H. pylori* (checked using BLAST (93)). Therefore substrate analogues that have significantly higher affinities for these proteins can be used.

However, this optimistic view of using hydrogenase research for the development of antibiotics is far from becoming a reality. Unfortunately, the antibiotic market is broken resulting in no new classes of antibiotics discovered since the 1980s (94–96). This is not because of a lack of ideas about new targets or candidates but because pharmaceutical companies do not have enough economic gains to stimulate the development and production of new antibiotics (95). In addition, basic research accounts only for a very small fraction of the cost associated with antibiotic drug development and therefore any basic academic research must be coordinated with pharmaceutical companies to make sure that we are not investing millions of public money in creating more pre-clinical candidates that no company cares to invest further (as too frequently happens!).

## 1.8 Cryogenic electron microscopy

### 1.8.1 Structural biology techniques and Cryo-EM

Structural studies provide invaluable information for the interpretation of molecular mechanisms and have fundamental applications in drug discovery (97). Over the past decades, X-ray crystallography has been the most widely used technique to obtain the structure of proteins, however, more recently single particle cryo-electron microscopy (cryo-EM) emerged as a competitive alternative with a rapidly growing number of structures released every year (Fig.1.10).



**Figure 1.10** Plot showing the number of PDB structures released per year obtained with X-ray crystallography (black), nuclear magnetic resonance (NMR) (red), and single particle cryo-EM (blue). Plot obtained using PDB and EMBD statistics (98, 99).

The adoption of cryo-EM in structural biology followed the so-called “Resolution Revolution” (100) brought about by significant advancements in hardware, such as direct electron detectors, coupled with algorithmic improvements in image processing. With these advancements, cryo-EM is now capable of routinely resolving macromolecular assemblies at resolutions between 2.5 and 4.0 Å (101), with some examples (less than 1% of the total released cryo-EM structures) below 2 Å and the best resolution being 1.25 Å obtained with a prototype electron

microscope with a reduced energy spread of the electron beam and reduced optical aberrations (102).

Although X-ray crystallography usually results in details beyond 3 Å (~ 40% of structures have a resolution  $\leq 2$  Å (103)), its application is limited when studying flexible complex macromolecules (104). The main reason for this limitation is the requirement for protein crystallization which is better suited for small rigid proteins. Even when the crystallization of complex flexible proteins is possible, these are forced into a fixed conformational state that maximises lattice energy.

On the other hand, cryo-EM records micrographs of the biological samples in fully hydrated conditions, allowing visualization of conformational variabilities. Cryo-EM enables visualization of samples that are difficult to crystallize utilizing significantly smaller amounts of sample (~1-2 µg per microscope grid) with minimal specimen preparation. However, cryo-EM is not suitable for imaging proteins smaller than ~50 kDa as these have few distinctive morphological features, complicating the particle alignment needed for protein reconstructions (105). To overcome this challenge, small proteins can be imaged bound in complexes with other proteins (e.g. nanobodies, ligands) or as fusion proteins, however, this significantly complicates the sample preparation process.

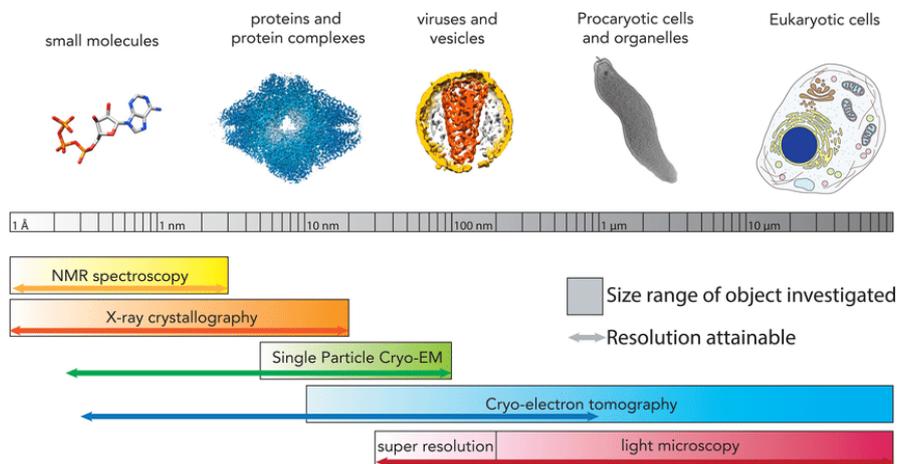
Nuclear magnetic resonance spectroscopy of proteins (NMR) is another technique that can provide information about the structure and dynamics of proteins, however, this is limited to smaller proteins (< 35 kDa) due to the complication of interpretation of data coming from larger proteins. An overview of these three structural techniques is provided in Table 1.6.

**Table 1.6.** Characteristics of X-ray crystallography, single-particle cryo-EM, and NMR for the determination of the 3D structure of proteins.

	<b>X-Ray crystallography</b>	<b>Single particle cryo-EM</b>	<b>Protein NMR</b>
Protein size	Best for rigid smaller proteins	Ideally > 100k Da Not possible < 50 kDa	Only < 35 kDa
Resolution	< 3 Å	Between 2.5 Å and 4.0 Å	n.a.
Sample needed	Few mg	Up to 10-20 µg	Few mg
Sample preparation	Very hard and time-consuming (requires crystallization)	Relatively simple	Relatively simple
Suitable for flexible proteins	No	Yes	Yes (small peptides)

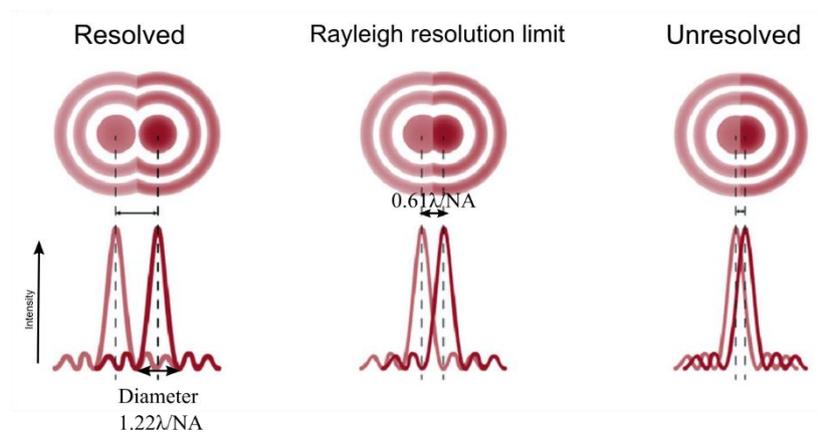
## 1.8.2 Using electrons to visualize proteins: theoretical resolution

Proteins and protein complexes are classified as nanoparticles (between 1–100 nm) and are far too small to be visualized with visible light (400-700 nm) (Fig.1.11).



**Figure 1.11** An overview of structural biology techniques and the biological objects they investigate. Image taken from Joshua Hutchings *et al.* (106)

To understand why this is the case the concept of resolution of a microscope needs to be introduced; this is defined as the smallest distance between two points on a specimen that can still be distinguished as two separate entities. Microscopes introduce blurring effects in which each point source is imaged as a bell-shaped object expressed by a point spread function (PSF) that can be approximately thought of as a Gaussian function. The closer any two points are the bigger the overlap between their PSFs making them less distinguishable (Fig.1.12)

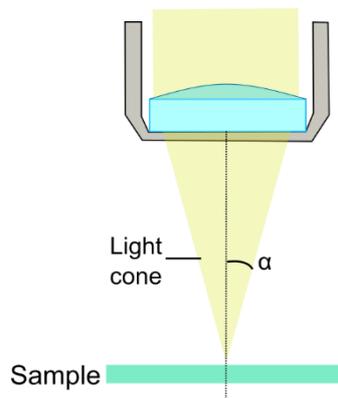


**Figure 1.12.** 2D-Point Spread Functions of two points source using a perfect imaging system with no aberrations. The greater the overlap of PSFs the less distinguishable points are. NA is the numerical aperture. Image adapted from Sebastian Dunst *et al.* (107).

When the maxima of the PSFs of any two points are at a distance greater than half the diameter of the bright central circle these are considered resolved (Rayleigh resolution limit in Fig. 1.12). Mathematically the resolution is therefore expressed as:

$$d = 0.61 \frac{\lambda}{n \cdot \sin(\alpha)} \quad (1.15)$$

Where  $d$  is the resolution,  $\lambda$  is the wavelength of light,  $n$  is the refractive index of the imaging medium (air, oil, vacuum), and  $\alpha$  is half of the cone angle of light from the specimen plane accepted by the objective (Fig. 1.13).  $n \cdot \sin(\alpha)$  is the numerical aperture (NA) and is equal to  $\sim 1$  for optical microscopes (108).



**Figure 1.13.** The numerical aperture of a microscope objective lens refers to the light-gathering ability of the lens and is calculated using half the angle of acceptance ( $\alpha$ ).

Formula 1.15 can be used to calculate the approximate theoretical resolution of a visible light microscope which is  $\sim 250\text{--}420$  nm. This is significantly higher than the dimension of any protein and as a result, proteins cannot be resolved (“seen”).

To resolve proteins, and even better, to resolve their detailed molecular structure we need to use probes with significantly smaller wavelengths so that the resolution is of about the same order of magnitude as what we want to visualize. One solution is to use X-rays which have significantly shorter wavelengths ( $\lambda$ ) than visible light as it is done in X-ray crystallography ( $\lambda = 1\text{--}1.5$  Å). Alternatively, accelerated electrons can be used; this is the case in cryo-EM.

The use of electrons to visualize small particles is the result of a century of discoveries, which started with Louis de Broglie who predicted the concept of wave-particle duality of all components of matter, including electrons (109). To calculate the wavelength of an electron the de Broglie Wave Equation can be used (Formula 1.16):

$$\lambda = \frac{h}{p} = \frac{h}{m_0 \cdot v} \quad (1.16)$$

Where  $\lambda$  is the wavelength of the electron,  $p$  is the momentum,  $h$  is the Plank's constant ( $6.626 \cdot 10^{-34} \text{J}\cdot\text{s}$ ),  $m_0$  is the rest mass of the electron ( $9.11 \times 10^{-31} \text{kg}$ ), and  $v$  is the velocity of the electron.

The velocity,  $v$ , of an electron can be calculated from its kinetic energy ( $E_k$ ):

$$E_k = \frac{1}{2} m_0 \cdot v^2 \quad (1.17)$$

The kinetic energy  $E_k$  is also equal to the charge of the electron ( $e = 1.6 \times 10^{-19} \text{C}$ ) multiplied by the applied accelerating voltage ( $V_a$ ). Therefore the velocity of the electron ( $v$ ) is equal to:

$$v = \sqrt{\frac{2 \cdot e \cdot V_a}{m_0}} \quad (1.18)$$

By substituting 1.18 in 1.16 we can derive that the wavelength of the electron ( $\lambda$ ) can also be calculated as:

$$\lambda = \frac{h}{\sqrt{2 \cdot m_0 \cdot e \cdot V_a}} \quad (1.19)$$

However, when electrons are accelerated to a significant fraction of the speed of light ( $c = 3 \cdot 10^8 \text{m/s}$ ), as it happens in electron microscopy ( $\sim 70\%$  of  $c$ ), then relativistic effects must be taken into account. According to the special relativity theory, the relativist velocity ( $v_{rel}$ ) can be calculated from the kinetic energy ( $E_k$ ) expression as:

$$v_{rel} = c \cdot \sqrt{1 - \frac{1}{\left(1 + \frac{E_k}{m_0 \cdot c^2}\right)^2}} = c \cdot \sqrt{1 - \frac{1}{\left(1 + \frac{e \cdot V_a}{m_0 \cdot c^2}\right)^2}} \quad (1.20)$$

Formula 1.20 can be used to calculate the relativistic momentum  $p_{rel}$ :

$$p_{rel} = \frac{m_0 \cdot v_{rel}}{\sqrt{1 - \frac{v_{rel}^2}{c^2}}} = \sqrt{2 \cdot m_0 \cdot e \cdot V_a + \left(\frac{e \cdot V_a}{c}\right)^2} \quad (1.21)$$

The momentum  $p_{rel}$  can be used to generalize the de Broglie Wave Equation (Formula 1.16) to account for relativistic effects:

$$\lambda = \frac{h}{\sqrt{2 \cdot m_0 \cdot e \cdot V_a + \left(\frac{e \cdot V_a}{c}\right)^2}} \quad (1.22)$$

Using formulas 1.20 and 1.22, it is possible to derive that for a typical cryo-EM microscope with an accelerating voltage of 200 kV, electrons reach a velocity of  $2.08 \cdot 10^8 \text{m/s}$  with an associated wavelength  $\lambda$  of 2.51 pm. Similarly, for a microscope with an accelerating voltage

of 300 kV,  $v_{rel} = 2.33 \cdot 10^8$  m/s and  $\lambda = 1.97$  pm. We have here demonstrated that the wavelengths of the electrons accelerated in typical electron microscopes are very small; to give a comparison this is ~60 times smaller than the radius of an atom of hydrogen (120 pm).

However, due to the geometry of the electron microscopes used in cryo-EM (TEM microscopes, discussed later), their numerical aperture is equal to ~0.01 (108). When using these values of numerical aperture and wavelengths in Formula 1.15, we can obtain a theoretical resolution of 1.5 Å in a 200 kV electron microscope and a resolution of 1.2 Å in a 300 kV electron microscope. In principle, we would be able to image at atomic resolution, “seeing” all the single atoms in a protein, however, lens aberrations and electron damage of the sample significantly worsen resolution resulting in experimental resolutions between 2.5 and 4.0 Å.

### 1.8.3 Cryo-EM sample preparation and vitrification

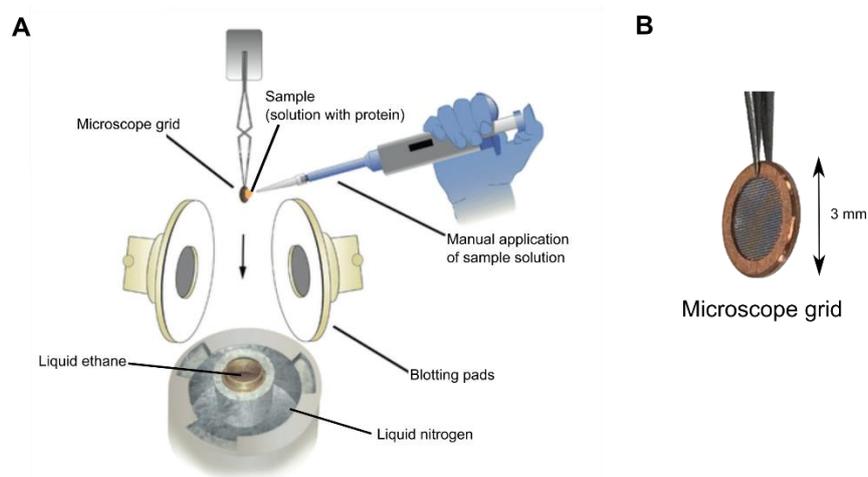
Initially, negative staining was used to image biological samples on electron microscopes. This technique involved the use of a thin layer of dried heavy metal salt to increase specimen contrast, however, low resolution and the loss of native conditions were major disadvantages. In 1981 Jacques Dubochet and Alasdair McDowell presented a method for preparing thin layers of vitrified ice for direct observation in the electron microscope (110). Since then the technique took off and it is now standard cryo-EM practice to image proteins in vitreous ice.

Vitreous ice is an amorphous form of ice in which water molecules retain their unstructured orientation of the liquid state. This allows the imaging of proteins as close as possible to their native conditions. Furthermore, the proteins immobilized inside this solid ice are protected from the vacuum inside the microscope while the overall cryogenic conditions provide important protection from beam damage.

The typical procedure (Fig.1.14) used to prepare a protein sample for imaging in vitreous ice consists of applying a small droplet (~2 µL) of a solution containing the protein (0.1-10 mg/mL) on top of a microscope grid which is pre-treated with glow discharge (a type of plasma) to remove contaminants and increase hydrophilicity. Then this is blotted to absorb the excess liquid leaving only a thin layer (~30-100 nm (111)) of solution on the grid. Then the grid containing the thin layer of the sample solution is rapidly plunged into liquid ethane which rapidly cools the solution forming a vitreous ice layer. The formation of vitreous ice requires

rapid cooling below its glass transition temperature (aprox.  $-135\text{ }^{\circ}\text{C}$  (112)) and therefore a cryogenic liquid with high heat capacity is needed. Among the cryogenics tested liquid ethane was shown to have the best relative cooling efficiency in cryo-EM studies (113).

As a note, I should also mention that blotting causes  $> 99.9\%$  loss of sample as most of the solution containing the protein ends up in the blotting pads. Furthermore, the process remains substantially manual and therefore significant sample variability is typically experienced. Combining automation with an alternative technique of sample application, such as jet spraying (114, 115), will allow minimization of sample loss and sample variability. However, currently, such automated devices remain prohibitively expensive for smaller labs.



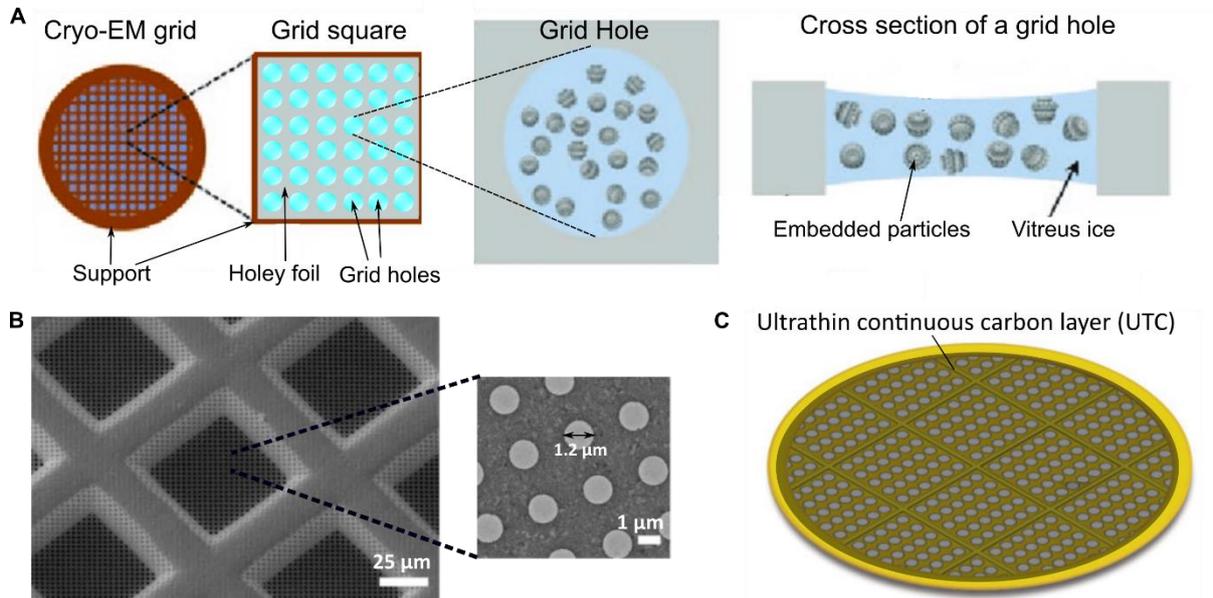
**Figure 1.14. a.** Sample preparation method in cryo-EM: a few  $\mu\text{L}$  of the solution containing the protein sample is applied on a microscope grid. Then the sample is blotted and rapidly plunged in liquid ethane. Image adapted from Märt-Erik Mäeots *et al.* (116) **b.** Photo of a cryo-EM grid (own image).

### *Notes on microscope grids*

Cryo-EM microscope grids are all composed of support covered by a thin holey foil with various geometries that affect ice thickness (Fig.1.15). Two main types of grids are available: grids with a gold holey foil applied on a gold support (e.g. UltrAuFoil®) and grids with a holey carbon film applied on a copper support (e.g. Quantifoil®). In addition, an ultrathin (2-3 nm) continuous carbon layer (UTC) can be applied to both grid types (Fig.1.15.c) to improve particle density and orientation by absorbing biomolecules across the entire hole, however, it also reduces image contrast.

Gold grids are significantly more expensive than copper ones ( $\sim 10\text{-}20\text{ }£/\text{grid}$  vs.  $\sim 5\text{-}8\text{ }£/\text{grid}$ ), however, they offer several advantages such as reduced beam-induced motion, reduced image

distortion from the accumulation of static, reduced beam-induced sample damage (positive charges are neutralized with secondary electrons generated by irradiating adjacent gold), and reduced mechanical strain from foil crinkling (117). Typically the use of gold grids improves the resolution of the 3D reconstructed map (118).

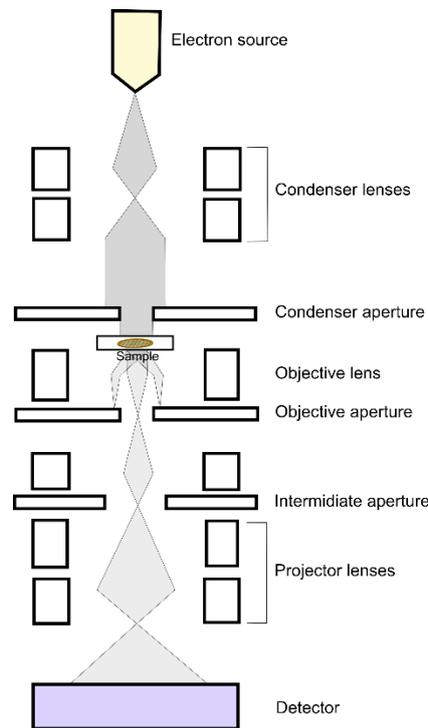


**Figure 1.15.** **a.** Cryo-EM grids are composed of metal support (gold or copper) which bars form a mesh that results in grid squares. A thin holey foil (gold or carbon) is applied on the support. The holes contain vitreous ice with embedded the particles to be imaged. Image adapted from Tiago R D Costa *et al.* (119). **b.** Scanning electron micrograph of a gold support and TEM image showing the perforated polycrystalline gold foil. Image adapted from Christopher J. Russo *et al.* (117). **c.** Ultrathin continuous carbon layer (grey) applied to a gold grid. Image adapted from Joel R Meyerson *et al.* (120).

#### 1.8.4 The electron microscope

Cryo-EM is a form of Transmission Electron Microscopy (TEM) in which a beam of electrons is transmitted through a specimen at cryogenic temperature to form an image. The beam of electrons is manipulated through electromagnetic lenses which can be compared by analogy to the optics of light microscopes. At the top of the microscope (Fig.1.16) there is an electron gun which generates an electron beam. Inside the column, where everything is under vacuum, a series of consecutive electromagnetic lenses shape and direct the electron beam onto the specimen. The interaction between the beam and specimen results in electron scattering events which are at the basis of contrast and therefore image formation (discussed later). The electrons that make it through the sample are again magnified by a series of electromagnetic lenses and

ultimately collected at the bottom of the column by a detector. The detector is typically a direct electron detector.



**Figure 1.16** Schematic representation of a typical TEM microscope. The electron beam (grey) is generated at the top of the microscope and is controlled by a series of electromagnetic lenses (squares). The first beam manipulation is performed by condenser lenses which parallelize it. The parallel beam hits the sample and the electrons that go through it are focused by objective lenses. Projector lenses then magnify and focus the image onto the electron detector.

### *Electron gun*

Field Emission Guns (FEGs) are the preferred choice in electron microscopes. These are composed of a sharply pointed emitter held at several kilovolts of negative potential relative to a nearby electrode. The electron beam produced by FEGs is smaller in diameter, more coherent and with greater brightness than conventional thermionic emitters such as tungsten filaments. Below the FEG is an accelerator stack that applies a fixed potential typically of 200 or 300 kV to accelerate the electrons to relativistic velocities.

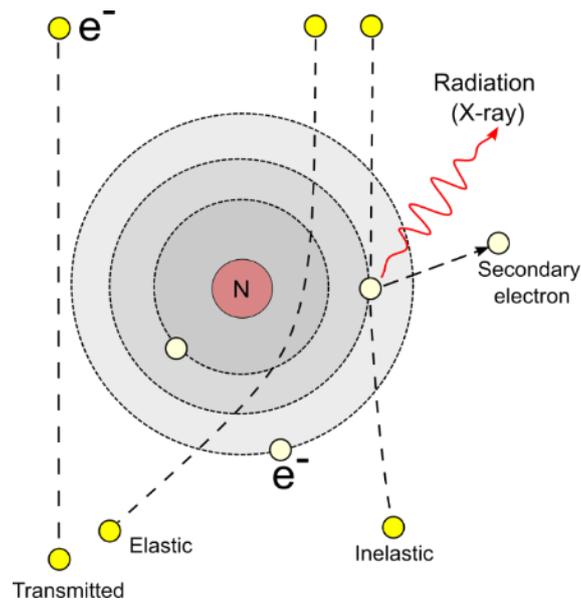
### *Detector*

Initially, projection images in TEM microscopes were recorded on photographic film making it a very time-consuming process. In the 1990s new types of detectors were introduced where a scintillator (e.g. a layer of phosphor) was combined with a charge-coupled device (CCD) camera resulting in straightforward digital image acquisition (electrons→photons→image).

However, the CCD detectors had a worse noise-to-signal ratio compared to photographic films due to electron scattering within the scintillator and charge-sharing between adjacent pixels (121). The subsequent introduction of direct electron detection devices significantly improved the resolution by directly converting electrons into electric current (electrons→image) (122). Direct electron detectors are composed of an array of sensitive pixels made of semiconductor layers (123); when a pixel is hit by electrons, the voltage of the diode diminishes proportionally to the amount of charge collected which, in turn, is proportional to the amplitude of the electron wave. This change in voltage is converted to a digital signal which results in a digital image where brighter pixels correspond to a higher amplitude of the electron wave.

### **1.8.5 Image formation in cryoEM: Electron interactions**

In cryoEM, an image is formed as the result of interactions between the incident electrons and the atoms of the specimen. Most of the electrons (90%) penetrate through the sample without any interaction while the remaining ones are scattered due to the interaction with the Coloumb potential of electrons and protons of the atoms in the specimen. The scattering can be elastic or inelastic depending on whether there is any energy loss during the interaction. If the electron interacts with the Coloumb potential of the nucleus (more than 1800 times heavier), then there is no energy loss and the electron bounces off with the same speed constituting an elastic scattering event. If instead, the electron interacts with other electrons in the atoms of the specimen, then some energy is lost in the form of radiation and the electron will be scattered inelastically (Fig.1.17). Inelastic scattering is particularly damaging to the sample due to secondary emissions (e.g., X-rays), ionization, or radicalization.



**Figure 1.17** Representation of elastic and inelastic electron scattering events. Interaction with the nucleus (N) of an atom results in elastic scattering while interaction with its electrons results in inelastic scattering.

Both elastic and inelastic scattering contribute to image formation in cryo-EM by providing contrast although with different mechanisms. Elastic scattering primarily results in phase contrast which contributes to most of the image contrast in cryo-EM (~90% (124)) while inelastic scattering only contributes to amplitude contrast where the electron is scattered to such an extent that does not reach the detector resulting in a dip in the intensity of the electron wave (number of electrons counted). The physical reason why amplitude contrast has a minimal contribution to the total contrast is that Cryo-EM specimens only scatter electrons weakly (few scattering events at low angles) due to being very thin (30-100 nm) and made of light atoms (H, N, C, O).

Furthermore, inelastic scattering in frozen amorphous samples is attributed to plasmons (collective excitation of valence electrons by the electric field of the imaging electrons) which tend to be delocalized (the extent of it is debated) and therefore it has a reduced contribution to high-resolution information compared to elastic scattering (125).

### ***Phase contrast***

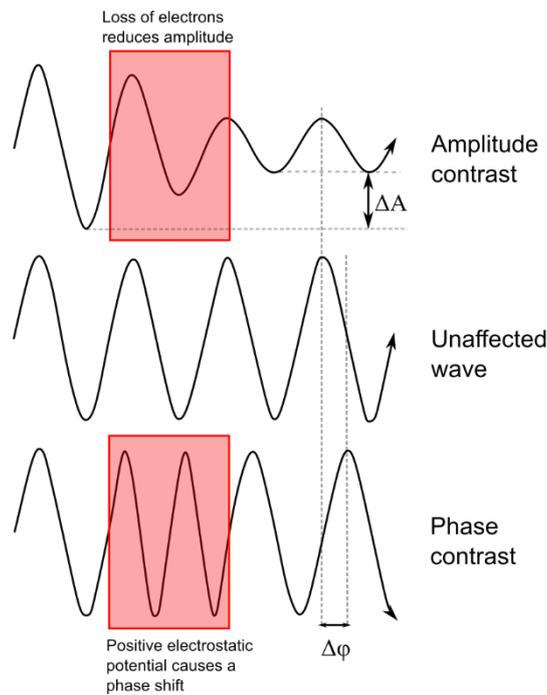
Phase contrast is caused by the phase shift that electrons experience when elastically scattered by their passage near the nucleus of the atoms in the sample (Fig.1.18). The longer the path of the electrons across the sample (i.e. thicker sample) and the higher the number of phase-shifting interactions. The total phase shift ( $\Delta\phi_{tot}$ ) is proportional to the mean inner potential of the bulk

material ( $V_0$ , a volumetric average of the electrostatic potential in a solid-state material (126)) and the sample thickness ( $d$ ) according to Formula 1.23 (127).

$$\Delta\varphi_{tot} = \sigma \cdot V_0 \cdot d \quad (1.23)$$

Where  $\sigma$  is a proportionality constant that only depends on the electron energy and has values of 0.73, and 0.65  $\text{mrad} \cdot \text{V}^{-1} \cdot \text{\AA}^{-1}$  for 200 and 300 keV electrons, respectively.

In particular, biological specimens have a higher inner potential (e.g. protein  $V_0 \sim 7.5\text{V}$ ) than the surrounding vitreous ice ( $V_0 \sim 5\text{V}$ ) causing the interacting electrons to have a more pronounced phase shift. This difference is at the basis of contrast formation as without it we would not be able to distinguish the signals coming from the biological sample (i.e. proteins, lipids, etc.) from the background signal of the vitreous ice.

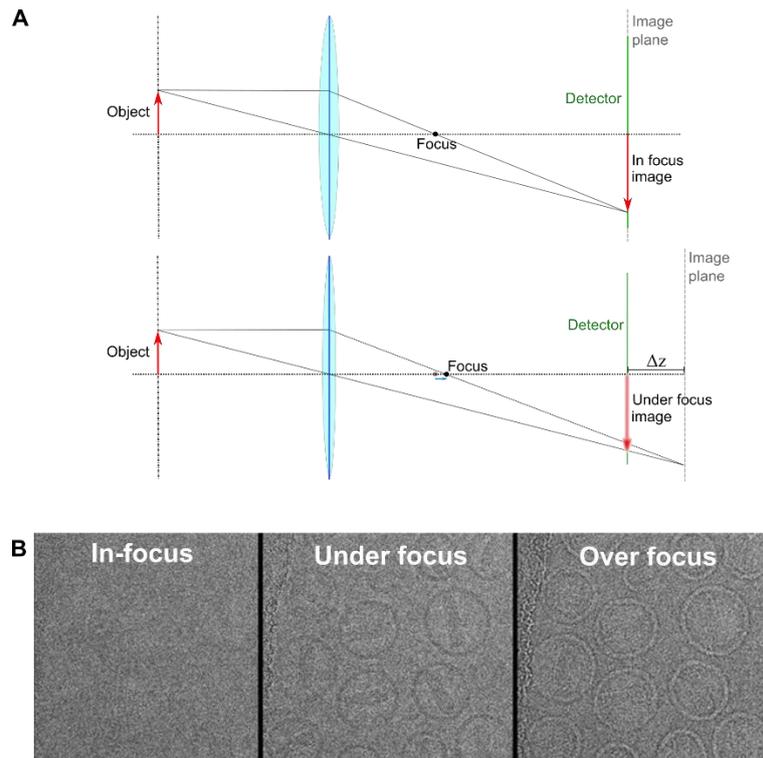


**Figure 1.18** Schematic showing the mechanism behind contrast. When the electrons are scattered to such an extent that they do not reach the detector a loss of signal can be observed that corresponds to a decreased amplitude in the electron wave. When electrons pass near an atomic nucleus they are elastically scattered and the corresponding electron wave is phase-shifted relative to the unaffected (corresponds to the unscattered) beam.

Directly measuring the phase shift of the electron waves exiting the sample is not possible (see Chapter 1.8.4, detectors), however, the constructive or destructive interference between the unscattered and elastically scattered waves results in a change in amplitude (intensity) which is possible to measure (125, 128). This change in amplitude due to wave interference will cause

the image of the sample to be “brighter” or “darker” in different regions depending on the amount of phase shift experienced.

To generate a measurable interference and therefore a phase contrast, the image needs to be collected out of focus (i.e. defocused for easier interpretation (128)) by slightly changing the objective lens current (Fig.1.19). This also causes a spread of information and thus blurring of the image at the expense of high-resolution structural information and therefore excessive defocus ( $\Delta z$ ) above  $2\mu\text{m}$  is generally avoided.



**Figure 1.19 a.** A detector positioned exactly at the image plane records an in-focus image. When the focal point is moved forward while keeping everything in the same position the detector records an under-focus image; in electron microscopy, this can be achieved by reducing the magnetic field of the magnetic lenses **b.** In focus, under focus, and over focus cryo-EM micrographs of Doxil, an anticancer drug composed of liposome packaged doxorubicin hydrochloride crystals. In-focus images only have amplitude contrast and therefore most of the contrast contribution (phase contrast) is lost. Image adapted from Linda E. Franken *et al.* (128).

### 1.8.6 Image formation in cryoEM: CTF function (pre-concepts)

The electrons transmitted after interaction with the sample are focused on the image plane (Fig.1.19.a). The position of the image plane relative to the position of the detector can be adjusted by changing the amount of current in the magnetic lenses. When the position of the

image plane coincides with the position of the detector the image recorded is in focus, otherwise, it is out of focus by a distance  $\Delta z$ .

As mentioned before, each sensor pixel in the detector can only measure the amplitude of the electron wave (see Chapter 1.8.4, detectors). Therefore to produce an information-rich image, different features of the sample must be differentiated by different localized amplitudes (intensities) of the overall recombined electron wave (scattered + unscattered components). This difference in amplitudes is primarily modulated by the phase difference between the unscattered and elastically scattered electron waves (phase contrast); when two waves with different phases recombine, the resulting wave will have a different amplitude. This phenomenon results in a transfer of contrast which can be negative or positive depending on whether there is constructive or destructive interference. The transfer of contrast is mathematically modelled by a contrast transfer function (CTF), however, before moving into its definition, some pre-concepts need to be introduced, namely the exponential representation of wave functions and the Fourier transformation.

### ***Exponential representation of wave functions***

A wave is a periodic function in time and space. For example, in an electromagnetic wave, the electric and magnetic field oscillates periodically while in an electron wave, it is the probability amplitude of finding the particle that oscillates. In particular, the wave function of an electron  $\psi(x, t)$  can be derived by solving the Schrodinger equation which solutions are complex sine and cosine waves:

$$\psi(x, t) = \psi_0 \left[ \cos\left(\frac{2\pi}{\lambda}x - 2\pi ft + \varphi\right) + i \cdot \sin\left(\frac{2\pi}{\lambda}x - 2\pi ft + \varphi\right) \right] \quad (1.24)$$

Where  $\psi_0$  is the amplitude,  $\varphi$  is the phase, and  $f$  is the frequency (velocity of the particle divided by its wavelength). In particular, the probability density  $\rho(x, t)$  of finding an electron in a location at a specific time is related to the value of  $\psi(x, t)$  (Formula 1.25), and therefore the charge felt by each detector pixel is proportional to the amplitude  $\psi_0$ .

$$\rho(x, t) = |\psi|^2 \quad (1.25)$$

The above wave function (Formula 1.24) can also be equivalently expressed as a complex exponential function:

$$\psi(x, t) = \psi_0 e^{i\left(\frac{2\pi}{\lambda}x - 2\pi ft + \varphi\right)} \quad (1.26)$$

This is justified by Euler's formula:

$$e^{ix} = \cos(x) + i \cdot \sin(x) \quad (1.27)$$

Similarly, the oscillating electric field in an electromagnetic wave can be expressed through a complex exponential, however, in this case, only the real (Re) part is considered:

$$E(x, t) = E_0 \cos\left(\frac{2\pi}{\lambda}x - 2\pi ft + \varphi\right) = E_0 \operatorname{Re}\left(e^{i\left(\frac{2\pi}{\lambda}x - 2\pi ft + \varphi\right)}\right) \quad (1.28)$$

Where  $E_0$  is the amplitude,  $\varphi$  is the phase, and  $f$  is the frequency (speed of light divided by the wavelength of the photon).

### ***Fourier transformation***

The Fourier Transform is an important image-processing tool that is used to decompose an image into a sum of complex exponentials, which represent sine and cosine waves (see above) of varying magnitudes, frequencies, and phases. The output of the transformation represents the image in the special frequency domain (also called Fourier domain), while the input image is the spatial domain equivalent (i.e. each pixel of a photo represents a specific spatial position). In particular, any image collected by a detector is discrete because it is composed of pixels and therefore a discrete Fourier transform is required. For simplicity, here only the continuous Fourier transform is presented as the fundamental concepts are similar.

The Fourier transform  $F(w_1, w_2)$  of a function  $f(x, y)$  of two spatial variables  $x$  and  $y$  is:

$$F(w_1, w_2) = \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{\infty} f(x, y) e^{-iw_1x} e^{-iw_2y} dx dy \quad (1.29)$$

Where  $w_1, w_2$  are spacial frequencies which are the coordinates of a 2D Fourier transform image. The Fourier transform image can be back converted into its original spatial domain image without loss of information (inverse Fourier transform, Formula 1.30).

$$f(x, y) = \frac{1}{4\pi^2} \int_{w_1=-\infty}^{\infty} \int_{w_2=-\infty}^{\infty} F(w_1, w_2) e^{iw_1x} e^{iw_2y} dw_1 dw_2 \quad (1.30)$$

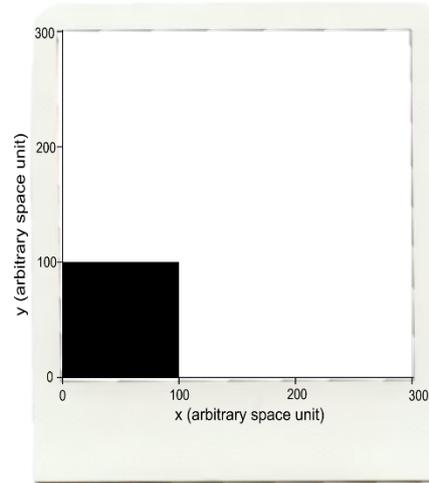
Where the  $\frac{1}{4\pi^2}$  can be demonstrated to derive from the expression of the delta Dirac function.

Two examples are provided to understand this transformation:

#### ***Example 1***

Imagine we have a non-discrete photo (i.e. infinitely high resolution) of a black square on a perfectly white background (Fig.1.20). The coordinates  $x$  and  $y$  are spatial coordinates as they

represent a defined length in the real object of the photo. The square can also be thought of as the function representing a single black pixel.



**Figure 1.20** photo of a black square on the corner of a white background. The coordinates  $x$  and  $y$  are in arbitrary space units (e.g.  $\mu\text{m}$ ,  $\text{mm}$ ,  $\text{cm}$ , etc.).

This photo can be described by a simple function  $f(x, y)$  that is equal to 1 (black) when  $x$  and  $y$  are between the values 0 and 100 (arbitrary spatial units) and everywhere else  $f(x, y) = 0$  (white). This is also the 2D version of a rectangular function.

Now we use Formula 1.29 to find the Fourier transform function  $F(w_1, w_2)$ :

$$F(w_1, w_2) = \int_{x=0}^{100} \int_{y=0}^{100} 1 \cdot e^{-iw_1x} e^{-iw_2y} dx dy = \int_{x=0}^{100} e^{-iw_1x} dx \int_{y=0}^{100} e^{-iw_2y} dy \quad (1.31)$$

We solve each integral term separately:

$$\begin{aligned} \int_{x=0}^{100} e^{-iw_1x} dx &= -\frac{1}{iw_1} e^{-iw_1x} \Big|_0^{100} = -\frac{1}{iw_1} (e^{-i100w_1} - 1) = -\frac{1}{iw_1} e^{-i50w_1} (e^{-i50w_1} - e^{i50w_1}) = \\ &= -\frac{1}{iw_1} e^{-i50w_1} [\cos(-50w_1) + i \cdot \sin(-50w_1) - \cos(50w_1) - i \cdot \sin(50w_1)] = \\ &= -\frac{1}{iw_1} e^{-i50w_1} [-2i \cdot \sin(50w_1)] = \frac{2}{w_1} \sin(50w_1) e^{-i50w_1} \end{aligned} \quad (1.32)$$

Similar steps can be followed to solve the other integral in  $y$  obtaining:

$$\int_{y=0}^{100} e^{-iw_2y} dy = \frac{2}{w_2} \sin(50w_2) e^{-i50w_2} \quad (1.33)$$

By substituting Equations 1.32 and 1.33 in Formula 1.31 we can derive the Fourier transform function  $F(w_1, w_2)$ :

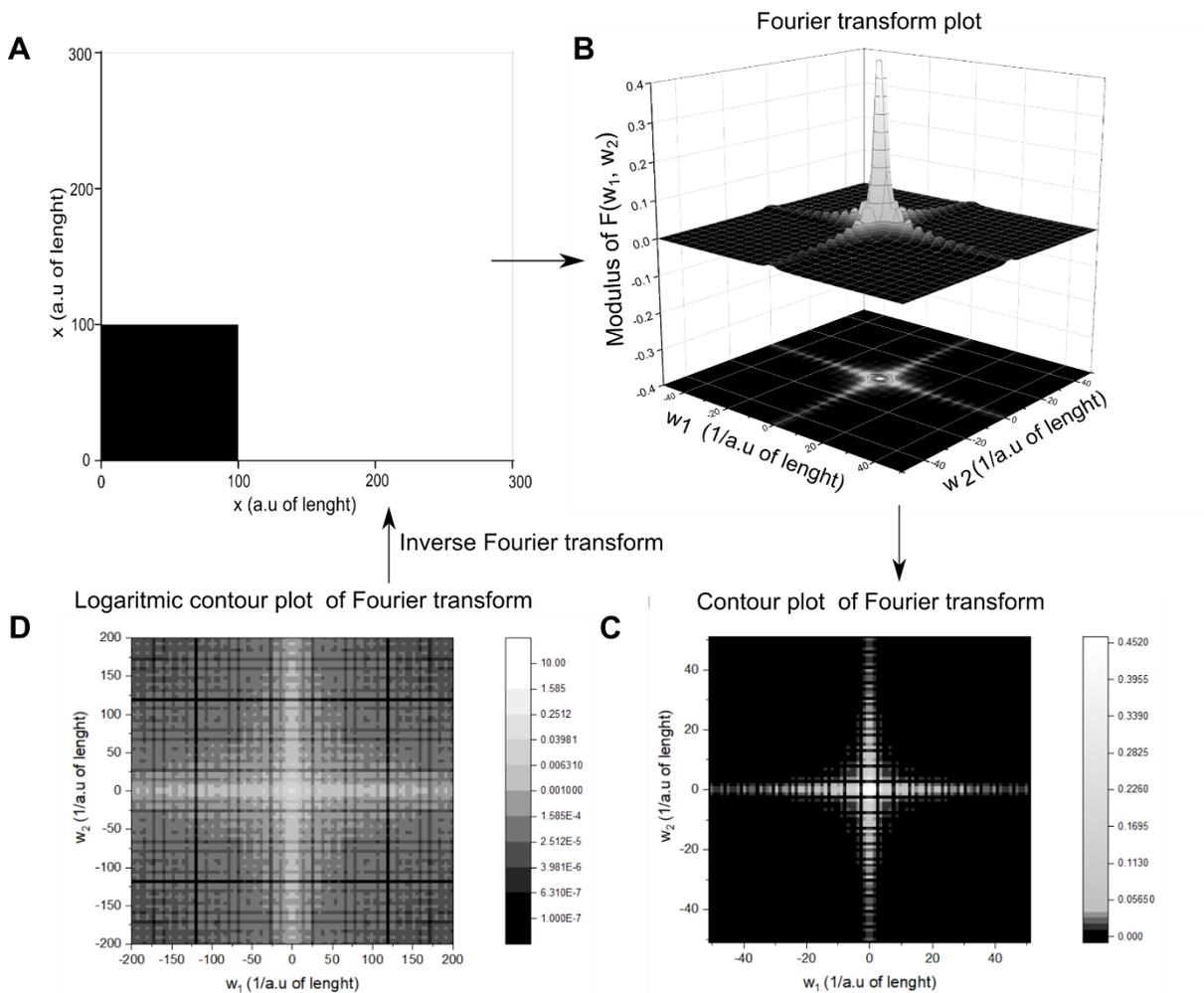
$$F(w_1, w_2) = \frac{4}{w_1 w_2} \sin(50w_1) \cdot \sin(50w_2) \cdot e^{-i50(w_1+w_2)} \quad (1.34)$$

The modulus of  $F(w_1, w_2)$  can be plotted in a 3D graph that has  $w_1$  and  $w_2$  as variables (called spatial frequencies) instead of the spatial coordinates  $x$  and  $y$  (Fig.1.21.b). We, therefore, need to calculate the modulus of  $F(w_1, w_2)$ :

$$|F(w_1, w_2)| = \text{abs}\left(\frac{4}{w_1 w_2} \sin(50w_1) \cdot \sin(50w_2)\right) \cdot |e^{-i50(w_1+w_2)}| = \text{abs}\left(\frac{4}{w_1 w_2} \sin(50w_1) \cdot \sin(50w_2)\right) \quad (1.35)$$

Now we can plot  $|F(w_1, w_2)|$  as a function of the spatial frequencies  $w_1$  and  $w_2$  (Fig.1.21.b).

The 2D contour plot (Fig.1.21.c) is typically referred to as Fourier transform image and it is typically plotted with the modulus in a logarithmic scale (Fig.1.21.d).



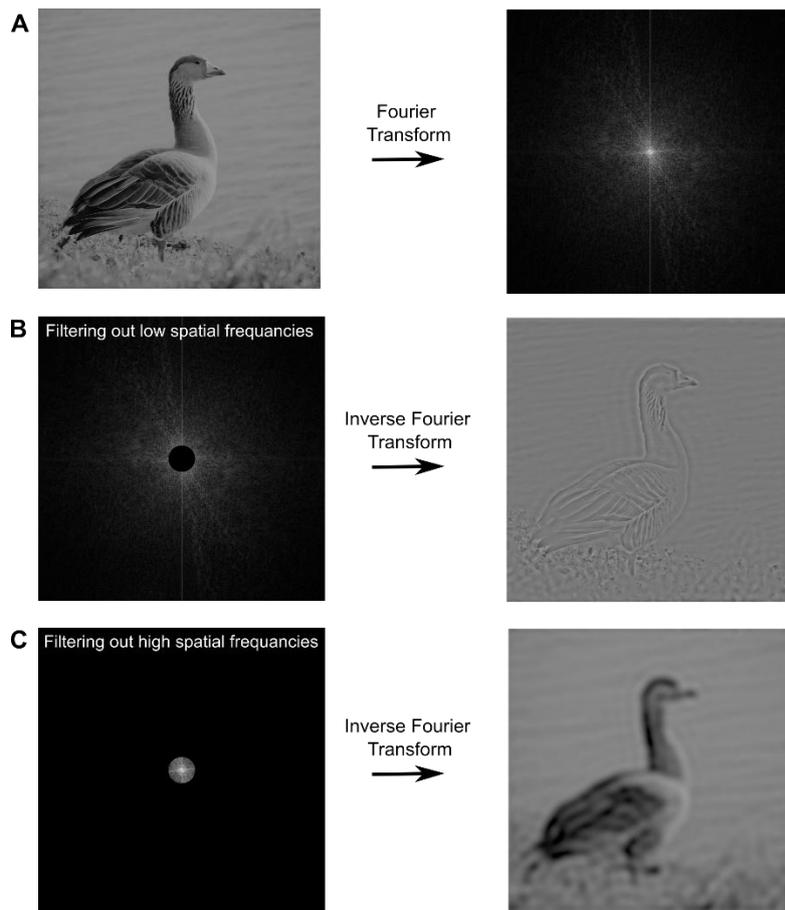
**Figure 1.21.a.** Spatial domain representation of a black square in a white background (e.g.a photo). **b.** Fourier transform plot of the black square image based on Formula 1.35. **c.** Contour plot of the Fourier transform function. **d.** Contour plot of the Fourier transform function with the modulus in a logarithmic scale. Note that the values of the spatial frequencies  $w_1$  and  $w_2$  go from  $-\infty$  to  $+\infty$ , however, the contour plot is only represented up to a certain range. **Note:**

*Changing the position of the square in  $a$  does not change the Fourier plot when the modulus is used. This can be easily demonstrated, however, it's outside of the scope.*

In a “complete” Fourier transform image the values of the spatial frequencies  $w_1$  and  $w_2$  go from  $-\infty$  to  $+\infty$ , however, the modulus of the Fourier transform gets progressively smaller and therefore lower frequencies (in absolute value) contain more image information than the higher ones. This justifies representing the Fourier transform image up to a limited range of spatial frequency values.

### *Example 2*

In this example, we take a picture of a goose and perform a discrete Fourier transform using image processing algorithms (129). As shown in Fig.1.22.a the Fourier transform image (contour plot of the logarithm of the modulus of the Fourier function) contains high and low spatial frequencies (in absolute value). When the low spatial frequencies are filtered out then most of the image information is lost (Fig 1.22.b). On the other hand, when the high spatial frequencies are filtered out (Fig.1.22.c) then the high-resolution features are removed but most of the image information is preserved.

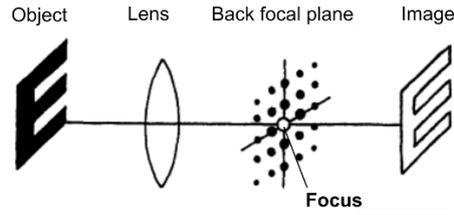


**Figure 1.22.a.** Photo of a goose (left) and its Fourier transform image (right). **b.** Photo of a goose (right) obtained by filtering out the low spatial frequencies information. **c.** Photo of a goose (right) obtained by filtering out the high spatial frequencies information.

### 1.8.7 Image formation in cryoEM: CTF function

Now that the exponential representation of wave functions and the Fourier transformation have been introduced we will better understand the CTF function.

When the electrons pass through the objective lens a Fourier transformation and an additional phase shift occur. In fact, there is a “Fourier Transform” relationship between the front and the back focal plane of a simple lens (Fig.1.23).



**Figure 1.23** The back focal plane image is the Fourier transform of the image. Image adapted from N. Hampp et al. (130)

As such, when only phase contrast is considered, the wavefunction on the back focal plane of the objective lens ( $I(\theta)$ ) can be represented as:

$$I(\theta) = \delta(\theta) + \Phi K(\theta) \quad (1.36)$$

Where  $\theta$  is the scattering angle between the transmitted electron wave and the scattered electron wave which is proportional to the spatial frequencies ( $\omega$ ):

$$\theta = \lambda \cdot \omega \quad (1.37)$$

$\delta(\theta)$  is a delta function representing the non-scattered, transmitted, electron wave.  $\Phi$  is the Fourier transform of the wavefunction's phase.  $K(\theta)$  is the phase shift incurred by the microscope's aberrations, also known as the Contrast Transfer Function (CTF).

$$K(\theta) = \sin\left(\frac{2\pi}{\lambda} \left[ \frac{C_s}{4} \cdot \theta^4 - \frac{\Delta z}{2} \cdot \theta^2 \right]\right) \quad (1.38)$$

Where  $\Delta z$  is the defocus and  $C_s$  is the spherical aberration of the objective lens (the outer parts of a lens do not bring the electrons into the same focus as the central part).

Using equation 1.37 and substituting it in equation 1.38 we can also derive the CTF as a function of spatial frequencies:

$$K(\omega) = \sin\left(\frac{2\pi}{\lambda} \left[ \frac{C_s}{4} \cdot \lambda^4 \omega^4 - \frac{\Delta z}{2} \cdot \lambda^2 \omega^2 \right]\right) \quad (1.39)$$

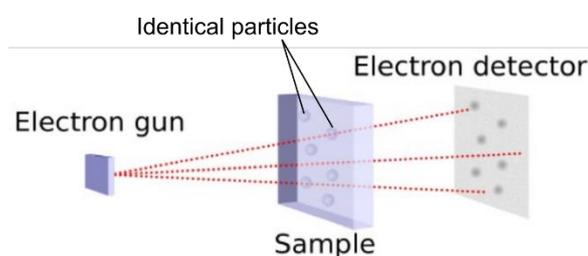
Therefore the wavefunction on the back focal plane of the objective lens ( $I(\theta)$ ) requires that  $K(\theta)$  is not null, otherwise, there would not be any contrast ( $I(\theta) = \delta(\theta)$ ). Furthermore the CTF function ( $K(\theta)$ ) shows that both defocus and spherical aberration of the lenses are necessary to produce contrast; if both are null then also  $K(\theta)$  is null resulting in no contrast and therefore no information can be obtained on the specimen imaged.

Note that more complex forms of the CTF function can be obtained when amplitude contrast and additional aberrations are taken into consideration.

### 1.8.8 Image processing: Single particle reconstruction

Protein molecules, whole viruses or other small individual structures such as liposomes can be imaged in vitrified ice using TEM microscopes. To include all these options the general term “particles” is typically used.

In order to reconstruct a 3D model (which is a density map) of a particle, multiple views and thousands of images of each view are required. Due to these constraints, only particles that are all identical to each other can be 3D reconstructed; this is possible for a purified protein or virus where a single microscope grid contains many thousands of these identical particles (Fig.1.24).



**Figure 1.24** Cryo-EM samples containing thousands of identical particles are imaged to obtain a 3D reconstruction. Image adapted from David Střelák *et al.* (131).

#### ***Reconstruction of 3D models of proteins***

When a vitrified solution containing the purified protein target is imaged, the resulting micrograph will be difficult to interpret because it is defocused and very noisy as low electron doses are used to minimize sample damage. Because of the low signal-to-noise ratio, thousands of protein molecules need to be imaged and averaged together to visualize meaningful features. Therefore, the protein should not be excessively flexible as this could result in far too many conformations and therefore not enough identical particles available to be averaged.

In addition, to reconstruct the 3D model, multiple views from different angles of the target protein are required. When protein molecules are randomly oriented within the ice layer multiple views will be spontaneously obtained, however, this is not always possible; in this latter case, the particles assume a preferential orientation and strategies such as the addition of detergents to cover the air-water interface during grid preparation or tilting the sample stage relative to the electron beam are utilized. In particular, imaging by tilting the stage has major limitations including longer data collection times and limited resolution caused by the lower

electron doses required to minimize electron damage (multiple exposures of the same sample at different angles and longer electron trajectories through the sample).

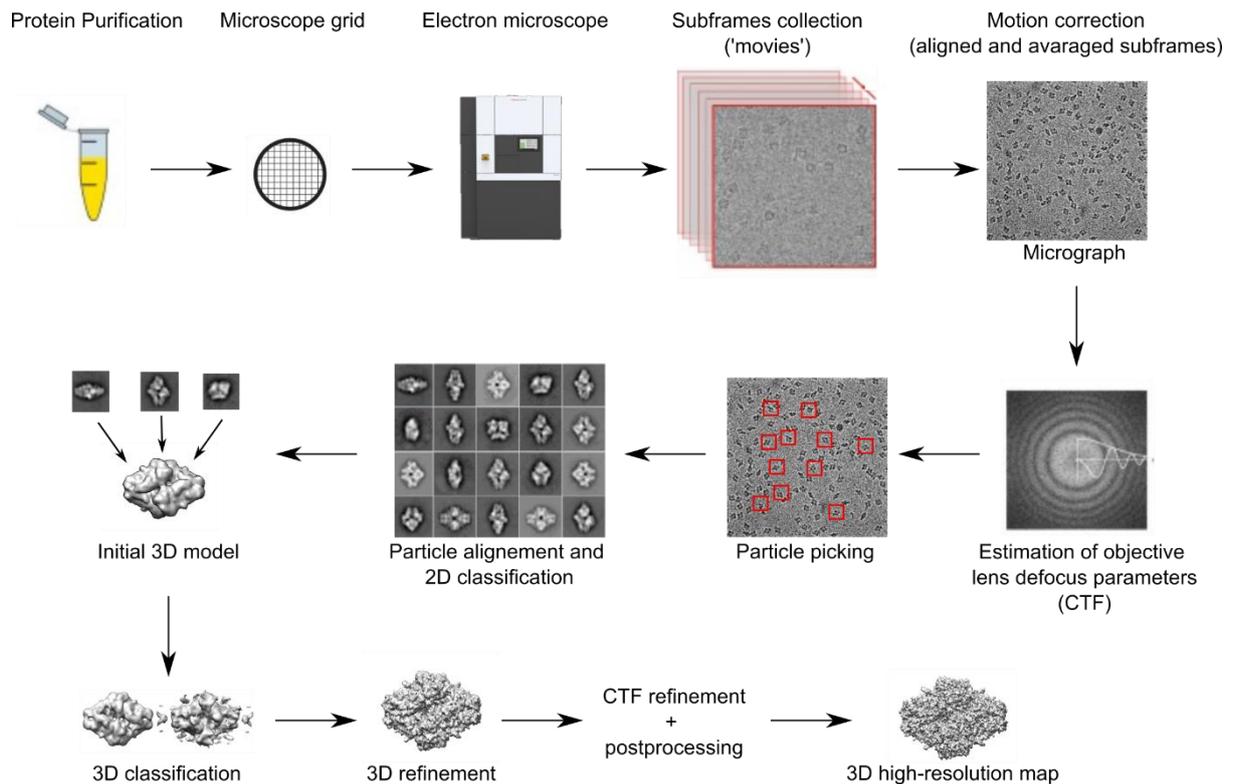
### ***Summary of requirements for single-particle reconstruction***

The reconstruction of a 3D model utilizing thousands of images of particles in multiple orientations is called single-particle reconstruction. The following requirements are essential for generating a 3D reconstruction:

1. Thousands of identical target particles need to be imaged.
2. The target particle should have a limited number of conformations.
3. Multiple views at different angles of the target particle are required; this is typically achieved when the particle is randomly oriented within the ice layer.

### **1.8.9 Image processing: Single particle reconstruction pipeline**

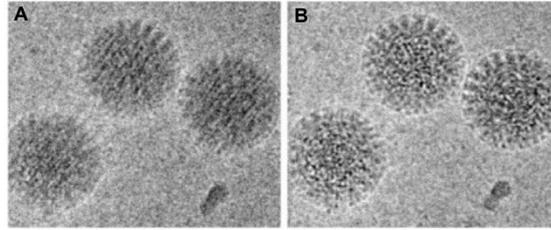
Reconstructing the 3D model of proteins from cryo-EM micrographs involves complex and computationally heavy image manipulations typically consisting of the following steps: motion correction, CTF estimation, particle picking, 2D classification, initial model, 3D classification, 3D refinement, CTF refinement, and post-processing (Fig.1.25).



**Figure 1.25** Cryoem image single particle reconstruction pipeline. Image adapted from Marta Carroni *et al.* (132).

### ***Motion correction***

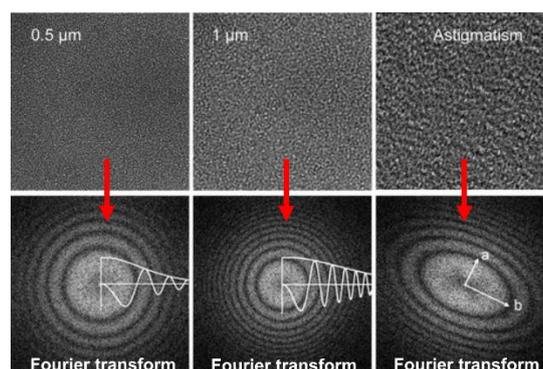
The image processing begins with the correction of the blurriness effect caused by the small unwanted movements of the stage (whole frame basis) and beam-induced local motions (per-particle basis). Beam-induced motions are caused by the deformation of the support foil and the ice layer when specimens are exposed to the electron beam which is thought to be driven by a build-up of positive charge on the specimen and radiolysis of the sample and vitrified embedding ice (133, 134). To allow for the correction of all these movements, modern direct electron detector cameras record a high frame rate (10–40 frames per second) 'movie' throughout the exposure; with this strategy, a single exposure is fractionated into a number of subframes with a sufficiently short duration such that the motions are negligible in each of them. A software package, such as MotionCor2 (135), then analyses these frames and estimates the local trajectories of the sample remapping each subframe. The subframes are then summed with a radiation-damage weighting to produce a motion-corrected micrograph (Fig.1.26).



**Figure.1.26** Close-up of a micrograph obtained by averaging the subframes of a movie with (A) and without motion correction (B). Motion correction allows for the recovery of high-resolution information. Image from Marta Carroni *et al.* (132).

### *CTF estimation*

The contrast of images recorded in cryo-EM is affected by lens aberrations and defocus (Chapter 1.8.7). The relationship between these factors and the contrast in the image is defined by the contrast transfer function (CTF) which needs to be accurately estimated to reconstruct a high-resolution 3D model. Softwares such as *ctffind4* (136) analyze the Fourier transform of motion-corrected micrographs and estimate lens aberrations and defocus. The Fourier transform of a cryo-EM micrograph is made of concentric rings called “Thon rings” (Fig.1.27), which are a distinctive feature of images obtained with electron microscopes. Thon rings can be fitted with the CTF function  $K(\omega)$  (Formula 1.39) and its parameters are estimated by cross-correlating them against a database of images with known CTF parameters. The dark and light rings represent spatial frequencies with negative and positive contrast transfer, respectively, as predicted by the oscillating sine CTF function (Formula 1.39).



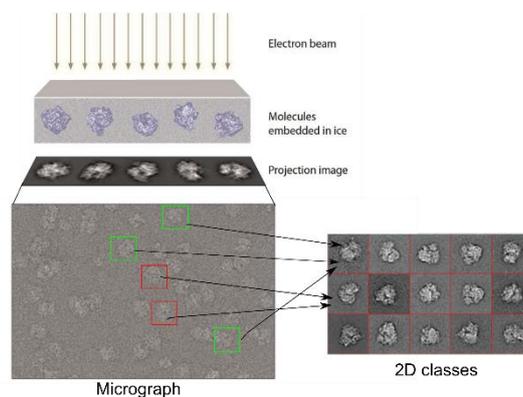
**Figure 1.27.** Images of carbon film micrographs, showing Thon rings and corresponding CTF curves. The left image was obtained at 0.5  $\mu\text{m}$  defocus, and the middle image was at 1  $\mu\text{m}$  defocus. The rings alternate between positive and negative contrast, as seen in the plotted CTF curves. On the right, is an example of an astigmatic image with characteristically distorted rings. Image adapted from E. V. Orlova *et al.* (137).

## ***Particle picking***

Detecting the locations of the particles in the micrographs is known as “particle picking”. Particle picking is performed automatically with or without a reference. If data of a similar protein to the target one are available then a reference can be constructed using previous 2D classes or maps generated from a PDB structure. Using a reference allows more precise picking. Reference-free particle picking can be performed by algorithms based on the Laplacian of Gaussian operator (138), where, contours around objects that are surrounded by a relatively consistent background are identified. Typically the particle-picking finds ten to hundred thousand particles in all the micrographs collected (typically hundreds to a few thousand micrographs are collected) and these images are cropped and stored for the next stages of image processing.

## ***2D classification***

After the particles are picked, they are classified based on orientation. The images of particles that have the same orientation are averaged together to produce an averaged image of the 2D class. In this stage, the user can select only the “sensible” 2D classes eliminating all the wrongly picked particles (e.g. noise, contaminants, etc.). The 2D classification can be repeated iteratively with the selected classes to minimize the presence of images that are not from the intact target protein. 2D classification is based on algorithms that rotationally and translationally align the images of particles. The images with nearly the same projection directions are then grouped in the 2D classification step (139) (Fig.1.28).



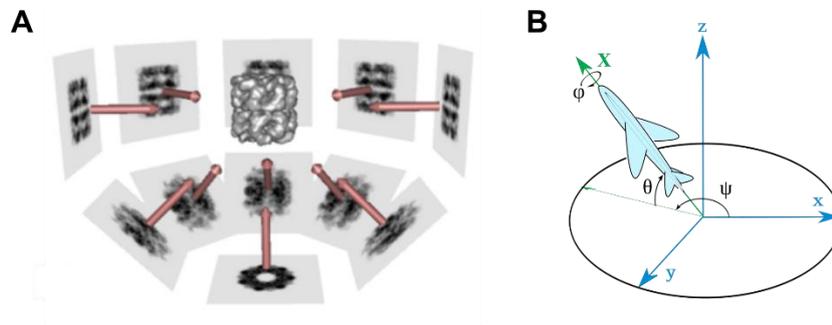
**Fig.1.28** The micrographs record a 2D projection of the particles. Images of particles representing the same projection are aligned and averaged to produce a 2D class image. Image adapted from Xabier Agirrezabala *et al.* (140).

### ***Initial model and 3D classification and 3D refinement***

The particles in the selected 2D classes can be used to reconstruct a 3D map of the target protein (Fig.1.29.a). Initially, an approximate 3D model is generated which is used in the next stage of 3D classification where the selected particles are classified into 3D classes. 3D classification is useful to further eliminate images of damaged proteins and to discriminate different conformations when these are present. The particles constituting a selected 3D class can then be used to produce a better 3D model in the 3D refinement stage.

To reconstruct a 3D model, a series of iterations are used starting with an initial guess of the Euler angles which describe the three-dimensional orientation of an object, Fig.1.29.b. The projections of the 3D model produced at the end of an iteration are used in the next iteration to correct the assigned Euler angles improving the model. When the Euler angles stop changing over iterations, a stable solution is reached which provides the final reconstruction.

It should be also noted that the relationship between the reciprocal of the resolution and the number of particles used to build a 3D model is proportional to the logarithm of the particle number (141).



**Figure 1.29.a.** 3D reconstruction from 2D class averages. **b.** Euler angles are used to describe the orientation of an object. Image adapted from Wikipedia (142).

### ***CTF refinement and postprocessing***

The 3D refined model can be further improved by re-estimating CTF parameters. Once higher resolutions of the 3D map are achieved ( $<4.5 \text{ \AA}$ ) as happens in 3D refinements, further aberrations of the microscope can become limiting (only defocus and large values of astigmatism have a significant impact at low resolution). Therefore a CTF refinement job can be used to further improve resolution. Examples of such additional aberrations are beam tilt, 3<sup>rd</sup> (trifoil) and 4<sup>th</sup> order aberrations. Additional postprocessing is typically needed to produce

the final 3D model which involves adding a mask to cover low threshold noise and map sharpening to highlight detailed features.

### **1.8.10 Resolution assessment**

The physical resolution limit of an electron microscope is determined by multiple factors from lens aberrations, stability of the microscope, electron wavelengths, sample nature, etc. One very important limiting factor is the sampling interval of the digital direct electron detector camera. The sampling interval is determined by the 'pixel size', which refers to the edge length of the sample area that gets projected onto each pixel (the larger the magnification of the microscope the smaller the 'pixel size'). To distinguish two points as individuals within a digital image, at least one empty pixel has to be between them. From this, the Nyquist frequency can be determined (Nyquist-Shannon Sampling Theorem), which is twice the pixel size and this value represents a threshold for the highest possible resolution achievable when all other factors are neglected. For example, if the pixel size of a set of micrographs is 0.60 Å/pixel then a resolution threshold of 1.2 Å is obtained simply by the intrinsic properties of digital cameras. This just calculated threshold is close to the resolution limit set by the electron wavelength in typical TEM microscopes (Chapter 1.8.2) and is therefore not limiting, however, an excessively high 'pixel size' (too low magnification) will make the camera the limiting factor in the achievable resolution.

The resolution limit set by the Nyquist frequency (twice the pixel size) is also useful as a comparison to the obtained resolution in the 3D reconstructed map. The closer the obtained resolution is to the Nyquist frequency the better the performance of the microscope and the image processing algorithms.

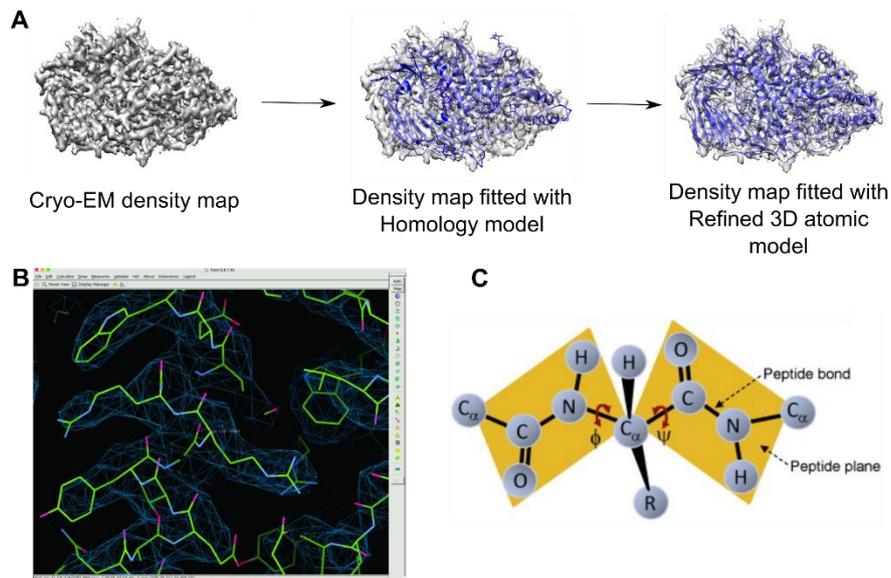
However, measuring the overall resolution of a 3D reconstructed map is not trivial. The resolution is a local property with some parts of the molecule more resolved than others. In the less resolved parts it is difficult to measure the resolution by observing distinct features and therefore a different mathematical strategy is needed. The solution is provided by the Fourier Shell Correlation (FSC). To perform an FSC the dataset is divided into two equal stacks, each of which is reconstructed into a 3D map. Both these structures are transformed into Fourier space and all spacial frequency shells are correlated. For low special frequencies, the correlation will be maximal (close to 1). With increasing special frequency, the correlation

declines and usually shows almost no correlation for Nyquist frequency. The most commonly used resolution is determined as the resolution where the FSC drops below 0.143 (143).

### **1.8.11 Building a 3D atomic model**

Once a 3D density map has been obtained, this can be used to build a 3D atomic model of the protein with all the amino acids and their side chains in place. This is also commonly referred to as “pdb structure” from “pdb” which is the file format containing all the atomic coordinates of the model. To build this atomic model, the protein sequence of the target protein is needed. This sequence is provided to online servers such as Phyre2 (144) and alphaFold (145) which produce an estimated atomic structure utilizing the ‘pdb structures’ of proteins with similar sequences available in public databases (RCSB PDB). All proteins have a degree of similarity to some other proteins as they are linked by evolution. This allows for the generation of these initial atomic models, also called homology models.

These initial atomic models are then fitted into the 3D density map generated from the cryo-EM dataset and manually manipulated to better fit the map (Fig.1.30.a). This manipulation is performed using software such as Coot (146) (Fig.1.30.b) which allows for the visualization of the 3D density map and aminoacid chain while providing additional features such as atomic clashes and Ramachandran analysis. Software such as Phenix (147) can then be used to automatically refine this manually constructed structure to minimize atomic clashes and reduce the number of amino acids with prohibited Backbone dihedral angles (Fig1.30.c) as predicted by the Ramachandran plot. Typically several rounds of refinements are used, where manual (Coot) and automatic refinement (Phenix) alternate until the stats (mainly clashes and Ramachandran outliers) stop improving.



**Figure 1.30. A.** The 3D atomic model is obtained by fitting and refining the homology model into the density map obtained from cryo-EM data. The image is purely for illustrative purposes with the density map generated from PDBID:1HTY **B.** An atomic model (green) fitted into the cryo-EM density (blue lines) map using Coot. The amino acid chain and side chains need to be manually moved to better fit the density map. Image taken from Toshio Moriya *et al.* (148) **C.** Dihedral angles  $\phi$  and  $\psi$  in a peptide chain and Ramachandran plot. Image adapted from Supratim Choudhuri (149).

## 2 Materials and methods

### 2.1 *Thermotoga maritima* hydrogenase methods

#### 2.1.1 Protein expression and purification

Protein expression and purification of this enzyme were carried out by a collaborator's lab (James Birrel). The details of the enzyme preparation are relevant to the discussion and are therefore here reported.

*Thermotoga maritima* hydrogenase (*TmHydABC*) was expressed heterologously in *E. coli* BL21(DE3)  $\Delta$ iscR and purified under anaerobic conditions, generating an 'apo' enzyme, containing all of the [2Fe–2S] and [4Fe–4S] clusters, but lacking the [2Fe]<sub>H</sub> subcluster of the H-cluster (the H-cluster is a core part of the active site, see Fig.1.3.b and Fig.3.22) in HydA (150, 151). In one preparation, the H-cluster was reconstituted using a synthetic [2Fe]<sub>H</sub> precursor (150, 152, 153) as *E. coli* lacks the required maturation proteins. Only minor structural differences were previously observed upon incorporation of the [2Fe]<sub>H</sub> subcluster (154) and the enzyme was shown to be active toward both oxidation and reduction of hydrogen (150).

Both preparations, with and without the [Fe]<sub>H</sub> subcluster, involved production under anaerobic growth conditions and purification in an anaerobic glovebox (Coy, 2% H<sub>2</sub> in N<sub>2</sub>) using Streptactin (IBA) affinity chromatography and size-exclusion chromatography (GE Healthcare) as previously described (150). Sample purity and quality were checked by sodium dodecyl sulfate–polyacrylamide gel electrophoresis and UV–vis spectrophotometry. Samples in 10 mM Tris–HCl, 150 mM NaCl, pH 8 were sealed inside bottles in the anaerobic glovebox to prevent oxygen exposure. These samples were then frozen at –80°C and subsequently transported in liquid nitrogen to our labs in York.

The protocol for the preparation of the enzyme without the [Fe]<sub>H</sub> subcluster was also used to produce purified reconstituted HydB subunits, which were used for inductively coupled plasma mass spectrometry (ICP-MS) analysis.

### **2.1.2 Inductively coupled plasma mass spectrometry**

Our cryo-EM analysis of apo-*TmHydABC* revealed a cofactor site in the HydB subunit of unclear nature and therefore a purified HydB sample was sent for metal analysis using inductively coupled plasma mass spectrometry (ICP-MS). A sample of the purified HydB subunit prepared as described above, was buffer exchanged into 10 mM MOPS pH 7 and concentrated to 621 μM, and a sample of 10 mM MOPS pH 7 was measured by Mikroanalytisches Laboratorium Kolbe (<https://www.mikro-lab.de/>). The samples were digested using a CEM Model MARS6 microwave digestion unit and measured on an Agilent Model 7900 ICP-MS.

### **2.1.3 Grid preparation and imaging**

#### ***TmHydABC without the [2Fe]<sub>H</sub> subcluster (apoenzyme)***

1.2/1.3 UltrAuFoil grids were glow discharged (PELCO easiGlow) for 90 s on each side using atmospheric gas before mounting in Vitrobot (model IV) tweezers (Thermo Fisher Scientific). Grids were prepared with minimal exposure to air using anaerobically frozen aliquots of ‘apo’ *TmHydABC*. These were individually defrosted and used. In this manner, *TmHydABC* was exposed to the air for a few seconds. The enzyme (without the highly oxygen-sensitive [Fe]<sub>H</sub> subcluster) seems to be stable under air for at least a few hours, determined as there were no visible spectral changes when the enzyme solution was exposed to air. Individual *TmHydABC* aliquots were defrosted and 2.5 μl immediately placed onto the grid, blotted, and plunged into liquid ethane. 12 grids were prepared, varying blot time from 2 to 4 s with 0.75–1.5 mg/ml protein; the blot force parameter was constant at –5. Following screening to optimize protein concentration and blotting parameters, cryo-grids could be consistently prepared with densely packed but non-aggregated particles where it was possible to see several different views of *TmHydABC* by eye. Following screening, a grid at 1 mg/ml protein concentration was selected for data collection on a Titan Krios microscope operated at 300 kV with a K2 detector and

energy filter. The energy filter was set to a 20 eV window. Three exposures were collected per hole, and the autofocus routine was run every 10  $\mu\text{m}$ . AutoCTF was used to correct for astigmatism and coma. 4790 movies of 48 frames each were collected. The total fluence was 57 electrons /  $\text{\AA}^2$ .

This grid preparation protocol was tested also for *TmHydABC* containing the  $[\text{Fe}]_{\text{H}}$  subcluster, however, it was clear from the screening that major degradation of the sample occurred. Therefore, anaerobic grid preparation was then attempted when anaerobic gloveboxes became available.

### ***TmHydABC without the $[\text{Fe}]_{\text{H}}$ subcluster bound to ferredoxin***

1.2/1.3 UltrAuFoil grids were glow discharged (PELCO easiGlow) for 90 s on each side using atmospheric gas before mounting in Vitrobot (model IV) tweezers (Thermo Fisher Scientific). Grids were prepared with minimal exposure to air using anaerobically frozen aliquots of ‘apo’ *TmHydABC* (1 mg/ml) and *Thermotoga maritima* ferredoxin (0.564 mM). These were individually defrosted and quickly mixed (1.4  $\mu\text{L}$  of ferredoxin in 100  $\mu\text{L}$  *TmHydABC*) to achieve a ratio of 1.3:1 of ferredoxin per mole of trimer ‘ABC’. 2.5  $\mu\text{l}$  of this mixture was immediately placed onto each grid, blotted, and plunged into liquid ethane. Blot time varied from 2 to 4 s with the blot force parameter constant at  $-5$ . Following screening, a grid was selected for the data collection using the on-site microscope; a Glacios microscope (ThermoFisher) operated at 200 kV mounted with a Falcon 4 direct electron detector. The data collection parameters were: one exposure per hole, total fluence of 50 electrons/ $\text{\AA}^2$ , exposure time of 8.75 s, defocus range  $-2$  to  $-1.4$   $\mu\text{m}$ , magnification of  $\times 150,000$ , spot size 5, calibrated ‘pixel size’ 0.934  $\text{\AA}/\text{pixel}$ , and objective aperture of 100  $\mu\text{m}$ .

A total of 3002 movies were recorded in EER format with 2107 raw frames per movie (155).

### ***TmHydABC with the $[\text{Fe}]_{\text{H}}$ subcluster (holoenzyme) with anaerobic grid preparation***

1.2/1.3 UltrAuFoil grids were glow discharged (PELCO easiGlow) for 90 s on each side using atmospheric gas before mounting in Vitrobot (model IV) tweezers (Thermo Fisher Scientific). The Vitrobot grid preparation device was previously placed inside an anaerobic glovebox together with defrosted aliquots of *TmHydABC* containing the  $[\text{Fe}]_{\text{H}}$  subcluster and *Thermotoga maritima* ferredoxin. These anaerobically prepared aliquots were opened only after being introduced inside the anaerobic glovebox to minimize *TmHydABC* oxygen exposure. In addition, solutions of sodium dithionite and NADH were placed inside the

anaerobic glovebox and left open to deoxygenate. The anaerobic glovebox was operated at room temperature and the oxygen concentration oscillated between 80 to 190 ppm (equivalent to 0.008-0.019%) during grid preparation. Significant static electricity was noted across the experiment with grids being difficult to handle (sticking to surfaces); this phenomenon was likely related to the plastic material of the glove box.

Each grid was prepared by placing 2.5  $\mu$ l of sample aliquots onto them, then blotted, and plunged into liquid ethane. In total 20 grids were prepared, varying blot time from 1.5 to 5 s; the blot force parameter was constant at  $-5$ . The relative humidity inside the Vitrobot varied between 70% and 100%. The grids were prepared using sample solutions containing 1 mg/ml 'holo' *TmHydABC* (equivalent to a concentration of trimer 'ABC' of 6.26  $\mu$ M) and other components as follows:

- 2 grids: 1 mg/ml *TmHydABC* (**collected**)
- 4 grids: 1 mg/ml *TmHydABC* + ferredoxin (100  $\mu$ M) (**collected**)
- 2 grids: 1 mg/ml *TmHydABC* + sodium dithionate (2 mM)
- 2 grids: 1 mg/ml *TmHydABC* + sodium dithionate (2 mM) + ferredoxin (100  $\mu$ M)
- 3 grids: 1 mg/ml *TmHydABC* + sodium dithionate (2 mM) + NADH (2 mM)
- 3 grids: 1 mg/ml *TmHydABC* + sodium dithionate (2 mM) + ferredoxin (50  $\mu$ M) + NADH (1 mM)
- 4 grids: 0.3 mg/ml *TmHydABC* + ferredoxin (30  $\mu$ M)

All samples containing sodium dithionate showed signs of protein degradation and therefore only the samples containing pure 'holo' *TmHydABC* and 'holo' *TmHydABC* with the addition of ferredoxin were selected for data collection. Data were collected using the on-site microscope; a Glacios microscope (ThermoFisher) operated at 200 kV and mounted with a Falcon 4 direct electron. The data collection parameters for the sample containing only 'holo' *TmHydABC* were: one exposure per hole, total fluence of 50 electrons/ $\text{\AA}^2$ , exposure time of 14.05 s, defocus range -1.8 to -0.6  $\mu$ m, autofocus routine run every 8  $\mu$ m, magnification of x 150,000, spot size 6, calibrated 'pixel size' 0.574  $\text{\AA}$  /pixel, and objective aperture of 100  $\mu$ m. Data were collected using AFIS mode. The data collection parameters for the sample containing only 'holo' *TmHydABC* + ferredoxin were: one exposure per hole, total fluence of 50 electrons/ $\text{\AA}^2$ , exposure time of 11.25 s, defocus range -2 to -2.2  $\mu$ m, autofocus routine run every 10  $\mu$ m, magnification of x 150,000, spot size 6, non-calibrated 'pixel size' 0.574  $\text{\AA}$  /pixel, and objective aperture of 100  $\mu$ m. Data were collected using normal mode.

A total of 1342 movies were recorded in EER format with 3381 raw frames for the holoenzyme and a total of 1200 movies with 2709 raw frames for the holoenzyme + ferredoxin sample.

## 2.1.4 Image processing

### *TmHydABC without the [2Fe]<sub>H</sub> subcluster (apoenzyme)*

The Relion pipeline was used for all image processing. Whole micrograph motion correction and damage weighting were performed using the implementation of MotionCor2 in Relion (156). Initial CTF values were determined with CTFFIND4 (136) and particles were picked using a low-resolution ( $\approx 10$  Å) preliminary dataset that was previously collected. The early stages of 2D and 3D classification used images with the original pixel size downsampled from 0.85 to 3.4 Å/pixel. Reference-free 2D classification was performed to classify the particles and remove broken particles that are most likely denatured at the air-water interface, common to most cryo-EM projects (111). It was clear there were large particles that had four lobes consistent with a tetramer of trimers and smaller particles, with high-resolution features. Any classes that showed high-resolution features in the 2D class averages were selected for coarse 3D classification, which effectively cleaned the dataset to only the tetramer of trimer particles, consistent with the gel filtration profile of the preparation. An initial model was generated in Relion and coarse 3D classification ( $7.5^\circ$  sampling) without symmetry being enforced was used to remove broken particles. Docking in the related structure of subunits Nqo1, Nqo2, and Nqo3 of complex I from *T. thermophilus* (5) showed that the particles had D2 symmetry, consistent with a tetramer of trimers Hyd(ABC)<sub>4</sub> arrangement. The particles were reextracted with the original pixel size of 0.85 Å/pixel and the 3D auto-refinement of these particles resulted in a 2.5 Å resolution structure when D2 symmetry was applied. To further improve the resolution, anisotropic magnification, trefoil, and fourth-order aberration parameters were refined; with astigmatism and defocus being fitted on a per-particle basis (157). Bayesian polishing was also performed (158). The map displayed the features expected at such a resolution, with rotamers of many side chains being clear and water molecules being visible in well-resolved regions. Refinement resulted in a 2.3-Å resolution structure when D2 symmetry was applied. The final calibrated pixel size was 0.824 Å.

To investigate the blurred bridging regions, symmetry expansion was used to separate the different conformations into classes. Here, particles with symmetry are transformed so that

each symmetry-related subparticle is overlaid; a mask is then applied so each subparticle can be treated independently for classification and refinement (159). The high-resolution D2 refinement was used as a starting point. As each lobe appeared independent of the others, symmetry expansion with D2 symmetry to match the core was attempted to separate the different conformations into classes but this was unsuccessful, resulting in maps no clearer than the first. However, when the same process was repeated using C2 symmetry much better results were obtained. To achieve this, the `reliion_symmetry_expand` command was used to apply a C2 symmetry operator to the particles in the `refined.star` file. A 20 Å low-pass filtered mask, generated from fitted atomic coordinates and expanded by 20 pixels with 6 pixels soft edge, was then applied to half of the complex containing two tightly connected *TmHydABC* protomers with a complete and connected electron transfer network. A clear bridging density was found to exist between two HydBC lobes in a subset of Hyd(ABC)<sub>2</sub> particles (total 39.1%). A tighter mask was then created that included exclusively the two ‘bridges’ densities in the Hyd(ABC)<sub>2</sub> unit (20 Å low-pass filter, 6 pixels soft edge), allowing a better 3D classification without losing any signal in the ‘bridges’. The resulting ‘bridged’ classes (bridge backwards and forward) were refined with C1 symmetry applying a 6-pixel soft edge mask that included the Hyd(ABC)<sub>2</sub> unit with two bridges, reaching a resolution of 2.8 Å for both classes. In this subset, half of the particles had the bridge forward with respect to the rest of the enzyme (i.e., bridging from A to B’) and the other had the bridge backwards (i.e., bridging from A’ to B), but none showed both the bridges with clear density. The bridge is formed by the C-terminus of HydA (containing one [2Fe–2S] cluster) from one protomer and the C-terminus of HydB (containing two [4Fe–4S] clusters) from the neighbouring protomer, thereby breaking the rotational symmetry between the two bridged lobes.

To explore the location of the HydB in the non-bridged class, a mask was created around the suspected area and used for classification and refinement. The improved map allowed an improved mask to be created for a final round of classification and refinement. The resulting map density is of insufficient quality for ab initio model building, but the strong FeS signals allowed the HydB CT-domain to be docked in place.

### ***TmHydABC without the [2Fe]<sub>H</sub> subcluster bound to ferredoxin***

The Relion pipeline was used for all image processing. Whole micrograph motion correction and damage weighting were performed using the implementation of MotionCor2 in Relion (156). Initial CTF values were determined with CTFFIND4 (136), and particles were picked

using the D2 refined map previously obtained (described above) as a reference. This map was rescaled to match the pixel size of this dataset (0.934 Å/pixel). The early stages of 2D and 3D classification used images downsampled from 0.934 to 2.45 Å/pixel. Reference-free 2D classification was performed to classify the particles and remove broken particles that are most likely denatured at the air-water interface. Significantly fewer intact particles were found relative to the previous dataset possibly due to worse prep, however, sufficient for continuing with the data processing pipeline. After 3D classification, a class showing the distinct four lobes of *TmHydABC*, was selected for further refinement using C1 symmetry that showed blurry regions approximately in correspondence of the ‘bridges’, however, in a different arrangement than previously obtained with the sample not containing ferredoxin. A mask in correspondence with this blurry region was therefore created from the fitted atomic coordinates of the bridge domain (previously obtained as described in the above section) and used for further 3D classification. The masked 3D classification was performed on a symmetry-expanded dataset; the `relion_symmetry_expand` command was used to apply a C2 symmetry operator to the particles in the refined. star file, effectively doubling the available particles. This 3D classification allowed the determination of a class of particles with more distinct “bridge” features and a potential density attributable to ferredoxin. This class was refined resulting in a 4.90 Å resolution map. This refined map was then rescaled to the original pixel size and refined again resulting in a 4.90 Å resolution map. This latest refined map shows important differences in the bridge confirmations compared to the previously obtained one (above section), however, the ‘bridge’ and potential ‘ferredoxin’ densities were too blurry to allow atomic model building. An approximate fitting based on cofactor signals was attempted instead.

### ***TmHydABC with the [2Fe]<sub>H</sub> subcluster (anaerobic preparation)***

The Relion pipeline was used for all image processing. Whole micrograph motion correction and damage weighting were performed using the implementation of MotionCor2 in Relion (156). Initial CTF values were determined with CTFFIND4 (136), and particles were picked using the D2 refined map previously obtained as a reference. This map was rescaled to match the pixel size of this dataset (0.574 Å/pixel). The early stages of 2D and 3D classification used images downsampled from 0.574 to 1.51 Å/pixel. Reference-free 2D classification was performed to classify the particles and remove broken particles that are most likely denatured at the air-water interface. No intact particles were found after rounds of 2D classification, however, some classes showed some high-resolution features and were, therefore, 3D classified

and refined. It was clear from the refinement that no significant interesting information could be obtained from this dataset which seemed to be lacking intact particles possibly due to imperfect anaerobic handling or phenomena related to the high static observed during grid preparation.

### ***TmHydABC with the [2Fe]<sub>H</sub> subcluster bound to ferredoxin (anaerobic preparation)***

The Relion pipeline was used for all image processing. Whole micrograph motion correction and damage weighting were performed using the implementation of MotionCor2 in Relion (156). Initial CTF values were determined with CTFFIND4 (136), and particles were picked using the D2 refined map previously obtained as a reference. This map was rescaled to match the pixel size of this dataset (0.574 Å/pixel). The early stages of 2D and 3D classification used images downsampled from 0.574 to 1.51 Å/pixel. Reference-free 2D classification was performed to classify the particles and remove broken particles that are most likely denatured at the air-water interface. After 6 rounds of 2D classification, only 2'000 particles with distinct features of *TmHydABC* were found. Due to the very low number of particles, no significant interesting information could be obtained from this dataset which similarly to the previously described one showed evidence of significant protein degradation.

## **2.1.5 Model building and validation**

### ***TmHydABC without the [Fe]<sub>H</sub> subcluster (apoenzyme)***

WinCoot (146) and Phenix (147) were used for model building and validation, and ChimeraX (160) was used for visualization and figure generation. A homology model generated based on bacterial complex I (5) was used as a starting point for model building. Here, the Nqo3 subunit of complex I is related to HydA, Nqo1 to HydB, and Nqo2 to HydC. The map density was sufficiently strong to allow ab initio building of the non-conserved regions of HydA and HydB in the well-resolved parts of the D2 map, however, without further classification, many parts of HydB and HydC were poorly resolved. Model refinement was performed using Phenix real-space refinement. Phenix automatically recognizes the ligation between FeS clusters and cysteines, without having to manually define these restraints or provide the correct definition of the FeS geometry (161).

The ‘bridge’ is formed from 91 residues of the CT of HydA and 61 residues of the CT of HydB. The HydA CT ‘bridge’ domain has homology with the CT of HndA from the NADP-reducing hydrogenase complex in *Desulfovibrio fructosovorans* (162) and 82 CT residues of *T. maritima* HydC. The HydB CT ‘bridge’ domain has homology with bacterial 2×[4Fe–4S] ferredoxin domains. In both cases, Phyre2 (144) was used to build a homology model from this information, which was further built into the density, combined with the model for the rest of the complex built from the D2 map, and refined.

### ***TmHydABC without the [Fe]<sub>H</sub> subcluster bound to ferredoxin***

Here, the model building of the bridge domain and ferredoxin was not possible due to the regions having poor resolution. However, the intense signals from the high electron scattering iron-sulfur clusters allowed fitting with the previously obtained atomic model of the ‘bridges’ and the atomic model of *T. maritima* ferredoxin (PDBID: 1ROF).

## **2.2 Molecular biology: strep-tagging hydrogenase-1**

### **2.2.1 Bacteria strains, plasmids, and media used**

**Table 1.6** Bacteria strains used. *E. coli* K12 MG1665 genotype: F-  $\lambda$ - ilvG- rfb-50 rph-1. *E. coli* DH5 $\alpha$  genotype: F- endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG  $\Phi$ 80dlacZ $\Delta$ M15  $\Delta$ (lacZYA-argF)U169, hsdR17(rK- mK+),  $\lambda$ -. *E. coli* DB3.1 genotype: F- gyrA462 endA1 glnV44  $\Delta$ (sr1-recA) mcrB mrr hsdS20(rB-, mB-) ara14 galK2 lacY1 proA2 rpsL20(Smr) xyl5  $\Delta$ leu mtl1. *E. coli* CF-001 genotype: HyaB::Strep (N-terminal) F-  $\lambda$ - ilvG- rfb-50 rph-1.

<b>Strain</b>	<b>Use</b>	<b>Source</b>
<i>E. coli</i> K12 MG1665	Strain to be genetically engineered to contain the strep-tagged Hyd-1	DSMZ catalog #18039
<i>E. coli</i> DH5 $\alpha$	Storing and expression of pEcCas plasmid	addGene Plasmid #73227
<i>E. coli</i> DB3.1	Storing and expression of pEcgRNA. ccdB insensitive.	addGene Plasmid #166581
<i>E. coli</i> DH5 $\alpha$	Competent cells, transformed with customized pEcgRNA	Thermo Fisher (Catalog Number EC0112)

<i>E. coli</i> CF-001	<i>E. coli</i> K12 MG1665 genetically engineered to contain the strep-tagged Hyd-1 (after curing)	Obtained here
-----------------------	---	---------------

**Table 1.7** Plasmid used

Plasmid	Use	Antibiotic resistance	Copy number	Growth temperature*	Source
pEcCas	Expression, CRISPR, Red recombinase	Kanamycin	Low	37 °C	addGene Plasmid #73227
pEcgRNA	Expression gRNA and ccdB toxin	Spectinomycin	High	37 °C	addGene Plasmid #166581
pUC19	Transformation efficiency comparison	Ampicillin	High	37	Thermo Fisher

\*note that sometime a temperature of 30 °C was selected instead but was not necessary. All steps can be performed at 37 °C.

**Table 1.8** Growth media used

Media	Composition	Supplier
Buffered Lysogeny broth (LB)	Caseine digest peptone (10 g/L) Sodium Chloride (10 g/L) Yeast extract (5 g/L) pH adjustment (pH 7.2. at 20 °C): TRIS/TRIS-HCL (1.5 g/L)	Melford ltd. LB buffered capsules
S.O.C	Caseine digest peptone (20 g/L) Yeast extract (5 g/L) 10 mM NaCl 2.5 mM KCl 10 mM MgCl <sub>2</sub> 10 mM MgSO <sub>4</sub> 20 mM glucose	Thermo Fisher

**Notes:** all media when mentioned are implied to be sterile. All bacteria manipulations were performed inside a laminar flow hood except centrifugations and transformation steps. Loops, pipettes, falcon tubes, and flasks were all sterile according to standard laboratory practices.

## 2.2.2 Glycerol stocks preparation from purchased plasmid-harboursing strains and K12 MG1665

### *Strains harbouring pEcCas and pEcgRNA*

pEcCas and pEcgRNA plasmids were purchased from addGene (<https://www.addgene.org/>) and arrived in bacterial stabs (*E. coli* DH5 $\alpha$  contains pEcCas plasmid and *E. coli* DB3.1

contains the pEcgRNA plasmid). An LB plate containing 50 ug/ml of kanamycin was streaked with bacteria colonies from the pEcCas stab. Another LB plate containing 50 ug/ml spectinomycin was streaked with bacteria colonies from the pEcgRNA stab. The two LB plates were left overnight at 37 °C.

The next day two 50 mL plastic falcon tubes were filled with LB medium (~5 mL). In one tube a stock solution of kanamycin was added reaching 50 ug/mL of antibiotic concentration, while in the other spectinomycin was added instead (50 ug/mL). The two tubes were inoculated with a single colony from the LB plates with the respective antibiotic resistance strain. The two tubes were left overnight in two shaking incubators at 37 °C.

The next day 20% glycerol stocks were prepared from the liquid cultures and stored at -80 °C. To prepare the glycerol stocks, 400 µL of 50% glycerol solution (sterilized using syringe filters) and 600 µL of LB liquid culture were added together in each screw-top tube. The screw-top tubes are gently mixed and placed at -80 °C for long-term storage.

### ***E. coli K12 MG1665***

*E. coli* K12 MG1665 lyophilized strain was purchased from DSMZ. The strain was revived according to the protocol which involved opening a glass ampule and rehydration with 0.5 ml of LB medium for 30 min. After the cells had been rehydrated, they were streaked into an LB plate and left overnight at 37 °C. The next day a single colony of K12 MG1665 was selected from the LB plate and inoculated in 5 mL LB. The liquid culture was left overnight at 37 °C. The next morning 20% glycerol stocks were made as described above.

### **2.2.3 Plasmids extraction**

The plasmids pEcCas, pEcgRNA were maintained and stored inside cells. Therefore plasmid extraction was used to retrieve purified plasmids for further manipulation and/or transformation steps.

A sterile loop or pipette tip was used to inoculate a 50 mL tube containing 20 mL of LB and 50 ug/ml kanamycin with a small amount of frozen material taken from the glycerol stock of *E. coli* DH5α harbouring the pEcCas plasmid. Similarly, a tube containing 10 mL of LB and 50 µg/ml spectinomycin was inoculated with a small amount of frozen material taken from the glycerol stock of *E. coli* DB3.1 harbouring the pEcgRNA plasmid. In this latter case, only 10

mL of medium are used as pEcgRNA is a high copy number plasmid, and therefore fewer cells are needed to obtain a sufficient amount of purified plasmid. These two tubes were left in an incubator overnight with shaking at 37 °C.

The next morning the GeneJET Plasmid Midiprep Kit #K0502 from Thermofisher was used to extract the plasmids. The extraction protocol involves harvesting the cells by centrifugation, resuspension with a solution containing RNase enzyme, SDS/alkaline lysis, and neutralization of the lysate to create appropriate conditions for the binding of plasmid DNA on the silica membrane of the GeneJET spin column. The supernatant after neutralization is loaded on the GeneJET spin column and after centrifugation washing steps are repeated. Finally, an elution buffer (10 mM Tris-HCl, pH 8.5) is added to the centre of the GeneJET spin column to elute the plasmid DNA. After centrifugation, the plasmid in the elution solution can be collected and directly stored at -20°C for future use.

Generally, around 30 µL of elution solution with a DNA concentration of 50-100 ng/µL as measured by a NanoDrop device was obtained for both plasmids. Therefore a yield of ~1.5-3 µg of purified plasmid was obtained for each prep.

#### **2.2.4 Making electrocompetent *E. coli* K12 MG1665 cells**

Making electrocompetent cells means preparing the cells to receive external DNA through electroporation. Generally, protocols to make electrocompetent cells involve several washing steps with ice-cold deionized water by repeated pelleting and resuspension to remove salts and other components that may interfere with electroporation. After 3 to 4 washes, the cells are finally pelleted and resuspended in 10% glycerol for storage. The protocol below presented was the one used here, however, variations of it will produce similar results as long as they include several washing steps.

An LB plate was streaked with a small amount of material taken from a frozen glycerol stock of *E. coli* K12 MG1665. The plate was left overnight at 37 °C. The next day, one colony was picked and used to inoculate a 50 mL tube containing 20 mL of LB. This liquid culture was incubated overnight at 37 °C with shaking.

The next day, the 20 mL of liquid culture appeared cloudy confirming bacterial growth. 1 mL of this culture was added to a flask containing 500 mL of LB and incubated at 37 °C with

shaking. After ~3h the culture reached an OD<sub>600</sub> of 0.6 (mid-exponential phase). The flask with cells was then put on ice for 15 min. The content was then transferred to a 1 L centrifuge tube and topped up with another ~500 mL of ice-cold sterile Milli-Q water. The centrifuge tube was centrifuged for 20 min at 5,000 x g and 2-4 °C. After centrifugation, the supernatant was discarded. The pellet at the bottom of the centrifuge tube was left undisturbed and another ~ 1 L of ice-cold Milli-Q water was added to the centrifuge tube. The tube was centrifuged again for 20 minutes, 5,000 x g, and 2-4 °C and the supernatant was discarded. Finally, a sterile ~ 1L of ice-cold 10% glycerol solution (prepared with Milli-Q water) was added to the centrifuge tube without resuspending the pellet, and centrifugation as described above was repeated.

The supernatant was discarded and a minimal amount of 10% sterile glycerol solution was added to resuspend the pellet. The OD<sub>600</sub> of a 1:100 dilution of the resuspension was measured to estimate the cell density in the original resuspension (*E. coli* cultures: OD<sub>600</sub> x 10<sup>8</sup> cells/mL) and if required additional 10% sterile glycerol solution was added to the resuspended pellet until a cell density of ~ 2·10<sup>11</sup> is reached. The resuspended pellet was then quickly transferred to screw-top tubes and stored at -80°C in 100 µL aliquots.

### **2.2.5 Transformation of *E. coli* K12 MG1665 with pEcCas using electroporation**

One aliquot of electrocompetent *E. coli* K12 MG1665 cells (prepared as described in chapter 2.2.4) was put in ice to defrost together with purified pEcCas plasmid (prepared as described in chapter 2.2.3). In addition, an electroporation cuvette (Cell Projects Ltd. EP-102, blue cap) was left in ice to chill.

As soon as the electrocompetent cells were defrosted, 70 µL of their suspension was pipetted in the electroporation cuvette together with 1.5 µL of the solution containing purified pEcCas plasmid.

Then, the electroporation cuvette was placed in an electroporation device (Bio-rad MicroPulser) and the parameters suitable for electroporation of *E. coli* cells were selected (2.5 kV, 5ms). Immediately after a “bip” sound indicating the end of the electroporation, 1 mL of LB medium was added to the electroporation cuvette. Then the entire content of the electroporation cuvette was transferred in a microcentrifuge tube and incubated at 30 °C for 30 min.

After 30 min, 50  $\mu$ L of cell suspension was streaked in an LB plate with 50  $\mu$ g/ml kanamycin and left overnight at 30°C. Only cells that survive kanamycin selection have the pEcCas plasmid. In addition, also the *E. coli* K12 MG1665 untransformed strain was tested in another LB plate containing kanamycin to confirm that the untransformed strain is not resistant to kanamycin (this can happen if the purchased strain is not what is advertised).

### **2.2.6 Making electrocompetent *E. coli* K12 MG1665 cells harbouring pEcCas plasmid**

One colony from the kanamycin selection plate (Chapter 2.2.5) was inoculated in a falcon tube containing 20 mL sterile LB with 50  $\mu$ g/ml kanamycin and incubated. The falcon tube was incubated overnight with shaking (180 rpm) at 30°C.

The next day an OD<sub>600</sub> of 0.8 was reached. Part of the liquid culture was used to make 18% glycerol stocks while the remaining was used to inoculate two large flasks containing 120 mL LB and kanamycin (50  $\mu$ g/mL). The two flasks were incubated at 30 °C with shaking and after ~ 3 h an OD<sub>600</sub> of 0.35 was reached (needs to be between 0.3 and 0.5). Sterile (syringe filtered) 2.5 mL of L-arabinose stock solution (235 mg/mL) was added to each flask to induce the lambda-red genes. After the addition of arabinose, the cells were incubated for 1 h.

After 1 h, the flasks' content was aliquoted in six 50 mL falcon tubes (40 mL each). These were centrifugated and washed twice with milli-Q ice-cold sterilized water, and then with ice-cold 10% glycerol. In all washing steps the centrifuge was run at 4500 x g for 10 min at 4°C. A minimal amount of 10% glycerol solution was added to resuspend the pellets. The resuspension was then aliquoted in 8 screw-top tubes each containing 50  $\mu$ L of it. These 50  $\mu$ L aliquots containing electrocompetent cells were stored at -80 °C.

### **2.2.7 Restriction digestion (optional)**

To verify that the *E. coli* K12 MG1665 cells were correctly transformed with the pEcCas plasmid, restriction digestion can be performed. Kanamycin selection as described in chapter 2.2.5 should ensure that only cells that have received the pEcCas plasmid containing a kanamycin resistance gene survive, however, it does not exclude the presence of other contaminants plasmids that confer resistance (it should be a rare eventuality). To double-check

that the plasmid incorporated in the cells is the desired one, restriction digestion can be used. Here restriction Digestion with BamHI-HF restriction enzyme (NewEngland BioLabs Ltd.) was performed.

A small number of transformed cells from the colony used to make electrocompetent cells (Chapter 2.2.6) was inoculated in 25 mL of LB in a falcon tube and incubated overnight at 30 °C with shaking. The next day GeneJET Plasmid Midiprep Kit #K0502 was used to extract the purified plasmid. The purified plasmid had a concentration of ~180 ng/μL.

The ‘original’ pEcCas purified plasmids from *E. coli* DH5α (Chapter 2.2.3) and the purified plasmids from transformed *E. coli* K12 MG1665 cells were both digested with BamHI-HF in two separate reactions. This is a high-fidelity (HF) enzyme engineered to have reduced star activity (i.e., the enzyme does not cut a sequence that is an imperfect match to its known target). The restriction digestion protocol used was the one suggested by the manufacturer for a 50 μL reaction changed by using half volumes of each component to perform a 25 μL reaction instead. Therefore, 2.5 μL of 10X rCutSmart Buffer, ~1 μg DNA (μL of plasmid-containing solution calculated from the measured DNA concentration), 2 μL BamHI-HF, and sterilized milli-Q water to reach a volume of 25 μL. This mixture was incubated for 15 min at 37 °C.

The two reaction batches after 15 min were run on an agarose gel to confirm the presence of identical bands and fragments compatible with the predicted digestion products.

### **2.2.8 Customization of the gRNA sequence in pEcgRNA plasmid**

The pEcgRNA plasmid needs to be modified so that it contains the sequence that can be transcribed to the desired gRNA. To perform the modification, oligonucleotides with the desired N20 sequence (designed using the CRISPR Design Tools [https://www.milliporesigmabioinfo.com/bioinfo\\_tools/faces/secured/crispr/crispr.xhtml#](https://www.milliporesigmabioinfo.com/bioinfo_tools/faces/secured/crispr/crispr.xhtml#)) to be inserted in the plasmid needs to be designed and purchased. These oligonucleotides are then annealed forming a double-stranded DNA (dsDNA) with sticky ends. Finally, the pEcgRNA plasmid is linearized and ligated with the dsDNA with the sticky ends. The resulting modified plasmid will be referred to as pEcgRNA(NTHyaB).

#### ***Oligonucleotides preparation***

Two complementary oligonucleotides with sticky ends (blue) were ordered from Sigma Aldrich after checking that no strong secondary structure was present ([https://www.bioinformatics.org/sms2/pcr\\_primer\\_stats.html](https://www.bioinformatics.org/sms2/pcr_primer_stats.html)). Their sequence is the following:

Oligonucleotide A: TAGTGAGCACTCAGTACGAAACTC

Oligonucleotide A': AAACGAGTTTCGACTGAGTGCTC

The oligonucleotides arrived dried in tubes containing around ~335 and 352 ng of material, respectively. The two tubes were centrifuged to retrieve eventual oligonucleotide powder that could have adhered to the tube surface and then milli-Q water was added to the tube to produce 100  $\mu$ M stock solutions.

### ***Oligonucleotides annealing***

The two complementary oligonucleotides (A and A') were annealed in a PCR machine to form a dsDNA with sticky ends. In a PCR tube were added 40  $\mu$ L of milli-Q water, 5  $\mu$ L of T4 ligase buffer, 2.5  $\mu$ L of oligonucleotide A, and 2.5  $\mu$ L of oligonucleotide A' reaching a total volume of 50  $\mu$ L. This reaction mix was placed in the PCR machine and incubated at 95°C for 5 min, followed by steps where the temperature decreased by 10°C every minute until a temperature of 15 °C was reached. In the final step, the temperature was held at 16°C for 10 min (163). The PCR tube was then placed in ice.

### ***Linearization of pEcgRNA plasmid***

Eco311 (BsaI) restriction enzyme (Thermo scientific 10 U/ $\mu$ L, 1000U, catalogue number: ER0291) was used to digest the pEcgRNA plasmid. In a microcentrifuge tube the following were added in succession: 6  $\mu$ L of sterilized milli-Q water (nuclease-free), 2  $\mu$ L of 10Xbuffer G, 11  $\mu$ L from a 96 ng/ $\mu$ L solution of purified pEcgRNA (therefore 1  $\mu$ g of plasmid), and 2  $\mu$ L of BsaI solution reaching a total of 21  $\mu$ L. The reaction was incubated for 2h at 37°C.

The Monarch® PCR & DNA Cleanup Kit was then used to eliminate all the enzymes from the reaction; The protocol provided by the supplier was used, however, 20  $\mu$ L of sterile milli-Q water was used instead of the provided elution buffer. This choice was made to avoid the elution buffer interfering with the following ligation step. The microcentrifuge tube containing the purified DNA was put on ice and used in the following ligation step.

### ***Ligation of linearized pEcgRNA plasmid with annealed oligonucleotides***

Ligation of linearized pEcgRNA plasmid with the annealed oligonucleotides was performed using the T4 ligase (Merck 100IU T4 ligase, 1 U/  $\mu\text{L}$ ).

5  $\mu\text{L}$  of the pEcgRNA linearized plasmid solution ( $\sim 0.25 \mu\text{g}$ ) and 2  $\mu\text{L}$  from the PCR tube containing the annealed oligonucleotides were added together in a new PCR tube. Then 3  $\mu\text{L}$  of the provided 10x ligation buffer followed by 3  $\mu\text{L}$  of T4 ligase solution were added as well. The reaction volume in the PCR tube was brought to 30  $\mu\text{L}$  by adding 17  $\mu\text{L}$  of sterilized milli-Q water. This reaction mixture was incubated overnight at 10°C (the recommended range was 4 to 16 °C) in the PCR machine. The next day the ligation reaction was considered completed and stored at -20 °C for later use in the next step. This tube contains pEcgRNA(NTHyaB).

### **2.2.9 Transformation of competent *E. coli* DH5 $\alpha$ cells with customized pEcgRNA plasmid**

The ligated customized pEcgRNA plasmid, pEcgRNA(NTHyaB), obtained as described above was defrosted and used to transform DH5 $\alpha$  competent cells (Thermo Fisher Catalog Number EC0112) using heat shock. The transformation protocol provided by the supplier was used with some adjustments. Transformation with pUC19 and ‘original’ pEcgRNA plasmid was performed as well as control experiments.

#### ***Heat-shock transformation protocol used:***

A 100  $\mu\text{L}$  stock of competent DH5 $\alpha$  cells was thawed on ice. Then the cells were gently mixed and three aliquots of 50, 25, and 25  $\mu\text{L}$  of competent cells were placed in chilled microcentrifuge tubes. The larger aliquot of 50  $\mu\text{L}$  was reserved for the transformation with pEcgRNA(NTHyaB), while the others to transformations with pUC19 and ‘original’ pEcgRNA plasmid.

1.5  $\mu\text{L}$  of pEcgRNA(NTHyaB) solution (DNA concentration measured = 95 ng/ $\mu\text{L}$ ) (Chapter 2.2.8), 1.5  $\mu\text{L}$  of ‘original’ pEcgRNA solution (DNA concentration measured = 95 ng/ $\mu\text{L}$ ) (Chapter 2.2.3), and 1  $\mu\text{L}$  of pUC19 plasmid solution (corresponding to 10 pg) were added to their respective microcentrifuge tubes containing the thawed competent cells. The competent cells were incubated in ice for 30 minutes and then heat-shocked by placing them in a water bath at 42 °C for exactly 30 seconds. The cells were then incubated in ice for 2 min and  $\sim 250$

μL of room-temperature S.O.C. medium (which is a rich medium) was added to revive them (250 μL for the pEcgRNA(NTHyaB) reaction and 275 μL for the pEcgRNA and pUC19 reactions). The three tubes with S.O.C medium were incubated for 1 hour at 37 °C with shaking (225 rpm).

After 1 h, plates were prepared by spreading various amounts of revived cell suspension from the three transformation reactions. Each plate contained the appropriate selection antibiotic: 50 μg/mL spectinomycin for the cells transformed with pEcgRNA(NTHyaB) and ‘original’ pEcgRNA and 100 μg/mL of ampicillin for the pUC19 transformation. The colonies were counted to determine transformation efficacy (e.g. 20 μL from the pEcgRNA(NTHyaB) reaction were plated and 21 colonies were counted on the selection plate; the transformation efficiency is calculated as:  $\frac{\#colonies}{total\ DNA\ added} \cdot \frac{total\ reaction\ volume}{volume\ plated} = \frac{21}{0.142\mu g} \cdot \frac{(250\mu L+50\mu L)}{20\mu L} = 2.2 \cdot 10^3 \#colonies/\mu g$ ).

### 2.2.10 Sequencing of customized pEcgRNA plasmid

The colonies from the selection plate containing the cells transformed with the pEcgRNA(NTHyaB) plasmid are expected to have acquired spectinomycin resistance. To make sure that the colony picked has the correctly modified plasmid, the portion of the plasmid containing the gRNA sequence sent for was sequenced (Eurofins Genomics Ltd.).

Six colonies were picked from the selection plate containing the cells transformed with the pEcgRNA(NTHyaB) plasmid and inoculated in six falcon tubes containing 20 mL LB and spectinomycin (50ug/mL). The cultures were incubated overnight at 37°C with shaking. From each tube, a 20% glycerol solution was made and stored at -80°C. The remaining culture volume was used to extract the plasmids. A total of 6 solutions containing purified plasmid were prepared using the GeneJET Plasmid Midiprep Kit (protocol similar to the one described in chapter 2.2.3). 15 μL of each purified plasmid (with eventual dilution) was sent to Eurofins for sequencing making sure that the DNA concentration fell within the recommended range of 50-100 ng/mL. The left-over six solutions were kept at -20°C.

The sequencing primer selected needs to be 200 bp downstream to the target gRNA sequence as the sequencing of the first and last bp is not very accurate.

Sequencing primer ordered: CTGATTCTGTGGATAACCGTATTAC

After the arrival of the sequencing results, a chosen glycerol stock containing the correctly modified plasmid was kept while the others were discarded. Similarly, only one left-over solution containing the purified plasmid was kept.

### **2.2.11 Transformation of electrocompetent *E. coli* K12 MG1665 cells harbouring pEcCas with pEcgRNA(NTHyaB) and template dsDNA**

Electrocompetent *E. coli* K12 MG1665 cells harbouring pEcCas (Chapter 2.2.6) were electroporated with the EcgRNA(NTHyaB) plasmid and template dsDNA and then selected by streaking on an LB plate containing both spectinomycin (50 µg/mL) and kanamycin (50 µg/mL). The pEcgRNA(NTHyaB) plasmid used was taken from the selected left-over solution kept at -20 °C (Chapter 2.2.10) or if more was needed it could be purified from the selected glycerol stock produced as described in Chapter 2.2.10. 500 ng of the template dsDNA sequence (644bp) was purchased from Integrated DNA Technologies and the IDT gBlock protocol was used to retrieve it (Gibson assembly can be used as an alternative). Resuspension in milli-Q water resulted in a solution with 42 ng/µl DNA concentration.

An additional electroporation control reaction was performed by electroporating only with the EcgRNA(NTHyaB) plasmid. After electroporation, the cells were streaked in LB plates with both spectinomycin (50 µg/mL) and kanamycin (50 µg/mL).

An additional control reaction was performed by streaking on an LB plate with both antibiotics a tiny amount from the stock of *E. coli* K12 MG1665 previously transformed with pEcCas (Chapter 2.2.6).

#### ***Electroporation protocol:***

Two aliquots of electrocompetent *E. coli* K12 MG1665 cells harbouring pEcCas (Chapter 2.2.6) were defrosted in ice together with the desired plasmids. As soon as all was defrosted, 50 µL of electrocompetent cells were pipetted in each chilled electroporation cuvette, and then 1.5 µL of plasmid was added (173 ng). In addition, 5 µL of dsDNA solution (corresponding to 210 ng) was added only to one cuvette. The cuvettes were electroporated (bio-rad MicroPulser) using *E. coli* parameters. Immediately after “bip” indicating the end of the electroporation, 1 mL of S.O.C medium was added to both cuvettes. The two electroporated cell batches were transferred into two microcentrifuge tubes at incubated at 30 °C for 30 min. After 30 min, 30

µl from a 1:10 dilution, 20 µL, 50 µl, 100 µL up to 200 µL of the cell suspensions were plated in LB plates with 50 µg/ml kanamycin and 50 µg/ml spectinomycin and incubated at 30 °C.

### **2.2.12 Colony PCR and sequencing of edited *E. coli* K12 MG1665 cells**

From the plate with the edited colonies (electroporation with EcgRNA(NTHyaB) + template dsDNA, Chapter 2.2.11), four colonies were picked and incubated overnight at 30 °C in a falcon tube with LB and 50 µg/ml kanamycin (pEcCas needs to be maintained as it is used to cure pEcgRNA). The next day glycerol stocks were made for each colony.

Then colony PCR was performed for each glycerol stock to amplify the portion of the genome in *E. coli* K12 MG1665 that was edited. The 1 kb PCR product was purified and sent for sequencing.

Primers were designed with the help of online tools (e.g. [https://www.bioinformatics.org/sms2/pcr\\_primer\\_stats.html](https://www.bioinformatics.org/sms2/pcr_primer_stats.html)) to check for properties such as melting temperature, per cent GC content, and secondary structure following standard guidelines for primer design (164).

#### ***PCR primers used:***

Forward: CAACGCCTATCGACAAAGTC 20 nt 50% GC Tm: 63°C  
Backwards: GCATCTAACACATCGATCCAG 21 nt 48% GC Tm: 63°C  
Anneal at: 64 °C  
100 µM stocks

#### ***Colony PCR protocol:***

A tiny amount for each glycerol stock of edited colonies was streaked in an LB plate containing 50 µg/ml kanamycin. In addition, the ‘original’ unedited *E. coli* K12 MG1665 strain was plated in an LB-only plate as a control experiment. The plates were left during the weekend at room temperature instead of standard overnight incubation at 30°C. One colony from each plate was then scraped with a pipet tip and resuspended in 100 µL of milli-Q water. The protocol for 50 µL PCR reaction using Q5® High-Fidelity 2X Master Mix (NewEngland Biolabs Ltd.) was used. 25 µL of the master mix was added to a mix of 20 µL from the 100 µL colony resuspension and 2.5 µl of primer forward and 2.5 µl of primer backward solutions (from diluted 10 µM stocks). The PCR tube containing this reaction mix was immediately placed on a PCR machine with the following program (Table 1.9):

**Table 1.9** PCR setting

<b>STEP</b>	<b>TEMP</b>	<b>TIME</b>
Initial Denaturation	98°C	1min
35 Cycles	98°C	10 seconds
	64°C	30 seconds (annealing temperature)
	72°C	30 seconds (1kb amplicon)
	72°C	2 minutes
Final Extension	72°C	2 minutes
Hold		4°C

All the PCR products were purified using Monarch® PCR & DNA Cleanup Kit. 20 µL of elution buffer was used. The purified DNA amplicon concentration was measured using a nanodrop device and diluted 1:10 to reach a DNA concentration of 5-10 ng/µL. 15 µL of the diluted amplicon and 100 µL of the two 10 µM primers stock were sent to the sequencing service. In this case, the PCR primers can also be used as sequencing primers.

### 2.2.13 Plasmid curing

After the sequencing results arrived (Chapter 2.2.12), a glycerol stock containing a correctly edited colony was selected and a tiny amount of frozen material was inoculated in 10 mL of LB with 50 µg/ml of kanamycin and 10 mM L-rhamnose (prepared from a filtered 1M L-rhamnose stock solution). L-rhamnose induces the transcription of a specific gRNA sequence that is contained in the pEcCas plasmid which targets the pEcgRNA plasmid which is cut by the Cas9 protein (gene in pEcCas plasmid). This culture was incubated overnight at 30 °C with shaking. 1 µL of the overnight culture was then inoculated in 10 mL of LB and incubated at 37°C for about 6h reaching an OD<sub>600</sub> of 0.620. In the absence of selection pressure provided by the antibiotics, some cells will spontaneously lose the low copy number pEcCas plasmid.

LB selection plates with 4% sucrose were then made and streaked with dilutions of this latest culture. The agar plates were incubated overnight at 30 °C. Only the cells that lost the pEcCas plasmid will survive the sucrose selection (the pEcCas plasmid contains the *sacB* gene resulting in the conversion of sucrose to levan, which is toxic to *E. coli*).

Three colonies were picked and grown in LB-only tubes for 7 h at 37 °C. After 7 h, glycerol stocks were made, and 4 µl from each tube was also inoculated in 5 mL LB with only 50 µg/ml of kanamycin, only spectinomycin (50 µg/mL), and no antibiotics (as a control). A glycerol stock that corresponded to cells that did not grow in any medium with antibiotics but which grew in the LB-only tube was kept. This glycerol stock contains the successfully genome-

edited *E. coli* K12 MG1665 cells cured from both plasmids. This edited strain was used to purify the strep-tagged hydrogenase-1 (Chapter 2.4)

## **2.3 Purification of His-tagged hydrogenase-1 from *E. coli***

### **2.3.1 Protein preparation and purification (initial method)**

#### ***Strains***

*E. coli* cells (K12 strain, LAF-003 mutant (165)), featuring a hexa-histidine tag at the 3'-terminus of the gene encoding the Hydrogenase-1 (Hyd-1) small subunit (HyaA) were used. This allows for the use of a His-trap column for simpler purification. The strain used was already available in the group and its characteristics have been previously described (165).

#### ***Growth conditions***

The histidine-tagged Hyd-1 is natively expressed and therefore larger quantities of cells are needed to produce a sufficient amount of purified protein. 24 L of growth medium were typically prepared using four autoclaved bottles of 6 L each. The LAF-003 strain was taken from the -80 °C freezer and a small sample was incubated in sterilized LB-agar plates without antibiotics. The plates were placed in a 37 °C incubator overnight. One single colony was then taken and inoculated in 100 mL of LB and then incubated with shaking at 37 °C for 6 h.

Subsequently, 20 mL of this solution with cells was added to each of the four 6 L bottles which were filled with autoclaved LB medium, sodium fumarate (4 g/L), and glycerol (5 mL/L) leaving only 5 mL headspace to ensure the establishment of an anaerobic environment. Both the filling of the bottles and inoculation of the colonies were performed in sterilized laminar flow. The 6 L bottles were incubated at 37 °C overnight without shaking until an OD<sub>600</sub> >1.5 was reached (18-20 h). The lid was left loosely closed to let excess fermentation gas escape.

#### ***Growth medium preparation details***

Unbuffered lysogeny broth medium (LB) composed of tryptone (10 g/L), yeast extract (5 g/l), and sodium chloride (10 g/l) was prepared in 2.3 L bottles. Tryptone (18 g), yeast (9 g), and sodium chloride (18 g) were added to each bottle before the bottle was filled to 1.8 L with

MilliQ water. The mixture was shaken by hand until the solutes dissolved, before being sterilized by autoclave.

A sodium fumarate 16% w/v solution was prepared by dissolving 96g of fumarate in 600 ml of MilliQ water using stirring and gentle heating. The solution was then vacuum filtered using a 0.2  $\mu\text{m}$  cellulose nitrate membrane. In addition, a 500ml of glycerol 50% v/v solution was prepared by mixing 250 ml of glycerol and 250 ml of MilliQ water and then autoclaved.

150 mL of 16% w/v of vacuum-filtered sodium fumarate and 60 mL of autoclaved 50%v/v glycerol were added to each 6L bottle and then filled with autoclaved LB media.

### ***Cell isolation and centrifugation***

After reaching an OD  $\text{OD}_{600} > 1.5$  the 24 L of cell culture was centrifugated at 20,000 x g for 20 min at 4 °C (Sorvall Lynx 4000 centrifuge, F-12-6x500 LEX rotor). After every 20 minutes, the supernatant was discarded and more culture was added so that the pellet grew larger each time. Typically, a 24 L growth results in 60-70 g of cell pellet which was collected and divided into two equal batches. One batch was stored at -20 °C for future use, while the other was lysed the same day.

### ***Resuspension of pellet and cell (and membrane) lysis***

The ~ 30-35 g cell pellet was resuspended to a total volume of 125 mL lysis buffer (0.15 M NaCl, 0.1 M Tris pH 7.6), and 25 g of sucrose was added to the suspension resulting in a 0.58 M sucrose concentration. The suspension was stirred at 4°C for 30 min. This was pelleted by centrifugation at 20,000 x g for 30 minutes at 4 °C. The pellet was suspended in 175 mL of ice-cold MilliQ water to lyse the cells by osmotic shock. The lysate was stirred in the fridge for at least 30 minutes. The solution was adjusted to 0.3 M NaCl, 3 % (w/v) Triton X-100, 0.1 M Tris pH 7.6. 1.5 mL of protease inhibitor cocktail (containing AEBSF, Leupeptin, and Pepstatin) and benzonase (80  $\mu\text{g/L}$ ) were added and stirred overnight at 4 °C. The following morning the mixture was divided into 200 mL aliquots (in 250 mL flasks) and each aliquot underwent sonication (Soniprep 150) on ice for 5 x 30 s, swapping aliquots in sequence between each sonication. Sonication was monitored by taking 1 mL aliquots after each cycle, performing centrifugation on these aliquots to monitor the reduction in the size of insoluble sediment. The 200 mL aliquots of cell debris were then combined and centrifuged at 35'000 x g for 40 minutes at 4 °C and the supernatant, which contains soluble proteins and the detergent (Triton X-100) dissolved membrane proteins, was collected.

### ***Affinity purification***

Two buffer solutions were prepared: buffer A (100 mM Tris pH 7.6, 300 mM NaCl, 30 mM imidazole, 0.1 % w/v DDM) and buffer B (100 mM Tris pH 7.6, 300 mM NaCl, 500 mM imidazole, 0.1 % w/v DDM). Both buffer solutions were sterilized by vacuum filtration through a cellulose nitrate membrane filter paper (0.2 µm; diameter 47 mm).

An ÄKTA Start 2 system with an attached HisTrap 5 mL Ni<sup>2+</sup> affinity column was equilibrated with buffer A. The supernatant was loaded onto the column with buffer A, at a flow rate of 5 mL/min. The flowthrough was collected, and the column was washed with 15 column volumes (CVs each of 5 mL) of buffer A. The bound enzyme was eluted using a linear gradient from 0 mM to 500 mM imidazole in buffer B at a flow rate of 4 mL/min (9 CV) and a further 3 CV of 100% buffer B were collected at the end. A total of 30 fractions of 2 mL each were collected, and UV 280 nm absorption post-Ni column was recorded throughout. Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyze the eluted fractions and confirm the presence of *E. coli* Hyd-1. The fractions containing Hyd-1 were transferred in dialysis tubing (6 – 8 kDa) and dialyzed overnight at 4 °C (dialysis buffer 0.15 M NaCl, 0.1 M Tris pH 7.6, 0.3% w/v DDM). After dialysis, a total volume of ~ 20-30 mL was collected and concentrated reaching ~0.5-1 mL using a centrifugal filter tube with a membrane pore size of 30 kDa. These concentrated samples appeared brown.

### ***Size exclusion chromatography (SEC)***

The concentrated sample after dialysis was further concentrated reaching ~150 µL of volume using 30 kDa filter tubes. A buffer solution, buffer C, was prepared using 100 mM Tris, 150 mM NaCl, 0.03% w/v DDM, and adjusted to pH 7.6. A Superdex® 200 16/600 column (124 mL) was flushed with milliQ water (70 mL) before being equilibrated in buffer C. An ÄKTA start system was used to load the highly concentrated sample. Protein was eluted from the column with buffer C at a flow rate of 0.5 mL/min. The presence of protein was monitored through UV-Vis absorbance at 280 nm, and 300 µL fractions were collected. SDS-PAGE was used to analyze these fractions.

## **2.3.2 Protein purification without detergent**

In an alternative protein preparation protocol, the purification stages starting from and including the HisTrap affinity chromatography were performed without the use of detergents.

This method is identical to the above described (Chapter 2.3.1) except that no DDM was used in any phase of the preparation. In addition, size exclusion chromatography was performed using a Micro kit for ÄKTA pure. 50 µL of samples prepared with this detergent-free method were loaded in a Superdex 200 Increase 3.2/300 chromatographic column using the Micro kit for ÄKTA pure 25 to separate the oligomeric states. The column volume (2.4 mL) was flushed with milliQ water before being equilibrated in buffer C (100 mM Tris, 150 mM NaCl, adjusted to pH 7.6.). Protein was eluted from the column with buffer C at a flow rate of 0.04 mL/min. The presence of protein and cytochrome subunit was monitored through UV-Vis absorbance at 280 nm and 420 nm (cytochrome Soret band), and 50 µL fractions were collected. The chromatographic trace was calibrated using aldolase (158 kDa) and thyroglobulin (669 kDa) standards.

## **2.4 Purification of strep-tagged hydrogenase-1 from *E. coli***

### **2.4.1 Protein preparation and purification with DDM**

#### ***Strains***

The genome-edited *E. coli* K12 MG1665 strain prepared as described in Chapter 2.2 was used. This strain encoded the sequence for a strep-tagged Hyd-1, with the sequence encoding for the strep tag placed at the N-terminus of HyaB. Similarly to the his-tagged purification (Chapter 2.3), Hyd-1 is natively expressed.

#### ***Growth conditions***

Typically the *E. coli* strain encoding for the strep-tagged Hyd-1 was anaerobically grown each time in 4.6 L of growth medium using two 2.3 L bottles. The edited strain was taken from the -80°C freezer and a small sample was incubated in an LB-agar plate without antibiotics. The plate was placed in a 37°C incubator overnight. One single colony was then taken and inoculated in 10 mL of LB and then incubated with shaking at 37 °C for 6 h. Subsequently, 3 mL of this solution with cells was added to each 2.3 L bottle containing LB, glycerol (5 ml/L), and fumarate (5 g/L). The 2.3 L bottles were filled with medium leaving minimal head space

to establish an anaerobic environment. The lid was loosely closed and the cultures were incubated for around 20h reaching an  $OD_{600} > 1.5$ .

### ***Growth medium preparation details***

Unbuffered lysogeny broth medium (LB) containing tryptone (10 g/L), yeast extract (5 g/l), and sodium chloride (10 g/l) was prepared in 2.3 L bottles. Tryptone (23 g), yeast (11.5 g), sodium chloride (23 g), and 11.5 mL of pure glycerol were added to each bottle before the bottle was filled to half volume with MilliQ water. The mixture was shaken by hand until the solutes dissolved and then sterilized by autoclave.

A sodium fumarate 17.6 % w/v solution was prepared by dissolving 88g of fumarate in 500 mL of MilliQ water only by stirring. The solution was then vacuum filtered using a 0.2  $\mu$ m cellulose nitrate membrane. 65.3 mL of this solution was added to the autoclaved bottles containing LB and glycerol. The bottles were then filled with autoclaved milliQ water leaving minimal head space (~2-3 cm). All these filling steps were performed inside a laminar flow hood.

### ***Notes on growth conditions screening***

Growth conditions with different components were screened (varying quantities of glycerol, sodium fumarate, sodium formate, and glucose) by growing cells in 2.3 L bottles filled leaving minimal head space to establish an anaerobic environment. The  $OD_{600}$ , final pH, and protein yield were determined for each growth. The best results were obtained when LB + 5 ml/L glycerol + 5 g/L Na fumarate was used resulting in a purified protein yield of 0.28 mg/L measured by calculating the area under the 280 nm absorbance in the affinity purification Akta trace.

### ***Cell isolation and centrifugation***

After reaching an  $OD_{600} > 1.5$  the 4.6 L of cell culture was centrifugated at 4392 x g (4,500 rpm) for 20 min at 4°C (Sorvall LYNX 6000 centrifuge, rotor F9-6x1000 LEX) using 1 L centrifuge tubes. Typically, a 4.6 L growth results in a total of 10-15 g of cell pellets. The cell pellets were resuspended using a total of 25 mL filtered Lysis solution (50 mM Tris, 25 mM NaCl, pH 7.5) with the addition of  $MgCl_2$  (5 mM) and protease inhibitors. The resuspended pellets (total ~35-40 mL) were collected in a falcon tube and stored at -80°C.

Note: the lysis solution was made using stocks of 500 mM Tris and 1 M NaCl. Concentrated HCl was added until a pH of 7.5 was reached and the resulting solution was vacuum-filtered using a 0.2 µm cellulose nitrate membrane. Then 25 mL were transferred in a falcon tube and mixed with 125 µL of a 1M MgCl<sub>2</sub> stock (syringe filtered) and one protease inhibitor tablet (CompleteMini EDTA-free Protease Inhibitor Cocktail, Sigma Aldrich). MgCl<sub>2</sub> is added as the DNaseI activity (later used) is dependent on the presence of Mg<sup>2+</sup> ions.

### ***Cell lysis***

The frozen tube containing ~35-40 mL of resuspended cell pellet was fully thawed. DNaseI was added to the tube from a 10 mg/mL stock reaching a final enzyme concentration of 20 µg/mL. Immediately after a spatula was used to add egg lysozyme powder. The tubes were then put in a cold room and left on a tube roller (for mixing) for 30 min. After 30min the tubes were lysed in a high-pressure homogenization cell disruptor (Constant Systems Ltd. CF1) using a pressure of 27 kpsi. The filtered lysis buffer (50 mM Tris, 25 mM NaCl, pH 7.5) was used for washing and retrieving all the lysate. The collected cell lysate was then poured into two small centrifuge tubes (50 mL).

### ***Membrane isolation***

The two 50 mL centrifuge tubes containing the cell lysate were centrifuged at 50,000 x g for 15 min at 4°C (Sorvall Evolution centrifuge, SS-34 rotor) to remove all cell components except soluble ones and membrane fragments. The resulting opalescent supernatant (displaying the Tyndall effect) was then poured into a 100 mL ultra-high-speed centrifuge and centrifuged at 100,000 x g for 2 h at 4°C (Beckman Coulter Optima L-100 XP Ultracentrifuge, Ty45Ti rotor). The supernatant was discarded and the translucent pellet containing the precipitated membrane fragments was resuspended using a few millilitres of lysis buffer. A Dounce homogenizer was then used to fully resuspend (vortexing should be avoided as damages the proteins). The resuspension (~ 4-5 mL) was then transferred into a falcon tube and stored at -80°C. 10 µl of the resuspension were saved and stored at -20°C to be used in a BCA assay to determine protein concentration (necessary for determining the dilution factor). Protein concentration was then determined.

### ***Membrane solubilization***

The falcon tube containing the resuspended membrane fraction was diluted with lysis buffer to a protein concentration of 5 mg/mL. The dilution factor was determined from the BCA assay

results which showed that resuspended membrane fractions had a typical protein concentration of ~ 40 mg/mL.

Then detergent n-Dodecyl- $\beta$ -D-Maltoside (DDM) was added to the diluted membrane resuspension (~ 35-40 mL depending on the sample) to a total concentration of 1% w/v DDM. The DDM was added from a 10% w/v defrosted stock.

The falcon tube was then placed in a cold room and left on a tube roller (for mixing) for 45 min to fully dissolve the membrane protein in DDM.

After 45 min, the solution containing the solubilized membrane proteins was transferred in a 50 mL centrifuge tube and centrifugated at 20,000 x g for 20 min at 4°C (Sorvall LYNX 6000 centrifuge, rotor F20-12x50 LEX). The supernatant was collected and centrifuged again with the same settings to make sure all the debris was gone. The supernatant after the second centrifugation was then filtered using a 0.2  $\mu$ m syringe filter. The filtrate was collected in a falcon tube and loaded on a StrepTrap column. Filtering is important to avoid clogging the affinity column.

### ***Affinity purification***

The filtrate produced as described above was loaded to the strepTrap column (StrepTrap™ HP 1 mL, Sigma Aldrich) using a micropump (Pharmacia P-1 Peristaltic Pump) and flowrate of 0.5 mL/min inside a cold cabinet (4°C). The procedure for loading the sample on the StrepTrap column is described in the column manufacturer's manual and involves a washing step with filtered milliQ water, followed by regeneration using 0.5 M NaOH, another washing step with milliQ water, re-equilibration with detergent-containing binding buffer (0.1% DDM 100 mM Tris-HCl, 150 mM NaCl, pH 8), and finally sample loading. All these steps are performed using a 0.5 mL/min flow rate and filtered solutions (0.2  $\mu$ m filter). During the sample loading step, the strep-tagged protein binds to the column resulting in a visible brown band.

The strepTrap column after sample loading is then removed from the micropump device and loaded onto an ÄKTA pure 25 system which was operated inside the cold cabinet (4°C). The tubing of the ÄKTA pure 25 system was previously washed and equilibrated with binding buffer.

The mounted column was washed with 10 column volumes (CVs each of 1 mL) of detergent-containing binding buffer (0.04% DDM, 100 mM Tris-HCl, 150 mM NaCl, pH 8) at a flow

rate of 0.5 mL/min. The bound enzyme was then eluted with detergent-containing elution buffer (2.5 mM desthiobiotin, 0.04% DDM, 100 mM Tris-HCl, 150 mM NaCl, pH 8) at a flow rate of 0.5 mL/min. 50  $\mu$ L fractions were collected during elution. Absorption at 280 nm and 420nm post-Strep column was recorded throughout to monitor for protein concentration and cytochrome concentration, respectively. The fractions that appeared more brown also corresponded to the recorded peak absorption.

The fractions of interest were collected and flash-frozen in liquid nitrogen and then stored at -80°C. The less concentrated fractions were concentrated by bench-top centrifugation using 50 kDa filter tubes (Amicon Ultra-0.5 Centrifugal Filter).

Note: On one occasion elution and binding buffers were prepared with 0.01% GDN instead of DDM (everything else remained the same).

#### ***Size exclusion chromatography (SEC)***

A buffer solution, buffer C, was prepared using 100 mM Tris, 150 mM NaCl, 0.03% w/v DDM, and adjusted to pH 7.8. A 50  $\mu$ L fraction from affinity purification was defrosted and loaded in a Superdex 200 Increase 3.2/300 chromatographic column using the Micro kit for ÄKTA pure 25 to separate the oligomeric states. Two column volumes (1 CV = 2.4 mL) were flushed with milliQ water before being equilibrated with 1.5 CV of buffer C. Protein was eluted from the column with buffer C at a flow rate of 0.03 mL/min. The presence of protein and cytochrome subunit was monitored through UV-Vis absorbance at 280 nm and 420 nm (cytochrome Soret band), and 50  $\mu$ L fractions were collected.

Note: On one occasion Buffer C was prepared with 0.005% LMNG instead of DDM (everything else remained the same).

#### **2.4.2 LMNG buffer exchange**

A fraction obtained from affinity purification was buffer exchanged with Lauryl Maltose Neopentyl Glycol (LMNG) detergent. First DDM was removed using a detergent removal spin column (Pierce™ Spin columns, 25 - 100  $\mu$ L sample volume) following the protocol provided by the manufacturer. Then LMNG was added from a 5% stock solution.

## 2.5 Analytical methods

### 2.5.1 Cell concentration

Light scattering of a cell culture depends on turbidity and can be therefore used to estimate cell density. Optical density measurements were performed by measuring sample absorbance (in a cuvette) using a 600 nm wavelength (OD<sub>600</sub>).

### 2.5.2 Determination of protein concentration

#### *Spectrophotometrically (A<sub>280</sub>)*

Protein concentration can be estimated by measuring the UV absorbance at 280 nm of aromatic residues. This method was used for measuring the protein concentration of post-affinity purification fractions.

#### *Bradford Assay*

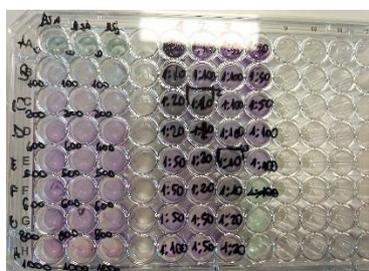
The Bradford Assay was used to determine the protein concentration in the his-tagged protein purification protocol (Chapter 2.3). In this method, the Bradford reagent (acidified Coomassie Brilliant Blue) binds to amino acids in proteins (preferably arginine (166, 167)) undergoing a colour change with the absorbance maximum moving from 470 to 595 nm. Therefore higher protein concentration samples will result in a higher absorbance at 595 nm. However, this assay is sensitive to detergents and was therefore here abandoned in favour of the BCA assay.

Protocol: A standard curve for protein concentration was created using serial dilutions of bovine serum albumin (BSA, 2 mg/mL) from 0 µg to 24 µg, diluted in MilliQ water in cuvettes. Coomassie Brilliant Blue R-250 dye (1 mL, 50 % v/v) was added to each cuvette. After 5 minutes, a UV-Vis spectrum was taken at a wavelength of 595 nm for each concentration of BSA, and the resulting absorbances were plotted. The protein sample was diluted in milliQ water, 1 mL dye was added and UV-Vis absorption at 595 nm was measured. The protein concentration was determined from the standard protein concentration curve.

#### *BCA assay*

A Bicinchoninic acid (BCA) assay (168) was used to determine the concentration of each of the fractions collected after size exclusion chromatography (SEC) and to measure protein concentration in the membrane fractions. The BCA assay is based on the reduction of  $\text{Cu}^{2+}$  ions to  $\text{Cu}^+$  ions in an alkaline solution in the presence of proteins and the subsequent formation of a purple complex between bicinchoninic acid and  $\text{Cu}^+$  ions. At lower temperatures, copper cysteine, cystine, tryptophan, and tyrosine residues are the primary reductive groups, while at elevated temperatures, the peptide bonds become the predominant contributors. Therefore performing the reaction at  $60^\circ\text{C}$  versus  $37^\circ\text{C}$  or room temperature reduces the variation in the response based on protein composition (167). The BCA assay is generally tolerant of ionic and nonionic detergents (169).

Protocol: The BCA assay was performed in a 96-well plate using commercially provided stock solutions (Sigma Aldrich, BCA protein assay kit).  $10\ \mu\text{L}$  of BSA standard at concentrations of 1.0, 0.8, 0.6, 0.4, 0.2, 0.1, and 0.05 mg/mL were deposited on a microwell plate and each standard was repeated in 4 wells. Other wells were filled with  $10\ \mu\text{L}$  of the samples with unknown protein concentration and dilutions of it.  $190\ \mu\text{L}$  of the working reagent was added to each well and incubated for 30 min at  $37^\circ\text{C}$ . The absorbance at 562 nm was measured for each well and the concentration of the fractions was determined using the standard curve calculated automatically with the SpectraMax software.



**Figure 2.1** Image showing an example of a BCA assay conducted to determine the concentration of the membrane fraction. Dilutions of membrane protein suspension on the left and BSA standard on the right.

### 2.5.3 Determination of DNA concentration

DNA concentrations were determined spectrophotometrically at 260 nm (extinction coefficient of  $50\ \text{ng}\cdot\mu\text{L}^{-1}\cdot\text{cm}^{-1}$ ), by placing  $1.5\ \mu\text{L}$  on the sample pedestal of a NanoDrop spectrophotometer after blanking. The presence of guanidine (or phenol) was monitored at 230 nm.

## 2.5.4 Gel electrophoresis

### *SDS-PAGE*

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) is commonly used as a method to separate proteins based on their molecular masses. Sodium dodecyl sulfate (SDS) is a reducing agent that in combination with protein denaturation with heating eliminates the influence of structure and charge. Proteins are therefore separated solely on the basis of differences in their molecular weight.

Protocol: SDS-PAGE was used to analyze the eluted fractions and confirm the presence of *E. coli* Hyd-1. The SDS-PAGE was carried out using an 8.75 % resolving gel. SDS reducing buffer (5 x; 3  $\mu$ L) and milliQ water (2  $\mu$ L) were added to 10  $\mu$ L of each fraction and heated at 96°C on a Techne Genius PCR machine for 4 minutes. The marker (10  $\mu$ L) used was unstained low-range SDS-PAGE standard (Bio-Rad) and was heated as well. The gels were run at 0.2 kV in running buffer (x 8) for 50 minutes and visualized by staining in Coomassie Brilliant Blue dye (R250).

In an alternative protocol, pre-cast acrylamide gels were used (BioRad). 3  $\mu$ L of x4 loading dye (containing Tris-HCL, glycerol, SDS, bromophenol-blue, and  $\beta$ -mercaptoethanol) was added to a few microliters of protein sample and milliQ was added reaching a total volume of 12  $\mu$ L per sample to be loaded on the gel. These 12  $\mu$ L samples were then denatured at 96°C. The polyacrylamide gel was placed on a Biorad tank filled with running buffer (5% SDS) and the 12  $\mu$ L samples were transferred to the gel wells in addition to 10  $\mu$ L of Biorad protein ladders. The gel was run at 150 V for 1 hour and then fixed in a fixing solution (10% acetic acid, 50% EtOH) and heated up in a microwave for 30 seconds followed by 10 min of resting. Then the gel was destained with Milli-Q water (microwave 30s wait 10min, microwave 30s, and wait 10min). After destaining, the gel was immersed in a magic dye solution and left overnight. The next day the gel was immersed in milliQ water for destaining and band visualization.

### *Native PAGE gel*

Native polyacrylamide gel electrophoresis (Native PAGE) is a non-denaturing electrophoretic method that is used to analyze intact proteins to determine their oligomeric state by weight separation. The electrophoretic mobility depends on the charge-to-mass ratio and on the physical shape and size of the protein.

Protocol: A pre-made Native PAGE (4 – 15 % acrylamide gradient gel) was loaded with a NativeMark™ unstained protein standard (10 µL), and 10 µL of hydrogenase sample together with 2 µL native dye were loaded. The 10 µL samples were taken from the fractions corresponding to the peaks in the size exclusion chromatogram. The gel was run at 0.2 kV for 40 minutes in running buffer (x 8, no SDS). The gel was fixed, stained with Coomassie Brilliant Blue R-250 dye, shaken for 20 minutes then left in destained overnight. The gel was photographed in a Syngene G: Box.

### **2.5.5 UV-Vis spectroscopy**

UV/vis spectra (250-700 nm) of His-tagged hydrogenase-1 were collected using a DeNovix (DS-11 FX+) spectrophotometer using standard quartz cuvettes with a path length of 1 cm.

UV/vis spectra (350-700 nm) of Strep-tagged hydrogenase-1 were collected with a spectrophotometer (Agilent Technologies, Carry Series UV-Vis spectrophotometer) using 50 µL quartz cuvettes with a path length of 1 cm. 50 µL of the protein sample was placed on the cuvette and their UV-Vis spectra were measured after blanking. 5-6 spectra were taken for each sample and averaged together. The sample was then reduced by adding 2 µL of sodium dithionite from a 1 M solution into the cuvette. The UV/Vis spectra of the reduced sample were then collected to observe the shift of the Soret band in the reduced heme cofactors.

### **2.5.6 Activity assay**

Hydrogen oxidation kinetics were monitored spectrophotometrically at room temperature inside an anaerobic glove box. The assay is based on the reduction of methylene blue mediated by a hydrogenase enzyme which oxidizes H<sub>2</sub> by transferring electrons to methylene blue. The oxidized methylene blue is blue while the reduced form is transparent, and therefore a progressive decolourization can be observed.

Protocol: The extinction coefficient at 626 nm of fully oxidized methylene blue was determined by measuring the absorbance of methylene blue solutions of different concentrations (0, 2, 5, 8, 10, 14, 22, and 25 µM) in mixed buffer (15 mM MES, CHES, TAPS, HEPES and Na acetate at pH 7.6) using a custom-built spectrophotometer (170). The measured value was 29541.01 M<sup>-1</sup>

$^1 \text{ cm}^{-1}$ . Then, a 2 mL aliquot of  $\text{H}_2$ -saturated solution ( $\text{H}_2$  was bubbled through for 1h) containing 25  $\mu\text{M}$  methylene blue in mixed buffer (pH 7.6) was injected into a cuvette (polystyrene, 3 mL volume, 10 mm optical path length) through a suba-seal lid. To the cuvette was added 10  $\mu\text{L}$  of enzyme solution (final enzyme concentration reached: 4 nM). The cuvette was placed inside the custom-built spectrophotometer. The  $\text{H}_2$ -oxidation activity was monitored via the concomitant reduction of methylene blue. All steps (except for the determination of the extinction coefficient of methylene blue) were performed inside an anaerobic glovebox filled with  $\text{N}_2$ .

## **2.6 Cryo-EM methods (hydrogenase-1)**

### **2.6.1 Grid preparation and imaging**

Multiple attempts were performed with different sample preparation methods, which will be discussed in Chapter 5. Depending on the preparation, 1.2/1.3 UltrAuFoil or 1.2/1.3 Quantifoil grids were used. Grids were glow discharged (PELCO easiGlow) for 90 s on each side using atmospheric gas before mounting in Vitrobot tweezers. Depending on the sample, different concentrations of Hyd-1 were used (1 to 10 mg/mL) and 2.5  $\mu\text{L}$  placed onto the grid, blotted (4°C, 100% or 70% humidity), and plunged into liquid ethane. Various blot times (1.5 to 5 s) and blot forces were used (-5 to +20) to achieve different ice thickness levels.

Data collection was performed on a Glacios microscope operated at 200 kV. One exposure was collected per hole with the total fluence maintained at 50.0 electrons/Å. A nominal magnification of 120K or 240K was used (depending on the dataset). Specific parameters for each dataset will be specified in the result chapter (Chapter 5).

### **2.6.2 Image processing**

The Relion pipeline was used for all image processing. Whole micrograph motion correction and damage weighting were performed using the implementation of MotionCor2 in Relion. Initial CTF values were determined with CTFFIND4 and the Laplacian-of-Gaussian (LoG) filter was used to select an initial set of particles. 2D classification was used in multiple rounds.

Strong preferred orientation was observed, nevertheless, 3D classification and refinement were attempted using a 3D map generated from a public pdb structure (PDBID:4GD3) as a template.

### 2.6.3 Preparation of PEGylated gold grids

In one preparation, PEGylated gold grids (1.2/1.3 UltrAuFoil) were made in order to increase particle density. This is an example of self-assembled monolayers, which can be used to functionalize gold surfaces by the binding of thiol groups to the gold atoms. Here (11-Mercaptoundecyl)hexa(ethylene glycol) (Sigma Aldrich) was used as a self-assembled monolayer. This compound contains a thiol head group and 6 PEG groups and will be here referred to as SH-PEG<sub>6</sub> for simplicity.

In this protocol, 200 mL of ethanol was deoxygenated by bubbling N<sub>2</sub> for 2 h. The bottle containing deoxygenated ethanol is then opened inside an anaerobic glovebox.

Microcentrifuge tubes containing 100 µL of the self-assembly reaction mixture were prepared inside the glovebox. To prepare them, 90 µL of pure deoxygenated ethanol and 10 µL of a stock solution of SH-PEG<sub>6</sub> (anaerobically prepared) were added to each microcentrifuge tube. The concentration of SH-PEG<sub>6</sub> in each tube was 20 µM.

1.2/1.3 UltrAuFoil grids were glow discharged 90 s each side using atmospheric gas and then transferred inside the anaerobic glovebox. Each grid was then transferred inside a microcentrifuge tube containing the self-assembly reaction mixture. The microcentrifuge tubes with the grids inside were closed and left for 2 days inside the glovebox to react.

After two days have passed, these grids were removed from the anaerobic glovebox and mounted on a Vitrobot device. 2.5 µL of hydrogenase-1 sample (1.2 mg/mL) were placed onto each grid, blotted (blot force -5, blot time of 10 seconds, 100% humidity, 4°C), and the grid plunged into liquid ethane. Data collection followed.

# 3 *Thermotoga maritima* hydrogenase: Experimental structural studies and thermodynamic modelling

## 3.1 Introduction

This introduction focuses on *Thermotoga maritima* hydrogenase (*TmHydABC*) and bifurcating enzymes and is followed by the experimental results obtained using cryo-EM. These results will be integrated with the existing literature to provide a mechanistic proposal of the function of this hydrogenase.

Results have been published (171).

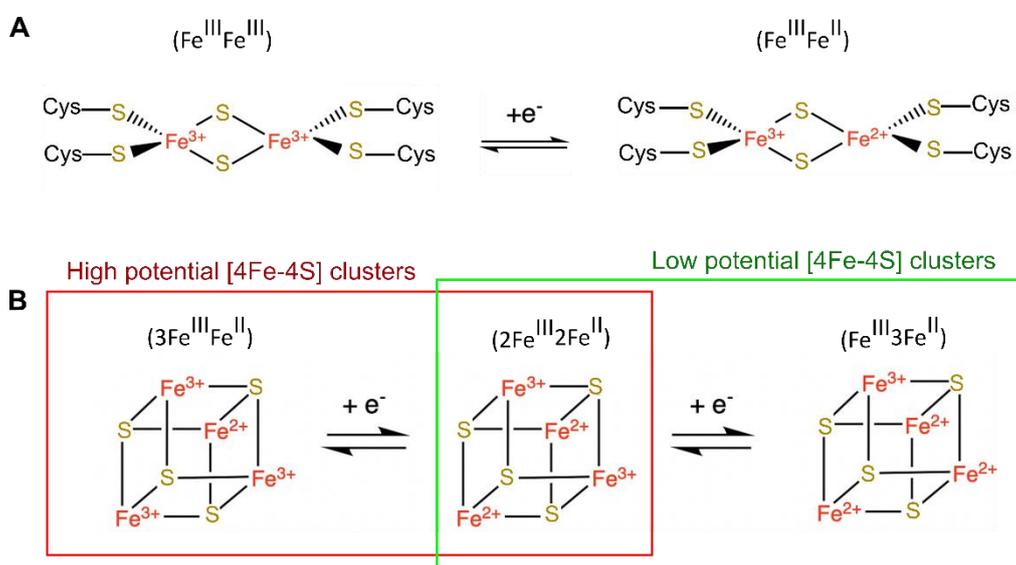
### 3.1.1 Iron-sulfur cluster redox reactions

In most hydrogenases, including *TmHydABC*, there are several iron-sulfur cluster cofactors, typically composed of two ([2Fe-2S]) or four iron centres ([4Fe-4S]) bridged by sulfide ions. These iron-sulfur clusters participate in the electron transfer chain by transferring one electron at a time to a neighbour iron-sulfur cluster within a 14 Å distance (11). During electron transfer, the oxidation state of the iron atoms varies between +2 and +3.

[2Fe-2S] clusters, physiologically, only exist in two redox states (Fig.3.1.A): based on formal charges, oxidised clusters contain two Fe<sup>3+</sup> ions (Fe<sup>III</sup>Fe<sup>III</sup>), whereas reduced ones (after receiving an electron) contain one Fe<sup>3+</sup> and one Fe<sup>2+</sup> ion (Fe<sup>III</sup>Fe<sup>II</sup>). [4Fe-4S] clusters can have three redox states (fig.3.1.B): based on formal charges, the most oxidised one has three Fe<sup>3+</sup> ions and one Fe<sup>2+</sup> ion (3Fe<sup>III</sup>Fe<sup>II</sup>), the intermediate state has two Fe<sup>3+</sup> ions and two Fe<sup>2+</sup> ion (2Fe<sup>III</sup>2Fe<sup>II</sup>), and the most reduced one has one Fe<sup>3+</sup> ions and three Fe<sup>2+</sup> ion (Fe<sup>III</sup>3Fe<sup>II</sup>). However, under physiological conditions, [4Fe-4S] clusters typically only function in two redox states: in higher potential [4Fe-4S] clusters, also known as (Hipip)-type, the redox state

varies between  $(3\text{Fe}^{\text{III}}\text{Fe}^{\text{II}})$  and  $(2\text{Fe}^{\text{III}}2\text{Fe}^{\text{II}})$  with redox couple potential ranging from +0.05 to +0.50 V (172, 173). In lower potential [4Fe-4S] clusters, also known as ferredoxin (Fd)-type, the redox state varies between  $(2\text{Fe}^{\text{III}}2\text{Fe}^{\text{II}})$  and  $(\text{Fe}^{\text{III}}3\text{Fe}^{\text{II}})$  with redox couple potential ranging from -0.10 to -0.70 V (174). Therefore under physiological conditions [4Fe-4S] clusters are single electron carriers. Typically, [4Fe-4S] clusters harboured in enzymes as part of electron transport chains (e.g. in complex I and hydrogenases) are lower potential [4Fe-4S] clusters (175).

The potential of the [4Fe-4S] clusters is tuned by the degree of hydrogen bonding between surrounding amino acids and sulfur atoms (176), and by histidine ligation which shifts the reduction potential of the cluster towards more positive values (177). More broadly, the potential is tuned by the local dielectric environment and coordination geometry of the ligands.

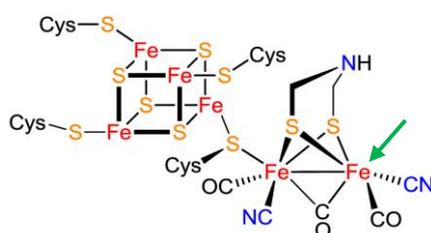


**Figure 3.1.** **A** Redox states of [2Fe-2S] clusters. **B.** Redox states of [4Fe-4S] clusters. The redox couples of high and low potential [4Fe-4S] clusters are shown inside red and green rectangles, respectively. Coordinating cysteine residues (Cys) can more rarely be replaced with histidine residues.

### 3.1.2 [FeFe]-hydrogenases: active site reactions

*Thermotoga maritima* hydrogenase (*TmHydABC*) is studied in this project; this enzyme is a [FeFe]-hydrogenase. [FeFe]-hydrogenases catalyse the reversible interconversion of protons ( $\text{H}^+$ ) and electrons to hydrogen ( $\text{H}_2$ ) and are considered to have the highest hydrogen evolution turnover numbers ( $k_{\text{cat}}$ ) among hydrogenases of up to  $10,000 \text{ s}^{-1}$  (178–180). They are

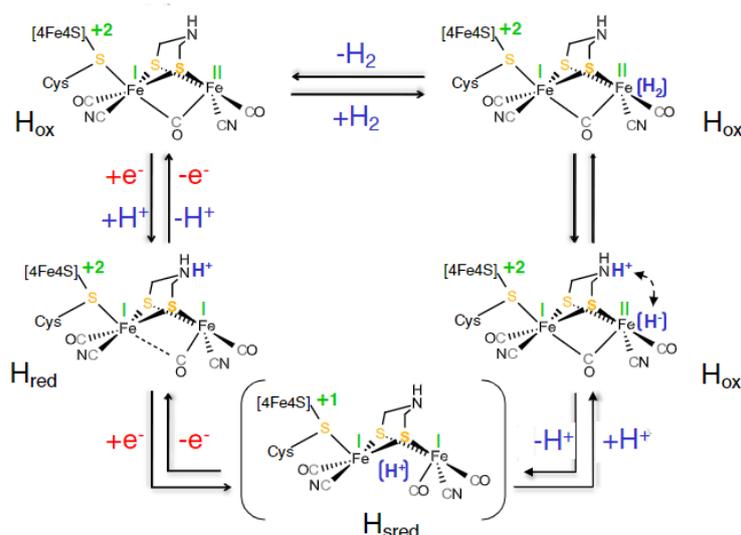
characterized by an “H-cluster” (the active site where H<sub>2</sub> is produced or oxidized) and accessory iron-sulfur clusters that transport electrons between active sites within the enzyme. The H-cluster is composed of a binuclear [FeFe] metal centre, which is covalently bound to a [4Fe-4S] cluster via a bridging cysteine residue of the protein (Fig. 3.2) (181). Each iron atom in the binuclear [FeFe] metal centre is coordinated to one CN<sup>-</sup> and one CO ligand (182). The two Fe atoms are connected through a bridging CO ligand and two bridging sulfur ligands of a bidentate 2-aza-propane-1,3-dithiolate (ADT) cofactor (152). The iron atom distal to the [4Fe-4S] cluster has an open coordination site where the substrate-binding site is located (green arrow in Fig.3.2) (183, 184).



**Figure 3.2** The active site structure of [FeFe] hydrogenase. The green arrow indicates the open coordination site. Adapted from Yu-Long Li *et al.* (184).

The H-cluster is buried inside a highly optimized protein scaffold, which tunes its catalytic efficiency and provides pathways for the transport of protons, electrons (electron transport chains), and gases (the substrate H<sub>2</sub> as well as inhibitors such as CO and O<sub>2</sub>) (185).

The H-cluster is cycled between states known as H<sub>ox</sub> ( [4Fe-4S]<sup>2+</sup>-Fe(I)Fe(II) ) and H<sub>sred</sub> ( [4Fe-4S]<sup>+</sup>-Fe(I)Fe(I) ) (186, 187). During the H<sub>2</sub> conversion, [FeFe]-hydrogenases pass through various oxidation states (188). In the active oxidized state (H<sub>ox</sub>), the [FeFe] metal centre is in a mixed valent [Fe(I)Fe(II)] state, and the [4Fe-4S] cluster is oxidized (2+). The H<sub>sred</sub> state is formed by two consecutive one-electron reductions of the [4Fe-4S] cluster and one iron centre coupled with protonation of the ADT (189). The active H<sub>ox</sub> state is restored by the release of an H<sub>2</sub> molecule (190, 191). Reverse reactions can be used to describe the H<sub>2</sub> oxidation steps. These species are shown in Fig.3.3 in the context of the catalytic cycle.

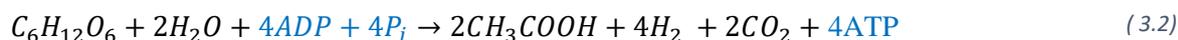


**Figure 3.3** Model of the catalytic cycle of [FeFe]-hydrogenases. Protonation of the ADT and one-electron reduction of the active oxidised state  $H_{ox}$  (top-left) results in the  $H_{red}$  state (bottom-left). The H-cluster in the  $H_{red}$  state is further reduced by receiving a single electron resulting in the  $H_{sred}$  state (bottom-center). The H-cluster in the  $H_{sred}$  state is protonated by a second  $H^+$  ion followed by the release of an  $H_2$  molecule. Although the catalytic cycle of [FeFe] hydrogenases is generally biased towards  $H_2$  production, it can also operate in the opposite direction with the oxidation of  $H_2$  to  $H^+$  involving a heterolytic splitting of hydrogen in  $H^+$  and  $H^-$  (right). Adapted from Martin Winkler *et al.* (192).

### 3.1.3 Hydrogenases in *Thermotoga maritima* metabolism

*Thermotoga maritima* is the most thermophilic eubacterium known and grows at temperatures of up to 90 °C in geothermally heated marine sediments (193) by a fermentative metabolism of simple and complex carbohydrates in which  $H_2$ ,  $CO_2$ , and organic acids (lactate and acetate) are the end products (194). The bacterium is strictly anaerobic.

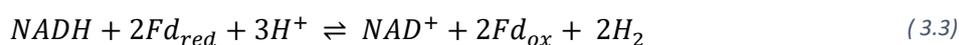
*T. maritima* employs the Embden-Meyerhof-Parnas (EMP) pathway (85% relative contribution), which is the most common glycolytic pathway, and the Entner-Doudoroff (ED) pathway (15% relative contribution) to convert glucose into pyruvate (Fig.3.4.B), which is then further converted into lactate (Equation 3.1) or acetate (Equation 3.2) (195).



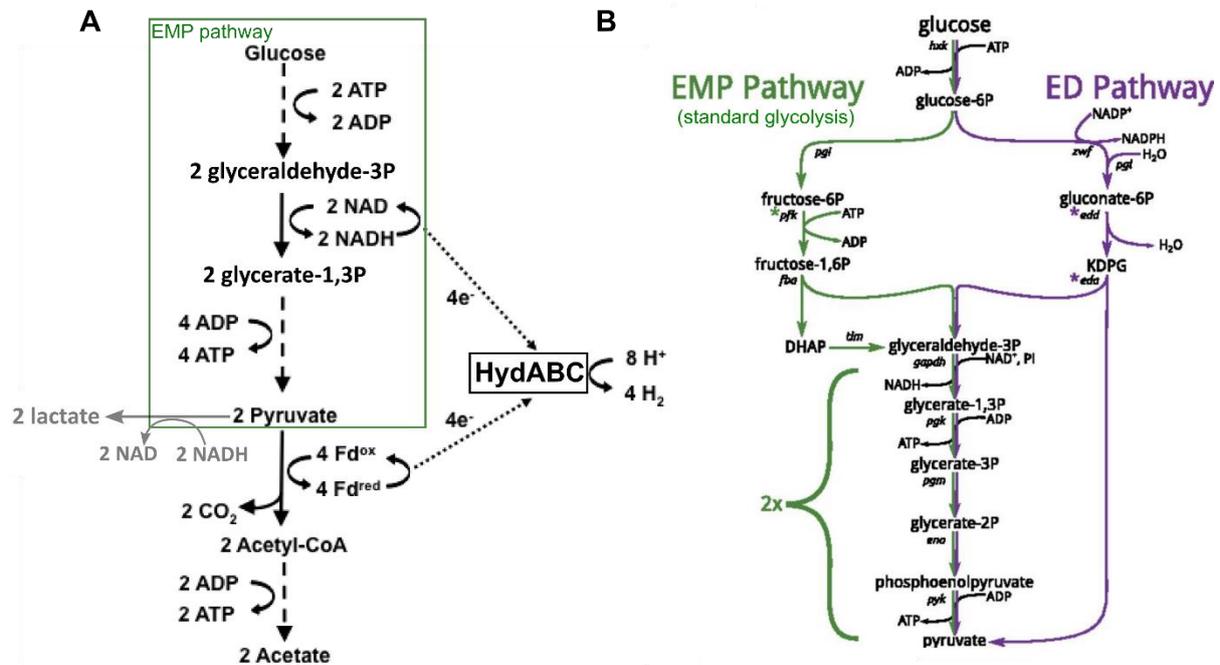
The predominant fermentation product in *T. maritima* is lactate (1.2 mol/mol glucose) (196, 197) and its production from pyruvate consumes all the NADH produced during glycolysis (balanced pathway). On the other hand, balancing the acetate fermentation pathway requires

the production of H<sub>2</sub> to get rid of the reducing equivalents from NADH and reduced ferredoxin (Fd<sub>red</sub>) (Fig.3.4.A). Only one ferredoxin from *T. maritima* was present in significant amounts during glucose fermentation, which was determined to be a one-electron donor/acceptor with a single iron-sulfur cluster (198, 199).

The production of H<sub>2</sub> in *T. maritima* is catalyzed by a single [FeFe]-hydrogenase located in the cytoplasm (*TmHydABC*) (194), which was shown to require the simultaneous presence of NADH and reduced ferredoxin (Equation 3.3) (200); in the absence of just one of these substrates, there was no significant catalytic activity.



In the above reaction, it is claimed that the exergonic oxidation of ferredoxin drives the endergonic oxidation of NADH. As electrons derived from two separate oxidation reactions are converged into a single reaction (H<sub>2</sub> evolution), *TmHydABC* is considered a confurcating enzyme; similarly, in the reverse reaction (observable *in vitro* (150)), the electron pairs derived from H<sub>2</sub> oxidation are thought to be split and transferred to both oxidized ferredoxin and NAD<sup>+</sup>, therefore constituting a bifurcating mechanism. When *TmHydABC* is referred to as bifurcating enzyme it is implied that it is also confurcating.



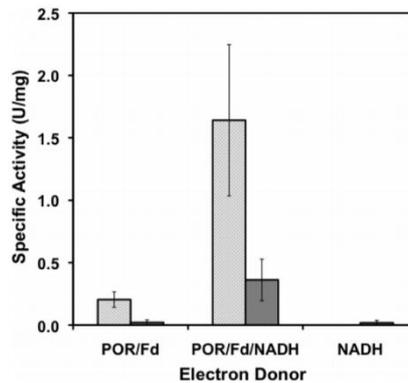
**Figure 3.4. A.** The role of the bifurcating [FeFe]-hydrogenase (*TmHydABC*) in the pathway of glucose-to-acetate conversion in *T. maritima*. Several consecutive enzymatic steps in the pathway are not shown and these are indicated by dashed arrows. In light grey, the alternative conversion to the fermentation product lactate is shown. Image adapted from Michael W. W. Adams *et al.* (200). **B.** A schematic of the EMP and ED pathways. The ED pathway is an alternative glycolytic pathway that only produces one ATP per glucose (half as much as the EMP pathway). The lower energy yield of the ED pathway is thought to be compensated by the lower amount of enzymatic protein required to catalyze pathway flux. Image adapted from Avi Flamholz *et al.* (201).

Only one additional hydrogenase has been isolated from *T. maritima*; this is a sensory [FeFe]-hydrogenase (*HydS*) and its exact role is unknown (202).

### 3.1.4 Discovery of the bifurcating *T. maritima* [FeFe]-hydrogenase and its properties

*T. maritima* [FeFe]-hydrogenase, *TmHydABC*, was first discovered in 1991 by Michael Adams and co-workers (194); they showed that the production of H<sub>2</sub> in *T. maritima* is catalysed by a single oxygen-sensitive hydrogenase located in the cytoplasm which was later better characterized and referred to as *TmHydABC*. In this study, Michael Adams and co-workers also showed that *T. maritima* hydrogenase had a substantially lower hydrogen evolution catalytic activity compared to mesophilic monomeric [FeFe]-hydrogenases, however, the affinity for H<sub>2</sub> in H<sub>2</sub> oxidation assays was determined to be comparable.

Importantly ferredoxin, the standard substrate of most known [FeFe]-hydrogenases, was determined not to function as an electron carrier to *TmHydABC* (194). The activity assay results in combination with the lack of activity in the sole presence of ferredoxin were the first hint of a more complex mechanism that was later discovered, in 2009, to require both NADH and ferredoxin to be present at the same time (Fig.3.5) (200).

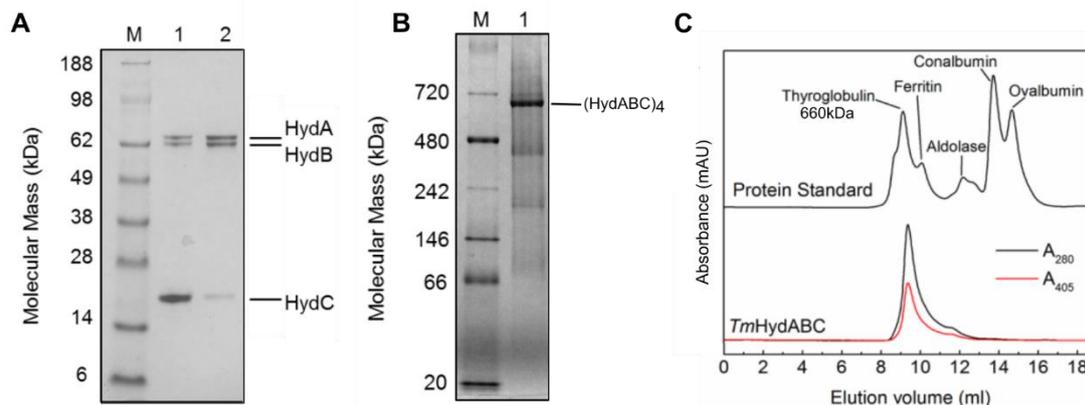


**Figure 3.5** Substrate dependence of the bifurcating [FeFe] hydrogenase of *T. maritima*. The specific activity in the standard H<sub>2</sub> production assay containing Fd/POR (Fd reduced enzymatically by its physiological partner POR), NADH, or Fd/POR and NADH was determined using the cytoplasmic extract (dark bars) and the purified hydrogenase (light bars). Taken from Adams MW *et al.* (200).

### ***Composition of T. maritima hydrogenase***

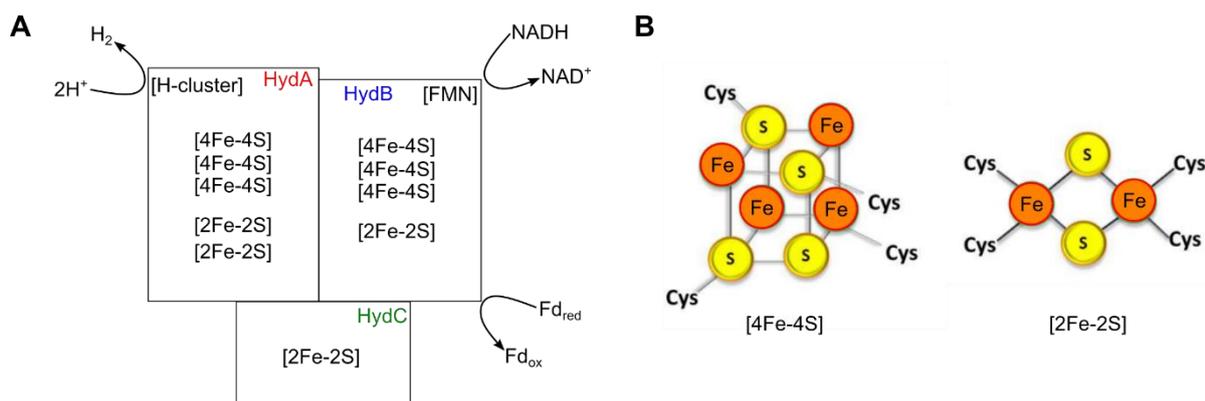
*T. maritima* hydrogenase was determined to be composed of three different subunits: HydA (74.2 kDa\*), HydB (69.8 kDa\*), and HydC (18.5 kDa\*) (Fig.3.6.A) (203). Native PAGE analysis and size exclusion chromatography showed the presence of a single oligomeric state (~650 kDa) at pH 8 (Fig.3.6.B, C), compatible with a tetramer of trimers (150), here indicated as (*TmHydABC*)<sub>4</sub>.

\*The molecular exact masses in brackets are calculated based on protein sequence (KEGG database, enzyme 1.12.1.4) and masses of iron-sulfur clusters.



**Figure 3.6.** **A** SDS-PAGE analysis of *T. maritima* hydrogenase holoprotein. Molecular weight markers (lane M), apo-*TmHydABC* after affinity chromatography (lane 1) and gel filtration chromatography (lane 2) **B**. Native PAGE analysis of purified apo-*TmHydABC* (lane 1). **C**. Analytical size exclusion chromatography of apo-*TmHydABC*. Images adapted from Chongdar N *et al.* (150).

From EPR spectroscopy, sequence, and metal analysis it was predicted that HydA, the largest subunit, harbours the H-cluster along with three additional [4Fe–4S] and two [2Fe–2S] clusters; HydB, the second largest subunit, was predicted to contain three [4Fe–4S] clusters, one [2Fe–2S] cluster, a flavin mononucleotide (FMN) binding domain, and an NADH-binding motif; HydC, the smallest subunits, was predicted to contain only one [2Fe–2S] cluster (Fig.3.7.A) (150, 204). These results were further confirmed in the cryo-EM structure determined in this project.



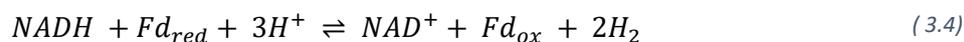
**Figure 3.7.A.** Predicted composition of the three subunits in *T. maritima* hydrogenase. The ferredoxin binding site is placed in HydB based on the cryo-EM structure obtained in this project (could not be predicted before). **B**. [4Fe–4S] cluster and [2Fe–2S] cluster. Coordinating cysteine residues can more rarely be replaced with histidine residues. Histidine ligation shifts the reduction potential of the cluster towards more positive values (177). Image adapted from Gentry-Dye Leslie (205).

### 3.1.5 Historical overview of proposed mechanisms for *T. maritima* hydrogenase

Based on the available spectroscopic, biochemical, and sequence information, multiple mechanisms were proposed over time, however, no consensus has been reached. Initially, it was not known the simultaneous requirement of both NADH and ferredoxin and only assays with cell extracts (not purified enzyme) with the addition of NADH showed measurable H<sub>2</sub> evolution activity (198). This led to a speculative electron transfer mechanism (Fig.3.8.A) which was only accurate in the assignment of the number and types of iron-sulfur clusters and NADH and H<sub>2</sub> binding sites to the three subunits (HydA, HydB, HydC).

In 2009, it was claimed that *TmHydABC* constituted a new class of bifurcating [FeFe]-hydrogenase in which the exergonic oxidation of ferredoxin is used to drive the unfavourable oxidation of NADH to produce H<sub>2</sub> (200). No detailed mechanism was proposed and the bifurcation site remained unclear (Fig.3.8.B).

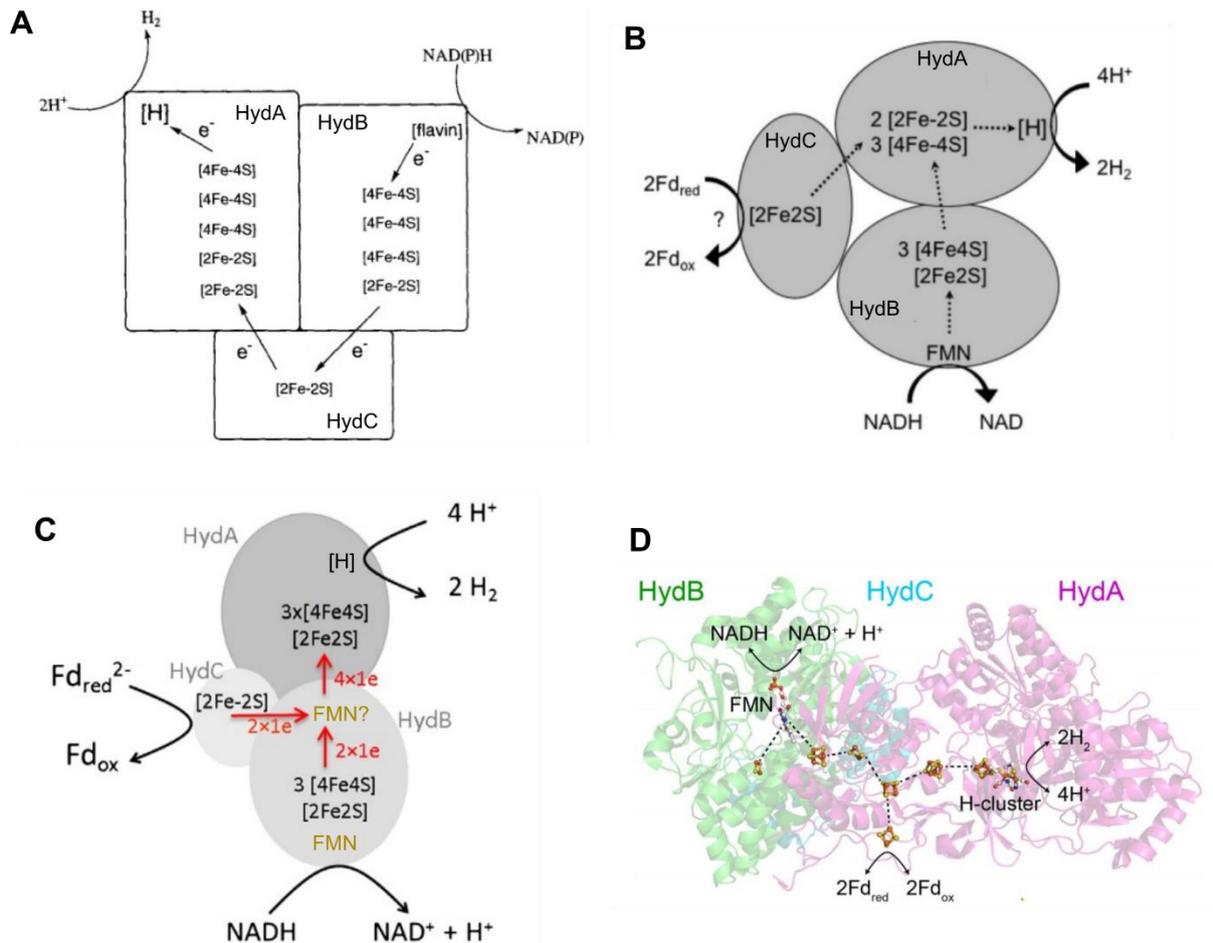
In a more recent literature review, in 2013, from Thauer and Buckel, an alternative model was presented where two FMN are bound to *TmHydABC* (206); one FMN is the site of NADH dehydrogenation and the other acts as flavin-based electron-bifurcation (Fig.3.8.C). The presence of a second FMN was postulated without experimental evidence. Thauer and Buckel also proposed that NADH and ferredoxin are oxidised in a 1:1 ratio based on Equation 3.4 referring to *clostridial*-type ferredoxins, while the ratio of NADH to *T. maritima* ferredoxin (that contains one single [4Fe-4S] cluster) is 1:2 as confirmed by recent experiments by Chongdar *et al.* (Equation 3.3) (150).



In a recent study (2019) by Chongdar *et al.*, recombinant and artificially matured *T. maritima* hydrogenase was characterized biochemically and spectroscopically (150). They showed that in addition to the already demonstrated electron confurcation (electrons are converged from NADH and Fd<sub>red</sub> to reduce protons to H<sub>2</sub>), also electron bifurcation (H<sub>2</sub>-dependent reduction of NAD<sup>+</sup> and Fd<sub>ox</sub>) occurs.

This study from Chongdar *et al.* also showed that the presence of a second FMN, as postulated by Thauer and Buckel, was unlikely based on FMN quantification combined with activity observations. Furthermore, Chongdar *et al.* suggested that the remaining single FMN is not the

site of bifurcation/confurcation based on sequence similarity between HydB and the NuoF (Nqo1) subunit of respiratory complex I which does not involve a bifurcation reaction. In addition, they proposed that the H-cluster in *TmHydABC* is unlikely the site of electron bifurcation as, based on EPR and FTIR data, it behaves similarly to the H-cluster in prototypical [FeFe]-hydrogenases. Based on this data and sequence similarity to complex I, Chongdar *et al.* proposed their speculative electron flow model (Fig.3.8.D), where three iron-sulfur clusters in HydA are the bifurcation site and ferredoxin interacts with a His-ligated [4Fe-4S] cluster of HydA. This model was accurate in predicting a single FMN, however, our cryo-EM data shows that the His-ligated [4Fe-4S] cluster cannot interact with ferredoxin as this is proximal to a second ABC trimer (171).



**Figure 3.8.** **A** Model of the *T. maritima* hydrogenase and its electron flow incorporating the available spectroscopic, biochemical, and sequence information available in 2001. [H] represents the H-cluster where proton reduction is thought to take place. Figure taken from M W Adams *et al.* (198). **B.** The proposed pathway of electron flow during the oxidation of reduced Fd and NADH by *T. maritima* hydrogenase. The dashed arrows indicate electron flow. Image taken from M W Adams *et al.* (200). **C.** Schematic representation of the structure and function of the HydABC complex from *Thermotoga maritima*. The enzyme is proposed to harbour a second flavin (the one with the question mark and for which there is no experimental evidence). The site of ferredoxin interaction was chosen arbitrarily. Image taken from Tauer and Buckel (206). **D.** The arrangements of the cofactors in *TmHydABC* are predicted based on the homology to the Nqo1 (HydB, green), Nqo2 (HydC, blue), and Nqo3 (HydA, pink) subunits of the structurally characterized complex I from *Thermus thermophilus*. Image taken from Chongdar *et al.* (150).

### 3.1.6 Homologous proteins to *T. maritima* HydABC

From genome sequence analysis, it became clear that there are two major types of [FeFe]-hydrogenases: trimeric putatively bifurcating hydrogenases and the more well-studied

monomeric ferredoxin-dependent [FeFe]-hydrogenases (200). Almost one-third of the known H<sub>2</sub>-producing anaerobes appear to contain trimeric bifurcating enzymes that have significant homology to *T. maritima* hydrogenase, although many of them also contain one or more homologs of the simpler ferredoxin-dependent hydrogenase (200).

Examples of trimeric bifurcating [FeFe]-hydrogenases that are homologous to *T. maritima* HydABC (*TmHydABC*) are HydABC from the acetogenic bacteria *Acetobacterium woodii*, *Thermoanaerobacter kivui* (207), *Ruminococcus albus* (208), and *Moorella thermoacetica* (209). Some homologous [FeFe]-bifurcating hydrogenases have additional small subunits nevertheless they maintain the overall similarity, for example, the tetrameric HndABCD hydrogenase from *Desulfovibrio fructosovorans* (210).

Some putatively bifurcating [NiFe]-hydrogenases are also homologous to *T. maritima* hydrogenase, for example, this is the case for NiFe-HydABCSL hydrogenase that reversibly oxidizes H<sub>2</sub> and couples the endergonic reduction of ferredoxin with the exergonic reduction of NAD<sup>+</sup> (211).

Additional homology can be found between subunits of *TmHydABC* and complex I (HydA and Nqo3, HydB and Nqo1, HydC and Nqo2) and between HydA and the monomeric [FeFe]-hydrogenase from *Clostridium pasteurianum*, CpI. This indicates common evolutionary origins (212).

### 3.1.7 Definition of electron bifurcation/confurcation

There are multiple variations of the definition of electron bifurcation/confurcation. Commonly electron bifurcation/confurcation is defined as the coupling of exergonic and endergonic redox reactions to simultaneously generate (or utilize) low- and high-potential electrons (213) or as a mechanism to drive a thermodynamically unfavourable redox reaction through direct coupling with an exergonic reaction (214).

However, moving forward I will use a more explicit definition of electron bifurcation/confurcation mainly for the following two reasons:

1. Electron pair splitting is sometimes not explicitly mentioned in the various definitions: endergonic and exergonic redox reactions could be coupled in alternative ways (currently not known in enzymes) that do not involve electron pair splitting. I argue the

splitting of electron pairs is instead the key requirement for the definition and therefore should always be specified.

2. It is implied that bifurcation/confurcation only couples an exergonic and an endergonic reaction. In reality, bifurcating/confurcating enzymes can also couple two exergonic reactions as will be clear from more explicit thermodynamical considerations (Chapter 3.1.11).

My version of the definition is:

*“Electron bifurcation (or confurcation) is a mechanism in which the spatial and/or temporal splitting (or converging) of pairs of electrons allows thermodynamic coupling of two reactions in an overall spontaneous process ( $\Delta G_{tot} < 0$ ).”*

Note that this is a general definition that is also valid for potential artificial systems. In enzymes, thermodynamic coupling requires spatial splitting (i.e. electrons from an electron pair proceeding to different spatial locations); in fact, any common intermediate cofactor in the pathway of electron transfer after electron pair splitting breaks the thermodynamic coupling (at equilibrium:  $E_{Aox/Ared} = E_{Cofactor\_in\_common} = E_{Box/Bred}$ ). In addition, as later discussed, kinetic coupling (also known as ‘stoichiometric’ coupling) is a **necessary but not sufficient** condition for thermodynamic coupling (also known as ‘energetic’ coupling) in both natural and artificial systems.

Because of the thermodynamic coupling, electron bifurcation (confurcation) can be used by enzymes (bifurcating/confurcating enzymes) to drive the reaction with the lowest electromotive force (213). In this sense, electron bifurcation/confurcation (214) is considered an alternative energy coupling mechanism to the well-known chemiosmotic coupling principle (215).

### 3.1.8 Potential inversion in electron bifurcation/confurcation

In all characterized and understood electron confurcation/bifurcation mechanisms a redox cofactor (e.g. FAD in EtfAB) or donor substrate (e.g. quinone in complex  $bc_1$ ) with three accessible oxidation states and with inverted potential acts as the bifurcating/confurcating centre. In a molecule or ion with inverted potentials (216), the loss of the first electron is less

favourable than the loss of a second one, which is opposite to what is normally encountered (normally the reduction potential of a redox species increases as more electrons leave).

Therefore, a fully reduced redox centre with inverted potentials requires a relatively higher potential acceptor to extract the first electron ('harder to get rid of') while the second electron ('easier to get rid of') can be extracted by a lower potential acceptor. Similarly, a fully oxidized redox centre with inverted potential requires that the first electron ('harder to receive') is donated by a relatively lower potential donor while the second electron ('easier to receive') can be donated by a higher potential donor. This splitting/converging of electrons towards/from higher and lower potential acceptors/donors is the key that allows the coupling of reactions, as will be clearer in the following subchapters.

In Chapter 3.1.11 I will show that potential inversion is a kinetic and not a thermodynamic requirement for bifurcation/confurcation.

*Examples of inverted potential species with three accessible oxidation states*

Flavin mononucleotide (FMN), like all flavins, has three accessible oxidation states: the fully oxidized FMN, the semi-reduced  $FMN^{\bullet-}$ , and the fully reduced  $FMNH^-$  (Equation 3.5). The semi-reduced flavin radical is less stable than  $FMNH^-$  causing the inverted potential phenomena.



The tabulated standard redox potentials show that the redox potential of the  $FMN /FMN^{\bullet-}$  couple is lower than the  $FMN^{\bullet-}/FMNH^-$  couple in agreement with the presence of inverted potential (Table 3.1) which is characteristic of all flavins when in solution and in most proteins. In some cases, semi-reduced flavins are stabilized by the protein environment, and inverted potentials are not observed, for example, this is the case in a flavodoxin (a small protein containing a flavin) from *A. fermentans* (217, 218).

**Table 3.1** Standard redox potentials of the FMN /FMN<sup>•-</sup> and FMN<sup>•-</sup>/FMNH<sup>-</sup> couples in water at pH = 7 as determined by R. Anderson (219). The values here reported are the estimate of the redox potentials for a typical flavin, as the values have very little variation between flavins (FAD, FMN, riboflavin) comparable to experimental errors.

Redox couple	E <sup>0</sup> (mV)
$FMN + 1e^- \rightarrow FMN^{\bullet-}$	-314
$FMN^{\bullet-} + 1e^- \rightarrow FMNH^-$	-124
$FMN + 2e^- \rightarrow FMNH^-$	$\frac{-314-124}{2} = -219$

Another example of a species with three accessible oxidation states and inverted potential is copper in aqueous solutions (Table 3.2). Removing a single electron from a metal Cu atom results in a destabilized Cu<sup>+</sup> ion while the subsequent extraction of a second electron results in a more stable Cu<sup>2+</sup> ion. Therefore Cu will get rid of a single electron less easily (higher reduction potential, +520 mV) than Cu<sup>+</sup> (lower reduction potential, +154 mV).

**Table 3.2** Standard redox potentials of the Cu<sup>2+</sup>/Cu<sup>+</sup> and Cu<sup>+</sup>/Cu couples in water. (220)

Redox couple	E <sup>0</sup> (mV)
$Cu^{2+} + 1e^- \rightarrow Cu^+$	+153
$Cu^+ + 1e^- \rightarrow Cu$	+521
$Cu^{2+} + 2e^- \rightarrow Cu$	$\frac{153+521}{2} = +337$

**\*Notes on nomenclature:** E<sup>0</sup> represents the redox potential in standard conditions (25°C, 1 atm, all species at 1M concentration). The redox potential in standard conditions but at pH = 7 ([H<sup>+</sup>]=10<sup>-7</sup>M instead of 1M) is usually indicated in biochemistry as E<sup>0'</sup>. Sometimes the numerically equivalent midpoint potential at 25°C, 1 atm, and pH=7 is used and indicated as E<sub>m</sub>'.

### 3.1.9 Classical models of electron bifurcation/confurcation

In all characterized enzymes, bifurcation and confurcation are reversible, and therefore the bifurcating centre also acts as a confurcating centre depending on reaction conditions. Therefore, typically, the centre is simply called a “bifurcating centre” even when referring to the confurcating reaction. Similarly, an enzyme observed to catalyze in both bifurcation and

confurcation directions is simply referred to as a “bifurcating enzyme”. To avoid confusion, I will use this convention but I will specify if the reaction is not reversible.

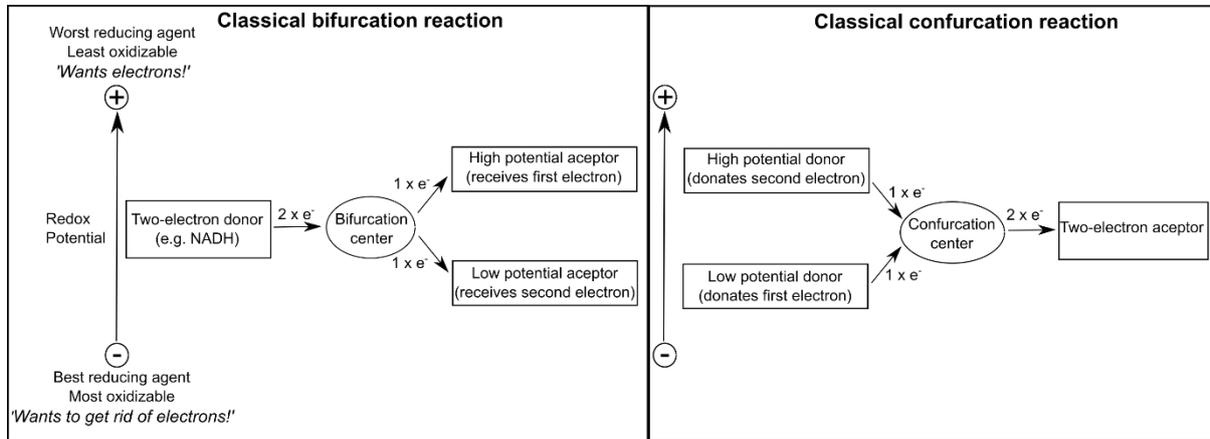
Currently, two types of electron bifurcation are known and sufficiently well understood: quinone-based electron bifurcation (QBEB) and flavin-based electron bifurcation (FBEB) (221) in which quinones and flavins, respectively, are the bifurcating centres. QBEB is only found in cytochrome *bc*<sub>1</sub> and analogous complexes (e.g. cytochrome *b<sub>6</sub>f* complex from plant chloroplasts and cyanobacteria) while FBEB is utilized by a more varied class of enzymes (221). In all characterized enzymes using QBEB, quinone (in the fully reduced form) is both a substrate and the bifurcating centre, while in all characterized enzymes using FBEB the bifurcating centre is a flavin cofactor which receives two electrons concertedly from NADH or NADPH. Other bifurcating enzymes, including *TmHydABC*, do not use NADH or NADPH as two-electron donors (in the bifurcation direction) but are suspected to use FBEB (debatable) although with a poorly understood mechanism. No enzyme is known to utilize flavin-based electron confurcation (FBEC) irreversibly and therefore FBEC is simply the reverse of FBEB. Sometimes the terms FBEB and FBEC are used interchangeably.

The characterized quinone- and flavin-based electron bifurcation/confurcation mechanisms were used to construct a more generalized model of electron bifurcation/confurcation (Fig.3.9) (221), which I will here call the ‘classical model’.

In the classical model of bifurcation (Fig.3.9, left), a fully oxidized bifurcating redox centre becomes fully reduced after receiving two electrons from a two-electron donor (221); then the fully reduced bifurcating redox centre donates its first electron (“harder to donate”) to the highest potential substrate; the now semi-reduced bifurcating redox cofactor donates its second electron (“easier to donate”) to the lowest potential substrate. Movements and rearrangements of parts of the enzyme prevent electron flow between the two acceptor substrates and ensure that the second electron cannot be transferred to another higher potential acceptor molecule.

Similarly, in a typical confurcation reaction (Fig.3.9, right), the fully oxidized bifurcating redox cofactor receives its first electron from the lowest potential acceptor and the second electron from the highest potential acceptor. Movements and rearrangements of the enzyme are necessary also in the confurcation direction.

The key aspect of the classical bifurcation model is that the bifurcation centre is fully reduced before splitting its two electrons into two spatially and temporally separate high and low potential pathways.



**Figure 3.9.** The classical model of a bifurcation (left) and confurcation (right) reaction. Note: the bifurcating centre can also be a substrate of the bifurcating enzyme (e.g. quinone in complex bc1), in this case, it binds to/is released from the bifurcating enzyme fully reduced.

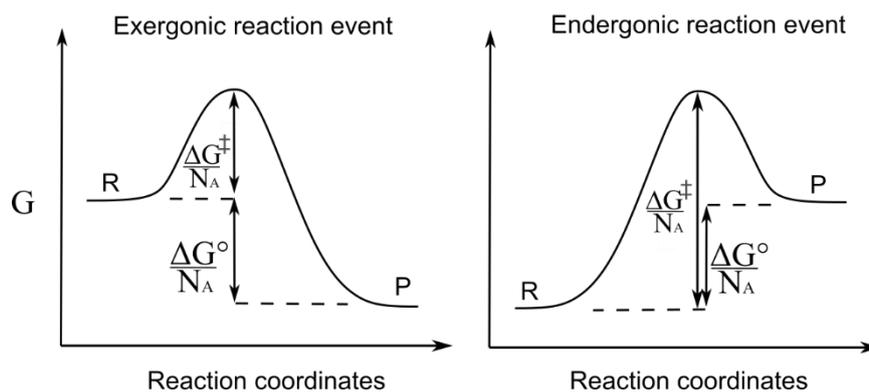
### 3.1.10 Useful general thermodynamic concepts

The reaction potential,  $E$ , is tied to the concept of Gibbs's free energy. In a reversible reaction/process at constant temperature and pressure, the variation of Gibbs's free energy ( $\Delta G$ ) is equal to the non-expansion work ( $w_{non-exp}$ ). Therefore, in a redox reaction at constant pressure at temperature, the  $\Delta G$  of the reaction can be calculated as:

$$\Delta G = w_{non-exp} = -nFE \quad (3.6)$$

Where  $n$  is the number of electrons transferred in the cell reaction or half-reaction and  $F$  is the Faraday constant ( $9.65 \times 10^4 \text{ C/mol}$ ).

When substituting  $E$  in Equation 3.6 with the standard potential of the reaction ( $\Delta E^\circ$ ) then the standard Gibbs free energy can be calculated ( $\Delta G^\circ$ ). The value of  $\Delta G^\circ$  provides an intrinsic measure of the change in Gibbs free energy in a single reaction event ( $\Delta G^\circ / N_A$ ) and can also be calculated from the standard Gibbs energy of formation. The reaction event is exergonic if  $\Delta G^\circ < 0$ , and endergonic if  $\Delta G^\circ > 0$  (Fig.3.10).



**Figure 3.10.** Exergonic (left) and endergonic (right) reaction events.  $\Delta G^\ddagger$ : Gibbs free energy of activation.  $\Delta G^\circ$ : standard Gibbs free energy.  $N_A$ : Avogadro number ( $6.02214076 \cdot 10^{23}$ ). R: Reactants (e.g.  $A_{ox} + B_{red}$ ). P: Products (e.g.  $A_{red} + B_{ox}$ ).

However, it is important to distinguish between the energetics of a single reaction event and those of the global reaction (a collection of single reaction events). In fact, when assessing whether a redox reaction (meant as a collection of reaction events) is spontaneous or not, the endergonicity or exergonicity of a single reaction event is not sufficient. In fact, whether a redox reaction is endergonic or exergonic depends on  $\Delta G$  which in turn depends on the redox potential of the reaction (E, Formula 3.6) which is strongly dependent on the concentration of reactants and products (Formula 3.7). If  $\Delta G < 0$  the reaction is exergonic and therefore spontaneous, if instead  $\Delta G > 0$ , the reaction is endergonic, and therefore an additional driving force is needed.

In simple terms,  $\Delta G^\circ$  influences the energy barriers making the reaction events in the exergonic direction faster/slower (Chapter 1.2.2) compared to the endergonic direction, while  $\Delta G$  takes into account the concentrations of reactants and products (e.g. a lot of slow reactions in the endergonic direction and too few fast reactions in the exergonic direction make the overall reaction in the endergonic direction exergonic).

In a reversible electrochemical reaction at constant temperature and pressure, the Nernst equation (Formula 3.7) can be used to calculate how the redox potential changes from the standard values based on concentration:

$$E = E^0(T) - \frac{RT}{nF} \ln Q_r \cong E^0(T) - \frac{RT}{nF} \ln \frac{[Red]}{[Ox]} \quad (3.7)$$

In this formula  $E^0(T)$  is the redox potential in standard conditions (1 atm, all species at 1M concentration) at the specified temperature T, R is the universal gas constant ( $8.31 \text{ J} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ ), T is the temperature in Kelvins, and  $Q_r$  is the reaction quotient that can be approximated by using the concentration of reduced and oxidized species in the reaction.

In particular when  $T = 298\text{K}$  ( $25^\circ\text{C}$ ), Formula 3.7 can be more conveniently expressed as:

$$E = E^0 - \frac{0.0592}{n} \log_{10} \frac{[Red]}{[Ox]} \quad (3.8)$$

Equation 3.8 will be useful in the following subchapter.

### 3.1.11 Thermodynamics in the classical model of electron bifurcation

Based on the concepts described above, I will model a classical bifurcation reaction mechanism. As described in Chapter 3.1.9, a classical bifurcating reaction requires a redox cofactor with three accessible oxidation states and inverted potential which receives two electrons from a donor substrate; the bifurcating redox cofactor then transfers sequentially one electron to a higher potential acceptor and a second electron to lower potential acceptor. A classical bifurcation reaction is also reversible (there isn't any mechanism that prevents bidirectional electron flow within the enzyme, i.e. confurcation), however, portions of the protein move to prevent electron flow ('electrical connection') between the two one-electron electron acceptor/donor substrates (Fig.3.13.B). Protein movements also prevent the transfer of the second electron from the bifurcating cofactor to a second high-potential acceptor molecule (bifurcation direction); or that a second low-potential one-electron donor molecule donates the second electron to the bifurcating cofactor (confurcation direction).

To make the model more concrete, I will choose FMN as a bifurcating centre, NADH as the two-electron donor, ferredoxin (Fd) as a lower potential one-electron acceptor\*, and cytochrome c (Cyt c) as a higher potential one-electron acceptor (other examples are a semi-reduced flavin or semi reduced quinone that becomes fully reduced, a Rieske cluster, etc.). As FMN, is a flavin, this is an example of FBEB.

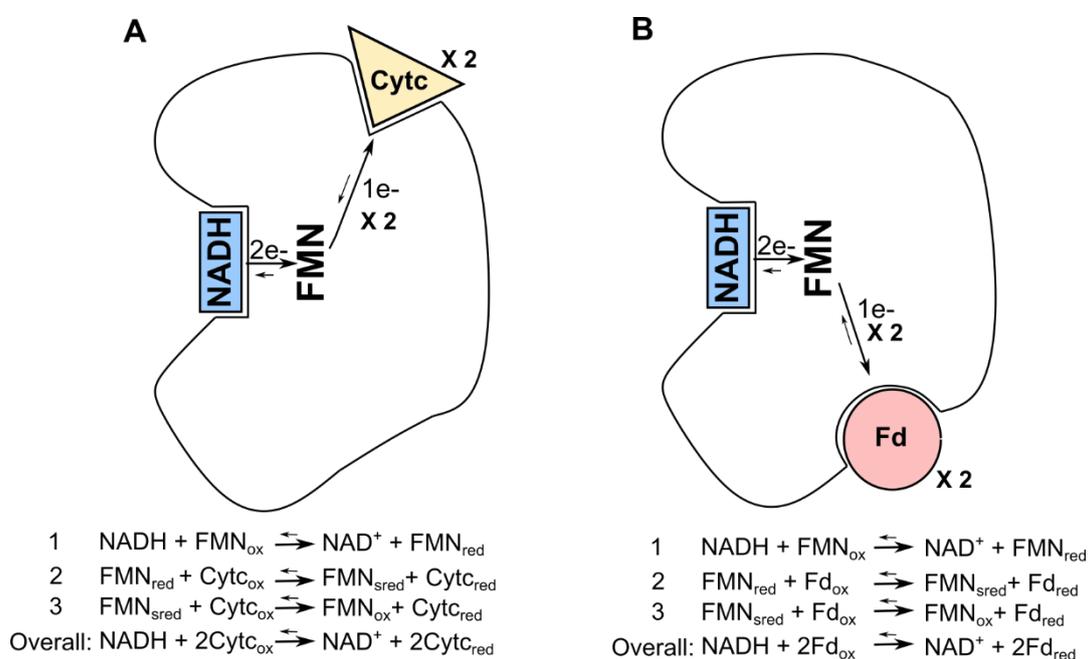
Initially, I will model a non-bifurcating version of the reaction (for comparison). Then I will model the corresponding bifurcation version. To conclude, I will present an example of an enzyme performing kinetic coupling but not thermodynamic coupling as this will be useful in my later discussions.

\* Some ferredoxins have two iron-sulfur clusters and therefore can sequentially receive two electrons (one for each cluster), however, the reduction potential remains almost constant in both separate reductions (227). Here for conceptual simplicity, a ferredoxin with a single iron-sulfur cluster is chosen (as is the case for *T. maritima* ferredoxin) and its reduction potential corresponds to the value of *C. kluyveri* ferredoxin ( $E^{\circ} = -420 \text{ mV}$ ). In general, the reduction potential of ferredoxins is low and varies between  $-500$  and  $-340 \text{ mV}$  (224).

## Non-bifurcating version

### Premises

Two different types of enzymes are present in the same solution (Fig.3.11). Both enzymes contain an FMN redox cofactor that receives two electrons concertedly from NADH becoming fully reduced ( $\text{FMN}_{\text{red}}$ ). In one enzyme the fully reduced FMN transfers sequentially both electrons to two molecules of oxidized ferredoxin ( $\text{Fd}_{\text{ox}}$ ) while in the other both electrons are sequentially transferred to two molecules of oxidized cytochrome c ( $\text{Cytc}_{\text{ox}}$ ). The reactions are reversible, and it is also assumed that any non-catalyzed direct redox reaction between NAD(H), ferredoxin, and cytochrome c is too slow and therefore negligible.



**Figure 3.11** Hypothetical non-bifurcating enzymes. **A.** NADH transfers two electrons concertedly to the fully oxidized FMN (cofactor). The fully reduced FMN then transfers one electron to a molecule of oxidized Cytc. The reduced Cytc molecule is released. Another molecule of oxidized Cytc binds and receives the second electron and is released. **B.** The same mechanism of enzyme a, however, with Fd instead of Cytc. Reactions are reversible as shown by the shorter arrows.

The reactions are imagined to be conducted at 25°C, atmospheric pressure, and buffered pH = 7. The ratio of NADH and  $\text{NAD}^+$  in the solution remains constant and has a value of 1:1 (e.g. by using a regeneration system (222)). Oxidized ferredoxin ( $\text{Fd}_{\text{ox}}$ ) and oxidized Cytc ( $\text{Cyt}_{\text{ox}}$ ) are added to the solution containing these hypothetical non-bifurcating enzymes (Fig.3.11). The protein environment significantly shifts the redox potentials of the bound cofactors and here the redox potentials of FMN as measured in complex I can qualitatively give an idea of

these changes (these values will change depending on the specific protein environment). The standard redox potential of all these species at pH=7 ( $E'^0$ ) is listed in Table 3.4.

### Analysis

When looking at the reduction potentials  $E'^0$  (Table 3.4) it is clear that in standard conditions and pH=7, the reduction of  $Fd_{ox}$  by NADH is endergonic ( $\Delta E'^0 = -420+320 \text{ mV} = -100 \text{ mV}$ ,  $\Delta G'^0 = nF\Delta E'^0 = +19.3 \text{ kJ/mol}$ ) while the reduction of  $Cytc_{ox}$  by NADH is exergonic ( $\Delta E'^0 = 254+320 \text{ mV} = 574 \text{ mV}$ ,  $\Delta G'^0 = -110.8 \text{ kJ/mol}$ ). However, these potentials significantly change based on concentrations, more precisely based on the ratio of reduced and oxidized species according to the Nernst equation (Equation 3.8).

When only  $Fd_{ox}$  and  $Cytc_{ox}$  are present (at the beginning), both reductions by NADH are exergonic and remain so until equilibrium is reached. When two or more reactions are in equilibrium their reduction potentials are equal (from  $\Delta G = 0$ , see Equation 3.6) and this can be used to calculate the concentrations at equilibrium.

It follows that based exclusively on the catalyzed reactions (Fig.3.11) all the following equalities must be satisfied at equilibrium (oxidized, semi-reduced, and reduced FMN indicated as  $FMN_{ox}$ ,  $FMN_{sred}$ ,  $FMN_{red}$  respectively):

$$\begin{aligned} E_{NADH/NAD^+} &= E_{FMN_{ox}/FMN_{red}} & E_{FMN_{sred}/FMN_{red}} &= E_{Fd_{ox}/Fd_{red}} \\ E_{FMN_{sred}/FMN_{red}} &= E_{Cytc_{ox}/Cytc_{red}} & E_{FMN_{sred}/FMN_{ox}} &= E_{Fd_{ox}/Fd_{red}} \\ E_{FMN_{sred}/FMN_{ox}} &= E_{Cytc_{ox}/Cytc_{red}} \end{aligned}$$

However,  $E_{FMN_{ox}/FMN_{red}} = \frac{1}{2}(E_{FMN_{sred}/FMN_{red}} + E_{FMN_{sred}/FMN_{ox}})$  and therefore by combining this with the above equalities we obtain that:

$$E_{NADH/NAD^+} = E_{Cytc_{ox}/Cytc_{red}} = E_{Fd_{ox}/Fd_{red}} \quad (3.9)$$

In Equality 3.9 the reduction potentials of the FMN couples are not included and this demonstrates that the redox potentials of the cofactor FMN do not influence the concentrations of reduced and oxidized Cytc and Fd at equilibrium, however, it can have a profound influence on the speed of the reactions. Equality 3.9 also demonstrates that the equilibrium concentrations in this non-bifurcating version would be identical to the ones in the absence of enzymes, however, the reactions would be kinetically too slow to sustain life.

Furthermore, equilibrium concentrations can be shown to be identical to the ones in which NADH can only transfer electrons sequentially due to the lack of a cofactor that can accept two

electrons (provable using equalities with the reduction potential of the  $\text{NAD}^+/\text{NAD}\cdot$ ,  $\text{NAD}\cdot/\text{NADH}$ ,  $\text{NAD}^+/\text{NADH}$  and couples); however, also, in this case, the reaction would be kinetically too slow to sustain life. In particular, NADH is an extreme case of potential inversion (see Table 3.4) with a very unstable semi-reduced radical, and therefore sequential electron transfer with standard biological redox partners is considered kinetically so slow to be relevant. In fact, in biochemical contexts, NADH is often referred to as an obliged two-electron donor/acceptor.

Having said this, we can continue with the calculation of the concentration at equilibrium.

By substituting Equation 3.8 in Equality 3.9, and the tabulated  $E'^0$  values for the  $\text{NAD}^+/\text{NADH}$  couple (-0.320 V) and  $\text{Cyt}_{\text{ox}}/\text{Cyt}_{\text{red}}$  (+0.254 V) it is possible to calculate the relative concentration of oxidized and reduced Cytc at equilibrium:

$$-0.320 - \frac{0.0592}{2} \log_{10}(1) = 0.254 - \frac{0.0592}{1} \log_{10} \frac{[\text{Cyt}_{\text{red}}]}{[\text{Cyt}_{\text{ox}}]} \quad (3.10)$$

The solution of Equation 3.10 is:

$$\begin{aligned} \frac{[\text{Cyt}_{\text{red}}]}{[\text{Cyt}_{\text{ox}}]} &= 10^{9.7} = 5 \cdot 10^9 \\ \frac{[\text{Cyt}_{\text{red}}]}{[\text{Cyt}]_{\text{tot}}} \cdot 100 &= \frac{[\text{Cyt}_{\text{red}}]}{[\text{Cyt}_{\text{ox}}] + [\text{Cyt}_{\text{red}}]} \cdot 100 \approx 100\% \end{aligned} \quad (3.11)$$

Similarly, using the tabulated  $E'^0$  values for the  $\text{NAD}^+/\text{NADH}$  couple and  $\text{Fd}_{\text{ox}}/\text{Fd}_{\text{red}}$  couple (-0.420 V) it is possible to calculate the relative concentration of oxidized and reduced Fd at equilibrium:

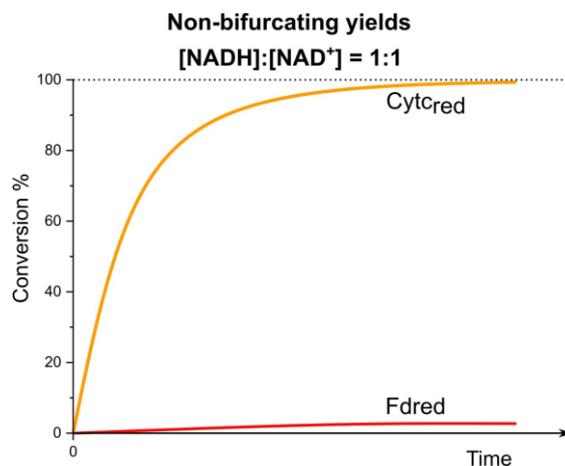
$$-0.320 - \frac{0.0592}{2} \log_{10}(1) = -0.420 - \frac{0.0592}{1} \log_{10} \frac{[\text{Fd}_{\text{red}}]}{[\text{Fd}_{\text{ox}}]} \quad (3.12)$$

The solution of Equation 3.10 is:

$$\begin{aligned} \frac{[\text{Fd}_{\text{red}}]}{[\text{Fd}_{\text{ox}}]} &= 10^{-1.7} = 0.020 \\ \frac{[\text{Fd}_{\text{red}}]}{[\text{Fd}]_{\text{tot}}} \cdot 100 &= \frac{[\text{Fd}_{\text{red}}]}{[\text{Fd}_{\text{ox}}] + [\text{Fd}_{\text{red}}]} \cdot 100 = 2.0\% \end{aligned} \quad (3.13)$$

The equilibrium concentrations of these non-bifurcating reactions are illustrated in Fig.3.12. The plot shows that at equilibrium almost all (~100%) Cytc is in the reduced form ( $\text{Cyt}_{\text{red}}$ ) while only 2% of Fd is reduced when the  $\text{NADH}/\text{NAD}^+$  ratio is kept constant at 1:1. Changing

the ratios of NADH: NAD<sup>+</sup> by changing the recycling system (in living cells this would be equivalent to a change in metabolism) can significantly change the equilibrium concentrations.



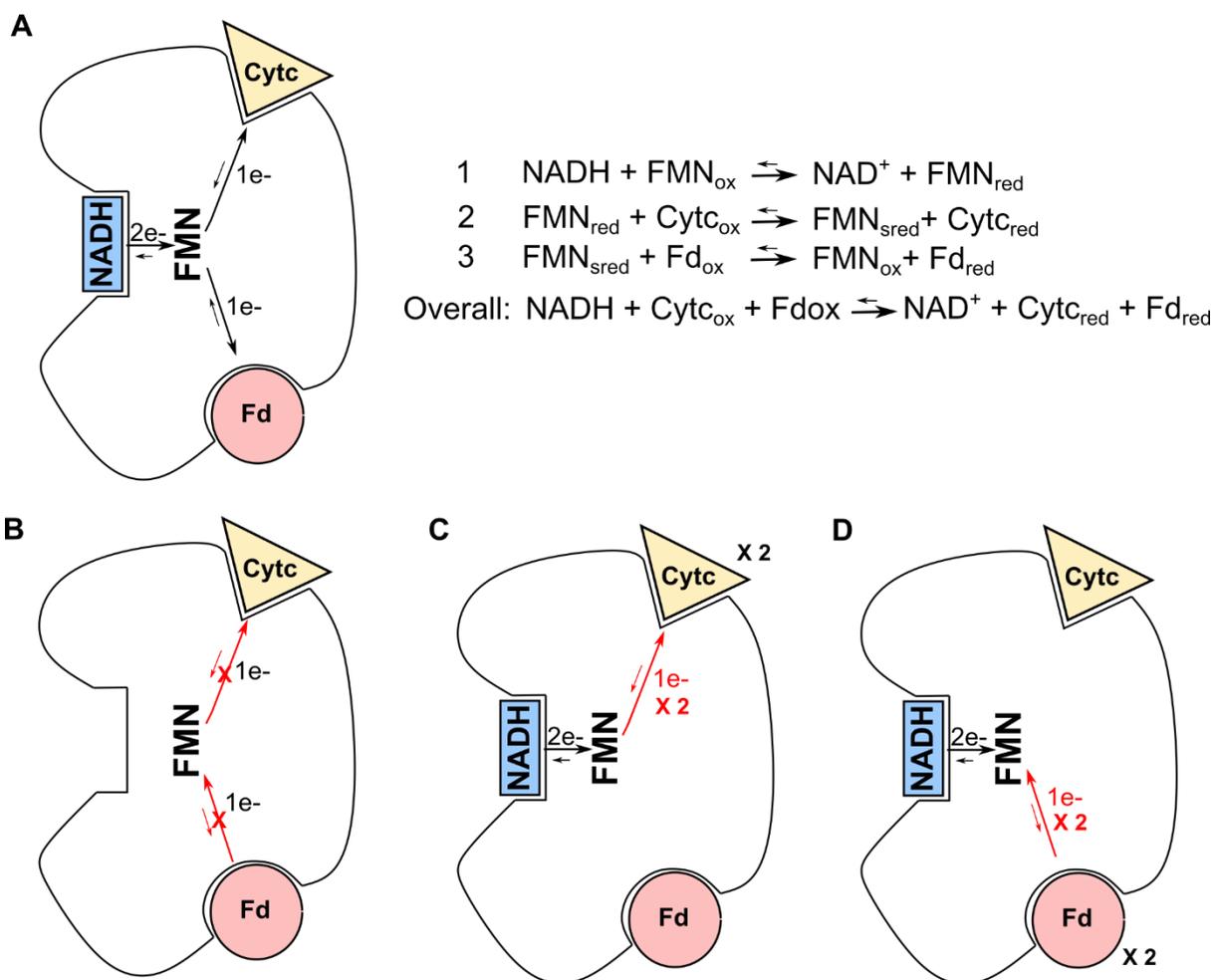
**Figure 3.12** Percentage of relative concentrations of Cytc<sub>red</sub> and Fd<sub>red</sub> compared to their respective total amount in solution (conversion %). Initially, only Cytc<sub>ox</sub> and Fd<sub>ox</sub> are added to the solution (0% conversion), after a certain amount of time (dependent on kinetic) the reaction reaches equilibrium. Here the NADH:NAD<sup>+</sup> ratio is kept constant at 1:1. Note: the shape of the kinetic plot is qualitative as thermodynamic analysis only provides information on the equilibrium concentrations.

### Bifurcating version

#### *Premises*

A single type of bifurcating enzyme is present in the solution (Fig.3.13.A). This bifurcating enzyme contains an FMN centre that receives two electrons concertededly from NADH becoming fully reduced (FMN<sub>red</sub>). FMN<sub>red</sub> then donates its first electron (“harder to donate”) to a cytochrome c molecule (Cyt<sub>c</sub>), which is the highest potential substrate; the now semi-reduced bifurcating redox cofactor donates its second electron (“easier to donate”) to ferredoxin (Fd) which is the lowest potential substrate (note: the donation of the first electron to Fd and the second one to Cyt<sub>c</sub> does not change the equilibrium concentrations, however, is kinetically significantly slower). The reactions are reversible (confurcation direction, shorter arrows in Fig.3.13.A), and it is also assumed that any non-catalyzed direct redox reaction between NAD, ferredoxin, and cytochrome c is too slow and therefore negligible. Catalyzed electron transfer between Fd and Cyt<sub>c</sub> is not possible as the bifurcating enzyme can physically block the ‘electrical connection’ by protein rearrangements (Fig.3.13.B). Protein rearrangements also ensure that the fully reduced FMN can only transfer one electron to a Cyt<sub>c</sub> molecule and one electron to an Fd molecule. The ability to transfer both electrons to or from two molecules of Cyt<sub>c</sub> (Fig.3.13.C) or two molecules of Fd (Fig.3.13.D) would result in the same non-bifurcating

equilibrium profile discussed above (Fig.3.12). In fact, in the bifurcating direction (tends to be with higher  $[NADH]/[NAD^+]$ ,  $[Cyt_{ox}]/[Cyt_{red}]$ , and  $[Fd_{ox}]/[Fd_{red}]$  ratios) the transferring also of the second electron to the higher potential substrate is favoured meaning that very little lower potential substrate will be reduced, and similarly, as soon as the reduced lower potential substrate concentration sufficiently increases electrons will backflow to FMN preventing further increase in concentration.



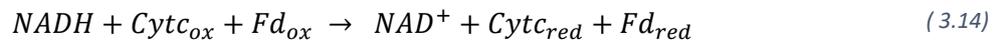
**Figure 3.13 A.** Hypothetical bifurcating enzyme. NADH transfer two electrons conertatedely to the bifurcation center FMN (cofactor). The fully reduced FMN then transfers sequentially one electron to oxidized Cyt (first electron) and oxidized Ferredoxin (second electron). The reaction is reversible (confurcation) as shown by the shorter arrows. **B.** Movements of portions of the enzyme prevent the transferring of electrons between the one-electron acceptors/donors Cyt and Fd. **C.** Movements and rearrangements of the enzyme prevent that both electrons from the fully reduced FMN are transferred to two molecules of Cyt (higher potential acceptor) or **D.** two molecules of Fd (lower potential acceptor) and the opposite direction (to a fully oxidized FMN).

As in the non-bifurcating version, the reactions are imagined to be conducted at 25°C, atmospheric pressure, and buffered pH = 7 with the ratio of NADH and NAD<sup>+</sup> in the solution

constant and equal to 1:1 (e.g. by using a regeneration system (222)). The same amount in moles of oxidized ferredoxin ( $Fd_{ox}$ ) and oxidized Cytc ( $Cytc_{ox}$ ) are added to the solution containing this hypothetical bifurcating enzyme (Fig.3.13). The standard redox potential of all these species at pH=7 ( $E'^0$ ) is listed in Table 3.4. Note that a similar analysis can be made also in the confurcation direction (not shown here) by using higher  $[NADH]/[NAD^+]$  ratios, and by adding reduced Fd ( $Fd_{red}$ ) and reduced Cytc ( $Cytc_{red}$ ).

### Analysis

As discussed already in the non-bifurcating version, in standard conditions and pH=7, the reduction of  $Fd_{ox}$  by NADH is endergonic ( $\Delta E'^0 = -100$  mV,  $\Delta G'^0 = +19.3$  kJ/mol) while the reduction of  $Cytc_{ox}$  by NADH is exergonic ( $\Delta E'^0 = 574$  mV,  $\Delta G'^0 = -110.8$  kJ/mol). However, in standard conditions, the overall reaction in which one Fd and one Cytc molecule are both reduced (Equation 3.15) is exergonic ( $\Delta E'^0 = -420+254+2\cdot 320$  mV = +474 mV,  $\Delta G'^0 = -91.5$  kJ/mol). Note that I say in standard conditions, actually for a portion of the reaction the reduction of  $Fd_{ox}$  by NADH is exergonic (Fig 3.14). Bifurcation reactions can also be achieved when based on standard potentials both reactions are endergonic as potentials change based on conditions (Nernst equation).



Similarly to the non-bifurcating version approach, I will calculate the conversion yields of Cytc and Fd at equilibrium ( $\Delta G = 0$ ) showing that coupling the reductions of Fd and Cytc enables the driving of the reduction of the lower potential substrate (Fd) even after reaching concentrations in which the reduction of  $Fd_{ox}$  by NADH becomes endergonic.

Based only on the catalyzed reactions (Fig.3.13) the following equalities must be satisfied at equilibrium. In red are the equalities that are not applicable as their corresponding reactions are blocked by the bifurcating enzyme (represented in Fig.13.B, C, D).

**Table 3.3** Equalities that must be satisfied at equilibrium (left) and equalities that are not (right) as they correspond to reactions blocked by the bifurcating enzyme.

Satisfied	Non-applicable (blocked by the enzyme)
$E_{NADH/NAD^+} = E_{FMN_{ox}/FMN_{red}}$	$E_{Cytc_{ox}/Cytc_{red}} = E_{Fd_{ox}/Fd_{red}}$
$E_{FMN_{sred}/FMN_{red}} = E_{Cytc_{ox}/Cytc_{red}}$	$E_{FMN_{sred}/FMN_{ox}} = E_{Cytc_{ox}/Cytc_{red}}$
$E_{FMN_{sred}/FMN_{ox}} = E_{Fd_{ox}/Fd_{red}}$	$E_{FMN_{sred}/FMN_{red}} = E_{Fd_{ox}/Fd_{red}}$

However,  $E_{FMN_{ox}/FMN_{red}} = \frac{1}{2}(E_{FMN_{sred}/FMN_{red}} + E_{FMN_{sred}/FMN_{ox}})$  and therefore by combining this with the above equalities ('satisfied') we obtain that:

$$E_{NADH/NAD^+} = \frac{1}{2}(E_{Cyt_{c_{ox}}/Cyt_{c_{red}}} + E_{Fd_{ox}/Fd_{red}}) \quad (3.15)$$

Equality 3.15 becomes identical to Equality 3.9 if only just one of the three non-applicable equalities (Table 3.3, red) becomes true. This means that the bifurcating enzyme must have tight control over the reactions otherwise thermodynamic coupling cannot be achieved and equilibrium concentrations are reached that are identical to the ones discussed in the 'non-bifurcating version'.

As already mentioned, donating the first electron to Fd and the second to Cytc would still allow coupling (Equation 3.15 doesn't change) as can be seen by simply mutually swapping  $E_{Cyt_{c_{ox}}/Cyt_{c_{red}}}$  and  $E_{Fd_{ox}/Fd_{red}}$  in the equalities of Table 3.3 (including red ones), however, this would result in a very slow reaction rate.

In addition, like in the non-bifurcating version, the reduction potentials of the cofactor FMN do not influence the concentrations at equilibrium however profoundly influence the kinetic; as a consequence, the potential inversion of the cofactor is just a kinetic and not a thermodynamical requirement for the coupling.

The first condition needed to calculate concentrations at equilibrium is given by Equality 3.15 which is substituted by the Nernst equation (Equation 3.8):

$$-0.320 - \frac{0.0592}{2} \log_{10}(1) = \frac{1}{2} \left( 0.254 - \frac{0.0592}{1} \log_{10} \frac{[Cyt_{c_{red}}]}{[Cyt_{c_{ox}}]} - 0.420 - \frac{0.0592}{1} \log_{10} \frac{[Fd_{red}]}{[Fd_{ox}]} \right) \quad (3.16)$$

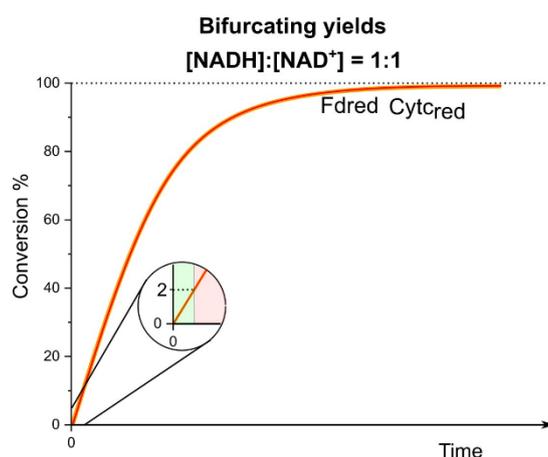
$$\text{Solution: } \frac{[Cyt_{c_{red}}]}{[Cyt_{c_{ox}}]} \cdot \frac{[Fd_{red}]}{[Fd_{ox}]} = 10^8$$

The second condition is given by the stoichiometry of the reaction and mechanism, which imposes that electrons are equally split between the two acceptors Cytc and Fd (if this wasn't the case, then at least one of the non-applicable equalities in Table 3.3 becomes true resulting in the non-bifurcating equilibrium). In this example, we start with equal concentrations of  $Cyt_{c_{ox}}$  and  $Fd_{ox}$ , and therefore at any point in the reaction, including at equilibrium,  $[Cyt_{c_{red}}]=[Fd_{red}]$  and  $[Cyt_{c_{ox}}]=[Fd_{ox}]$ . Combining this condition with Equation 3.16:

$$\frac{[Cyt_{c_{red}}]}{[Cyt_{c_{ox}}]} = \frac{[Fd_{red}]}{[Fd_{ox}]} = 10^4 \quad (3.17)$$

$$\frac{[Cyt_{red}]}{[Cyt]_{tot}} \cdot 100 = \frac{[Fd_{red}]}{[Fd]_{tot}} \cdot 100 = 99.99\%$$

The equilibrium concentrations of this bifurcating reaction are illustrated in Fig.3.14. The plot shows that at equilibrium almost all (99.99%) of Cyt<sub>c</sub> and Fd are in the reduced form when the NADH/NAD<sup>+</sup> ratio is kept constant at 1:1. Changing the ratios of NADH:NAD<sup>+</sup> by changing the recycling system can significantly change the equilibrium concentration. Compared to the non-bifurcating version, the thermodynamic coupling allows more electrons to be transferred to the lower potential acceptor (here Fd) and less to the higher potential acceptor (here Cyt<sub>c</sub>).



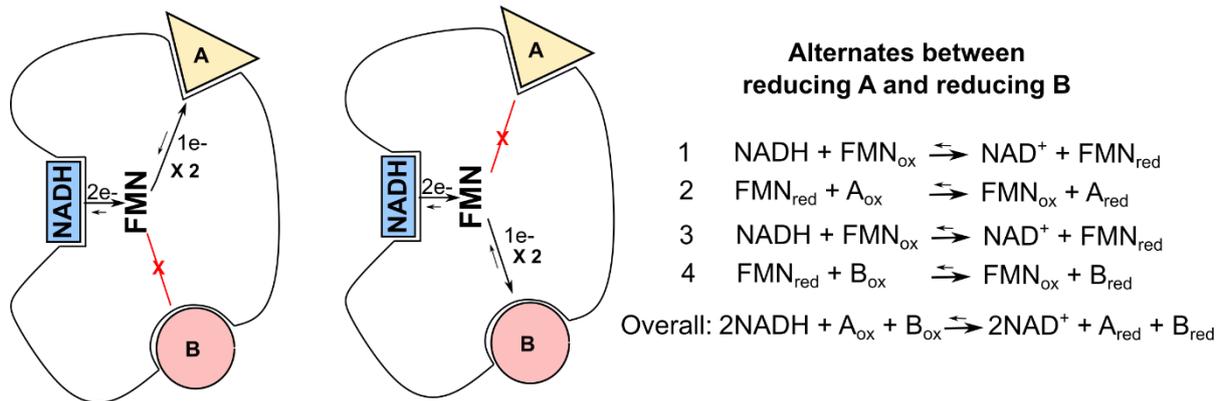
**Figure 3.14** Percentage of relative concentrations of Cyt<sub>c</sub><sub>red</sub> and Fd<sub>red</sub> compared to their respective total amount in solution (conversion %) vs. reaction time. Initially, only Cyt<sub>c</sub><sub>ox</sub> and Fd<sub>ox</sub> are present in solution (0% conversion) with the same concentration; after a certain amount of time (dependent on kinetic) the reaction reaches equilibrium. When the NADH:NAD<sup>+</sup> ratio is kept constant at 1:1 there is an almost complete conversion of Cyt<sub>c</sub><sub>ox</sub> and Fd<sub>ox</sub> into Cyt<sub>c</sub><sub>red</sub> and Fd<sub>red</sub> (99.99%). The reduction of Fd remains exergonic until a 2% conversion is achieved (green area). Note: the shape of the kinetic plot is qualitative as thermodynamic analysis only provides information on the equilibrium concentrations.

### Kinetic coupling without thermodynamic coupling (false bifurcation)

In the above bifurcating version, kinetic coupling is a necessary condition for thermodynamic coupling. However, I will demonstrate that kinetic coupling is not a sufficient condition for thermodynamic coupling and therefore for bifurcation.

For example, kinetic coupling can also be achieved by an enzyme that through protein rearrangements allows two electrons to be sequentially transferred from the FMN cofactor to substrate A, and after that, another two electrons are sequentially transferred to substrate B (note that in less simplified mechanisms A and/or B can be redox centres within the enzyme that then transfer both electrons to a substrate). However, this enzyme does not allow another

molecule of substrate A to bind and receive electrons unless substrate B has been reduced and released and vice versa (Fig.3.15). This is an example of a possible kinetic coupling mechanism that does not result in thermodynamic coupling, however, in certain conditions may appear as a false bifurcation (both substrates A and B are necessary). Upon close inspection, a false bifurcation mechanism has been proposed both for *TmHydABC* (171) and *HydABC* from the acetogenic bacteria *Acetobacterium woodii* and *Thermoanaerobacter kivui* (207). I will demonstrate this in Chapter 3.4.3.



**Figure 3.15** Multiple kinetic mechanisms can result in a false bifurcation where kinetic coupling does not produce thermodynamic coupling. All the possible false bifurcation mechanisms have in common that the two electrons received from the donor (in this case NADH) are first transferred both to the acceptor A and then another molecule of the donor (here NADH) binds and the two electrons are transferred to the acceptor B. The reaction can only proceed by reducing A and B in sequence: NADH binds  $\rightarrow$  A is reduced ( $2e^-$ )  $\rightarrow$  NADH binds  $\rightarrow$  B is reduced ( $2e^-$ )  $\rightarrow$  NADH binds  $\rightarrow$  A is reduced ( $2e^-$ ), etc. The reaction cannot proceed in the absence of either A or B (both are required).

Based only on the catalyzed reactions (Fig.3.15) the following equalities must be satisfied at equilibrium.

$$E_{\text{NADH}/\text{NAD}^+} = E_{\text{FMN}_{\text{ox}}/\text{FMN}_{\text{red}}}$$

$$E_{\text{FMN}_{\text{ox}}/\text{FMN}_{\text{red}}} = E_{\text{A}_{\text{ox}}/\text{A}_{\text{red}}}$$

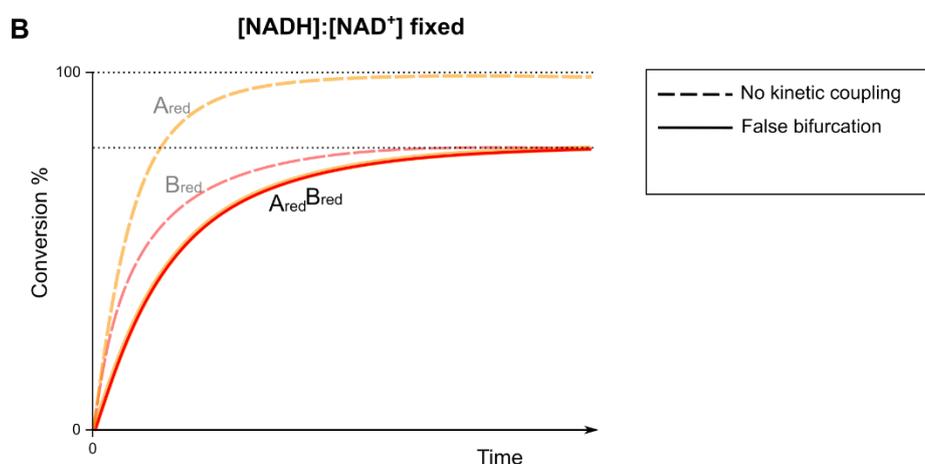
$$E_{\text{FMN}_{\text{ox}}/\text{FMN}_{\text{red}}} = E_{\text{B}_{\text{ox}}/\text{B}_{\text{red}}}$$

As a consequence, at equilibrium:

$$E_{\text{NADH}/\text{NAD}^+} = E_{\text{A}_{\text{ox}}/\text{A}_{\text{red}}} = E_{\text{B}_{\text{ox}}/\text{B}_{\text{red}}} \quad (3.18)$$

The above equality is analogous to the non-bifurcating version discussed previously. However, in this case, the specific type of kinetic coupling requires that both oxidized substrates A and B are present at the same time and are reduced alternately (Fig.3.15); as a result, the reaction can only proceed until both reductions remain exergonic and stops as soon as one reduction

becomes endergonic (Fig.3.16). The lower reduction potential substrate prevents the higher potential substrate from being further reduced and therefore true thermodynamic equilibrium is never achieved.



**Figure 3.16** Percentage of relative concentrations of the higher potential  $A_{red}$  and lower potential  $B_{red}$  compared to their respective total amount in solution (conversion %) vs. reaction time. Initially, only  $A_{ox}$  and  $B_{ox}$  are present in the solution (0% conversion) with the same concentration. In the reaction without kinetic coupling (dashed lines) both A and B reach their equilibrium concentration similarly to the non-bifurcating example discussed. In case kinetic coupling is present that does not result in thermodynamic coupling (‘false bifurcation’, continuous line) only  $B_{red}$  reaches its equilibrium concentration. Qualitative plot.

**Table 3.4** Standard redox potentials of the  $FMN_{ox}/FMN_{sred}$ ,  $FMN_{sred}/FMN_{red}$ , and  $FMN_{ox}/FMN_{red}$  couples in water at pH = 7 were determined by R. Anderson (219). Other redox couples reported are  $NAD^+/NAD^{\bullet}$ ,  $NAD^{\bullet}/NADH$ ,  $NAD^+/NADH$  (223) and  $Fd_{ox}/Fd_{red}$  (224). The reduction potentials of FMN species in Nqo1 (complex I) are reported at pH 7 and determined by Tomoko Ohnishi and co-workers (225). The reduction potential of the cytochrome was obtained from Lee Rodkey and Eric G. Ball (226).

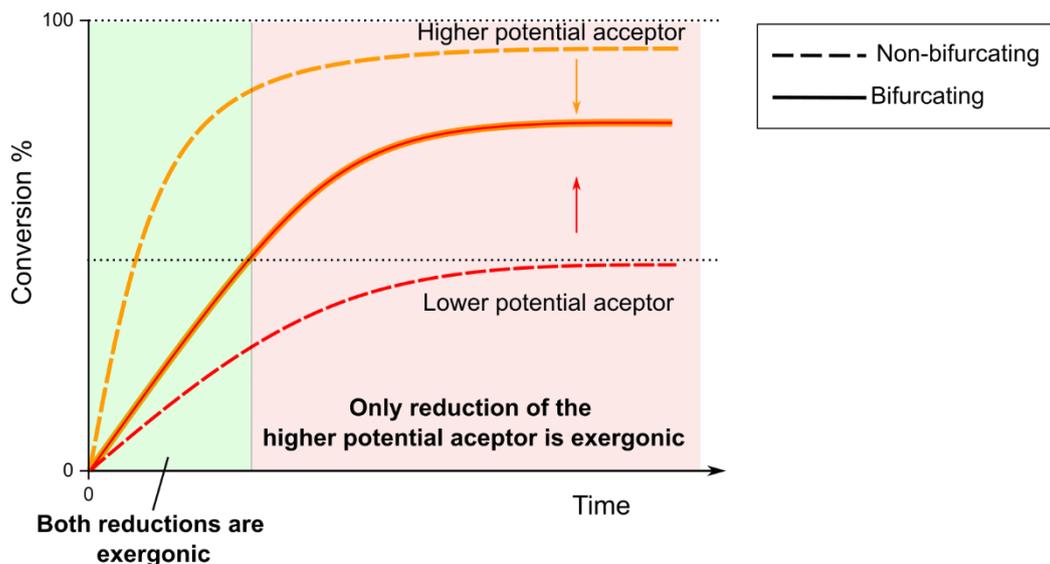
Redox couple	$E^{\circ}$ (mV) (solution, pH = 7)	$E^{\circ}$ (mV) pH = 7 Nqo1 (complex I)
$FMN_{ox} + 1e^{-} \rightarrow FMN_{sred}$	-314	-387
$FMN_{sred} + 1e^{-} \rightarrow FMN_{red}$	-124	-293
$FMN_{ox} + 2e^{-} \rightarrow FMN_{red}$	-219	-340
$NAD^{+} + 1e^{-} \rightarrow NAD^{\bullet}$	-922	
$NAD^{\bullet} + 1e^{-} \rightarrow NADH$	+282	
$NAD^{+} + 2e^{-} + H^{+} \rightarrow NADH$	-320	
$Fd_{ox} + 1e^{-} \rightarrow Fd_{red}$ (C.kluyveri ferredoxin)	-420	
$Cyt\ c\ (Fe^{3+}) + 1e^{-} \rightarrow Cyt\ c\ (Fe^{2+})$	+254	

## Conclusions of the thermodynamic analysis

Some conclusions and generalizations can be drawn from the presented example of a bifurcating and corresponding non-bifurcating reaction:

1. Bifurcation (and confurcation) is the thermodynamic coupling of two reactions with different electrochemical potentials. This coupling is preserved also while both corresponding non-bifurcating reactions are exergonic (Fig.3.14, green area).
2. Kinetic coupling is a condition necessary but not sufficient for thermodynamic coupling. Kinetic coupling requires dynamic protein rearrangements and movements.
3. The reduction potentials of the bifurcating cofactor (including the presence or not of inverted potential) only affect the kinetics unless the reaction becomes so slow to be comparable with the non-catalyzed redox reaction. Similarly, in the bifurcation (or confurcation) direction, donating (or accepting) the first electron from (to) the higher potential substrate and the second electron from (to) the lowest potential substrate or vice versa only affects the kinetics.

A qualitative plot of a general classical bifurcating reaction is presented in Fig.3.17 which shows how electron bifurcation affects the equilibrium concentrations of the higher and lower one-electron acceptor substrates.



**Figure 3.17** Qualitative plot showing the percentage of relative concentrations of reduced higher and reduced lower potential one-electron acceptors compared to their respective total amount in solution (conversion %) vs. reaction time. Initially, the one-electron acceptors are present in solution only in the oxidized form (0% conversion) with the same concentration; after a certain amount of time (dependent on kinetic) the reaction reaches equilibrium. In the plot, it is evident the kinetic coupling from the convergence of the two non-bifurcating graphs into one. The thermodynamic and kinetic coupling is maintained also while both reductions are exergonic (green area).

### 3.1.12 Electron bifurcation/confurcation in characterized enzymes

The first known example of electron bifurcation was that of the complex bc<sub>1</sub> of the respiratory chain. This enzyme uses ubiquinol as the two-electron donor, Rieske iron-sulfur protein as the higher reduction potential electron acceptor, and cytochrome b as the lower reduction potential electron acceptor (228). The quinol-based electron bifurcation (QBEB) was first proposed by Peter Mitchell in 1975 (229) to explain the mechanism of this enzyme.

In 2008, flavin-based electron bifurcation (FBEB) was discovered from the observation that in *C. kluyveri* the endergonic (based on standard redox potentials) reduction of ferredoxin with NADH is coupled to the exergonic reduction of crotonyl-CoA to butyryl-CoA by the cytoplasmatic Butyryl-CoA Dehydrogenase/Etf complex (EtfAB-Bcd) (230). Since then, a wide range of complexes containing Electron-Transferring Flavoproteins (**EtfAB**) was found to reversibly utilize electron bifurcation as a coupling mechanism (221).

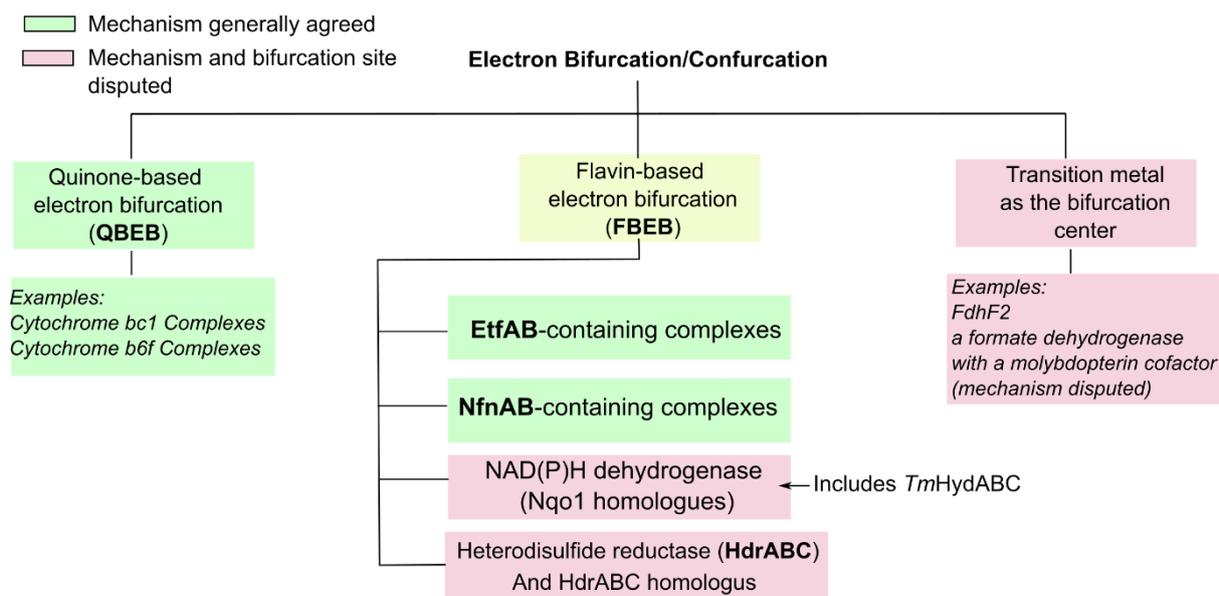
A few years later, in 2010, a confurcating NADH-dependent reduced ferredoxin:NADP oxidoreductase (**NfnAB**) in *C. kluyveri* was discovered (231). This enzyme catalyzes the reversible (therefore also bifurcation) reduction of 2 NADP by 1 NADH and 2 reduced ferredoxin utilizing two flavins as bifurcating centres. Later additional NfnAB bifurcating enzymes were found in other organisms (221).

Electron bifurcation in EtfAB- and NfnAB-containing complexes is the best characterized and understood mechanism among all flavin-containing bifurcating enzymes and fits the classical bifurcation/confurcation model (discussed in Chapter 3.1.9) although with some additional steps as required by the stoichiometry (Chapter 3.1.13). In these complexes, NADH or NADPH transfers two electrons concertedly to the flavin cofactor through hydride transfer.

Other classes of enzymes containing a flavin cofactor were proposed to be using an FBEB mechanism, implying that (at least one of) their flavin cofactor is the bifurcation centre which receives two electrons from a substrate other than NADH or NADPH (namely formate, H<sub>2</sub>, and F<sub>420</sub>H<sub>2</sub>). However, their mechanism is still debated including whether the flavin is their bifurcation center. A summary of all classes of enzymes utilizing or suspected to utilize FBEB is provided in Table 3.5.

In addition to quinones and flavins, transition metals with inverted potentials have been proposed to be bifurcating centres (232), however, this is still poorly understood.

An overview of these recognized types of electron bifurcation/confurcation mechanisms is provided in Fig.3.18.



**Figure 3.18** Types of electron bifurcation/confurcation mechanisms and enzymes utilizing them. Electron bifurcation/confurcation is mainly divided into three types depending on the bifurcation centre: QBEB, FBEB, and transition metal-based (disputed). Four unrelated families of enzymes are considered to utilize FBEB, however, only the mechanism of EtfAB- and NfnAB-complexes are generally agreed upon (green).

**Table 3.5** Electron-bifurcating flavoprotein complexes that have been characterized (221). The flavins in EtfAB- and NfnAB-containing complexes are generally agreed to be bifurcation centres while in the other enzymes, this is debated.

Confirmed or suspected enzyme complexes using FBEB					
Class	Enzyme complex	Two-electron donor	Higher potential electron acceptor	Lower potential electron acceptor	Bifurcating centre
<b>EtfAB-containing complexes</b>	Bcd-EtfAB	NADH	Crotonyl-CoA	Fd/fld	1 FAD
	CarC-CarED	NADH	Caffeyl-CoA	Fd	1 FAD
	LctD-LctCB	NADH	Lactate	Fd	1 FAD
	FixCX-FixBA	NADH	Ubiquinone	Fld	1 FAD
<b>NfnAB</b>	NfnAB	NADPH	NAD <sup>+</sup>	Fd	1 FAD
<b>NAD(P)H dehydrogenases (Nqo1 homologues)</b>	HydABC <i>(TmHydABC belongs here)</i>	H <sub>2</sub>	NAD <sup>+</sup>	Fd	FMN suspected
	HytABCDE-FdhA	H <sub>2</sub>	NADP <sup>+</sup>	Fd	FMN suspected
	HylABC-FdhF2	HCOO <sup>-</sup>	NAD <sup>+</sup>	Fd	FMN suspected
<b>HdrABC complexes and homologues</b>	MvhADG-HdrABC	H <sub>2</sub>	CoM-S-S-CoB	Fd	FAD suspected
	FdhAB-HdrABC	HCOO <sup>-</sup>	CoM-S-S-CoB	Fd	FAD suspected
	HdrA2B2C2	F <sub>420</sub> H <sub>2</sub>	CoM-S-S-CoB	Fd	FAD suspected
	MetFV-HdrABC-MvhD	NADH	Methylene-H <sub>4</sub> F	?	FAD suspected

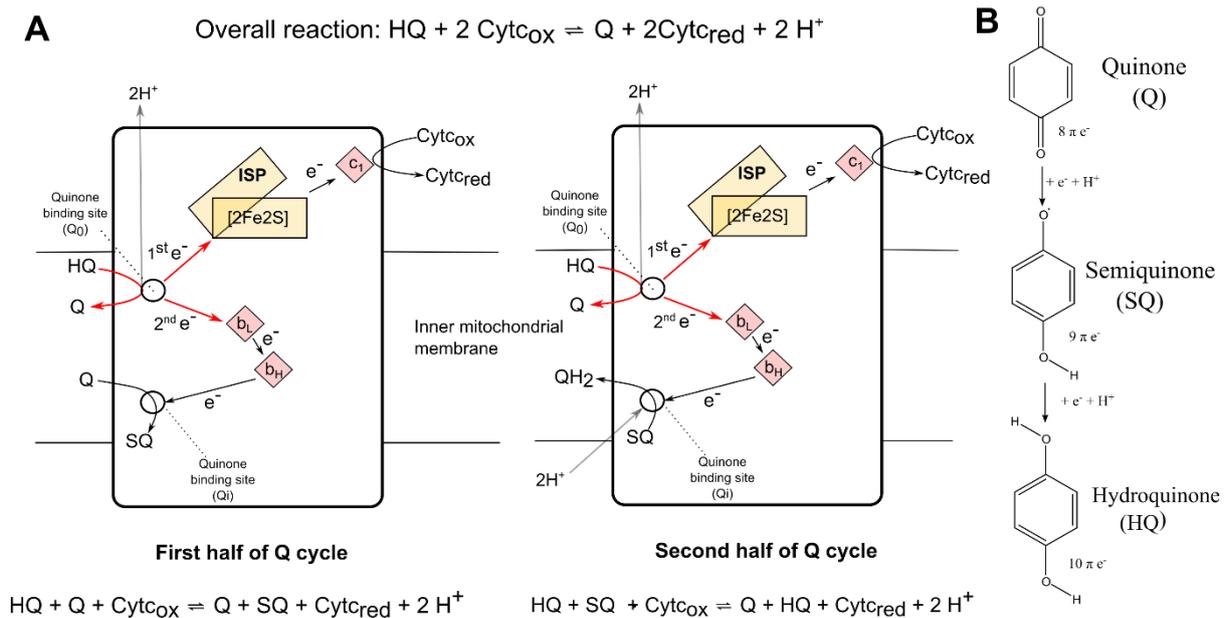
**Bcd-EtfAB:** Butyryl-CoA dehydrogenase-electron-transferring flavoprotein. **CarC-CarED:** Caffeyl-CoA reductase-EtfAB. **LctD-LctCB:** Lactate dehydrogenase-EtfAB. **FixCX-FixBA:** UQ reductase-EtfAB. **NfnAB:** Transhydrogenase. **HydABC:** [FeFe]-Hydrogenase. **HytABCDE-FdhA:** [FeFe]-Hydrogenase-CO<sub>2</sub> reductase. **HylABC-FdhF2:** Formate dehydrogenase. **MvhADG-HdrABC:** [NiFe]-hydrogenase-heterodisulfide reductase. **FdhAB-HdrABC:** Formate dehydrogenase-hetero-disulfide reductase. **HdrA2B2C2:** F<sub>420</sub>H<sub>2</sub>-dependent heterodisulfide reductase. **MetFV-HdrABC-MvhD:** Methylene-H<sub>4</sub>F reductase. **Fd:** ferredoxin. **Fld:** flavodoxin.

### 3.1.13 Examples of electron bifurcation in characterized enzymes

Characterized examples of electron bifurcation for which the classical model (Fig.3.9) successfully provides a mechanistic understanding is provided. An example for each of the three 'green' sub-classes in Fig.3.18 is presented.

#### *Cytochrome $bc_1$ complex bifurcation (QBEB)*

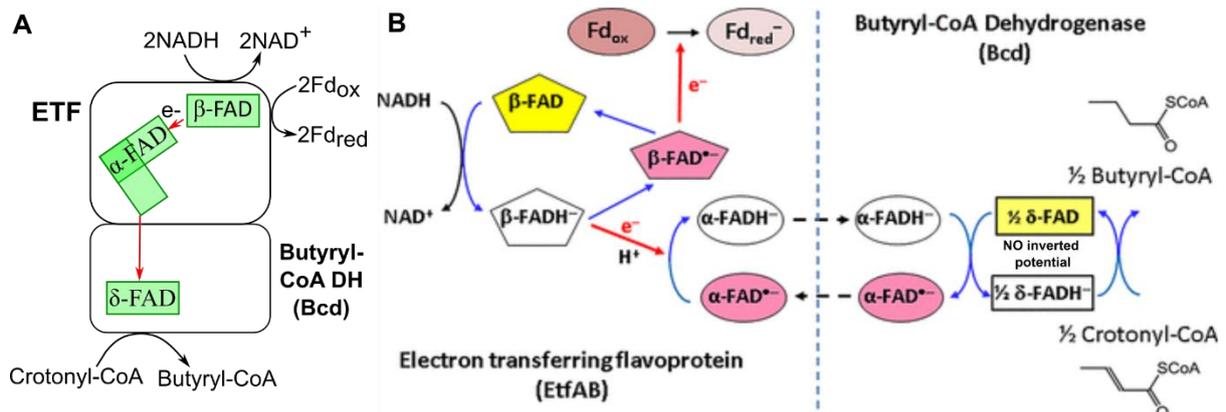
The first bifurcation pathway discovered was proposed in cytochrome  $bc_1$  complex (233). This enzyme transfers electrons from ubiquinol to cytochrome c and links this electron transfer to the establishment of a proton gradient across the inner mitochondrial or bacterial plasma membrane (234). Cytochrome  $bc_1$  utilizes a quinone-based electron bifurcation (QBEB) where the hydroquinone (HQ) bound at the outer Q binding site ( $Q_o$ ) is the starting point of bifurcation (Fig.3.19, red arrows). The higher potential one-electron acceptor is the Rieske iron-sulfur protein (ISP) and the lower potential one-electron acceptor is low potential cytochrome b (Cytochrome  $b_L$ ) (Fig.3.19) (221). During the bifurcation at the  $Q_o$  site (Fig.3.19, red arrows), hydroquinone ( $QH_2$ ) transfers one electron to the high-potential Rieske cluster ( $[2Fe2S]$ ) and becomes an intermediate semiquinone radical (SQ). The semiquinone species is considered highly unstable and reduces heme  $b_L$  very rapidly before it can react with dioxygen to generate superoxide. The bifurcated electrons are then transferred from the Rieske ISP and from cytochrome  $b_L$  to the final acceptors: the one-electron acceptor cytochrome c and an oxidized quinone molecule (Q) at the inner Q binding site ( $Q_i$ ) which becomes semi-reduced (first half of the Q cycle, Fig.3.19). The  $Q_i$  site preferentially stabilizes the semiquinone species (SQ) such that the couples HQ/SQ and SQ/Q have a reasonably close reduction potential (235). The SQ at the  $Q_i$  site is fully reduced to HQ after a second electron bifurcation event at the  $Q_o$  site (second half of the Q cycle, Fig.3.19). It is established that domain movement of the Rieske iron-sulfur protein (ISP) is fundamental to controlling electron flow.



**Figure 3.19.** **A** Mechanism of the Q-cycle in cytochrome  $bc_1$  complex. *First half of the cycle:* a hydroquinone molecule (HQ) binds at the  $Q_o$  site and transfers the first electron to a [2Fe-2S] cluster in the head of the Rieske iron-sulfur protein (ISP) which moves towards  $c_1$ . The second electron with less reducing power is donated to the lower potential cytochrome  $b_L$ . This split electron pair (red arrows) travels through the two different electron transfer pathways toward their final acceptors: oxidized cytochrome  $c$  ( $\text{Cyt}_{\text{ox}}$ ) and an oxidized quinone molecule (Q) at the  $Q_i$  site. *Second half of the cycle:* A second HQ molecule binds at the  $Q_o$  site and, similarly to the first half of the cycle, donates its first electron to the [2Fe-2S] cluster and second electron to cytochrome  $b_L$  (bifurcation, red arrows). The split electron pair travels through their distinct electron transport chains ending up reducing a second cytochrome  $c$  molecule and the SQ at the  $Q_i$  site to HQ. **B.** Conversion between quinone (Q), semiquinone (SQ), and hydroquinone (HQ).

### *Butyryl-CoA Dehydrogenase/Etf complex (EtfAB-Bcd) (FBEB)*

Butyryl-CoA Dehydrogenase/Etf complex (**EtfAB-Bcd**) was shown to couple the oxidation of NADH ( $E'^0 = -320 \text{ mV}$ ) to the simultaneous reduction of the lower potential acceptor Fd ( $E'^0 = -500 \text{ mV}$ ) and the higher potential acceptor of crotonyl-CoA ( $E'^0 = -10 \text{ mV}$ ) during acetate, ethanol (*C. kluyveri*) or glutamate (*A. fermentans*) fermentation (213). A scheme of Etf/Bcd mechanism and electron flow is shown in Fig 3.20.A. This shows that the enzyme operates under flavin-based electron bifurcation (FBEB) where the bifurcating centre is the  $\beta$ -FAD molecule.



**Figure 3.20** **A.** Proposed bifurcating electron transfer mechanism for Etf/Bcd complex in *Acidaminococcus fermentans*, in which the  $\beta$ -FAD in domain II of the EFT accepts electrons from NADH and bifurcates them to Fd and to the  $\alpha$ -FAD in the mobile domain.  $\alpha$ -FAD donates electrons to Bcd, which reduces crotonyl-CoA to butyryl-CoA. The summarized mechanism scheme is based on Fig.10 from (236). **B.** Scheme of electron bifurcation in the EtfAB-Bcd complex. NADH reduces  $\beta$ -FAD to  $\beta$ -FADH<sup>-</sup>, which bifurcates. One electron goes to  $\alpha$ -FAD<sup>-</sup> and the formed semi-reduced  $\beta$ -FAD<sup>•-</sup>, reduces ferredoxin (Fd). The semi-reduced  $\alpha$ -FADH<sup>-</sup> swings over to Bcd and transfers one electron to  $\delta$ -FAD. Repetition of this process yields to the fully reduced  $\delta$ -FADH<sup>-</sup>, which reduces crotonyl-CoA to butyryl-CoA. (218). Image taken from (221).

In Fig.3.20.B the electron flow of this Etf/Bcd system is shown in more detail; The enzyme uses classical bifurcation in which the  $\beta$ -FAD cofactor has an inverted potential and receives concerted two electrons (hydride transfer) from NADH becoming fully reduced ( $\beta$ -FADH<sup>-</sup>).  $\beta$ -FADH<sup>-</sup> then donates the first electron (“harder to donate”) to the higher potential acceptor which is the stabilized semi reduced  $\alpha$ -FAD<sup>•-</sup> cofactor; the now semi-reduced bifurcating redox cofactor ( $\beta$ -FAD<sup>•-</sup>) donates its second electron (“easier to donate”) to the lower potential acceptor which is the [4Fe-4S] cluster of ferredoxin.

After domain rearrangement, one electron flows from  $\alpha$ -FADH<sup>-</sup> to  $\delta$ -FAD in the Bcd subunit generating  $\delta$ -FADH<sup>-</sup>. Repetition of this process affords a second reduced ferredoxin and the returned  $\alpha$ -FAD<sup>•-</sup> becomes reduced to  $\alpha$ -FADH<sup>-</sup>, which donates one electron to  $\delta$ -FADH<sup>-</sup>. Finally, the generated fully reduced  $\delta$ -FADH<sup>-</sup> transfers a hydride to crotonyl-CoA yielding butyryl-CoA (237). These additional steps to the classical bifurcation are required as crotonyl-CoA is a two-electron acceptor requiring concerted electron transfer.

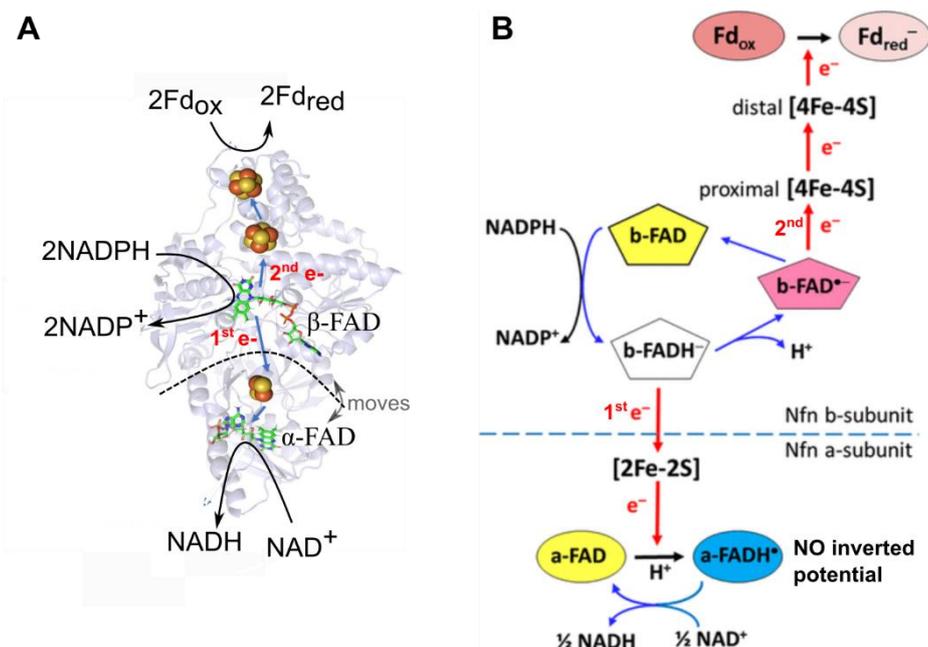
### ***NADH-dependent reduced ferredoxin:NADP oxidoreductase (NfnAB)***

The NADH-dependent reduced ferredoxin:NADP oxidoreductase (**NfnAB**) catalyzes the reversible reduction of 2 NADP by 1 NADH and 2 reduced ferredoxin utilizing two flavins as bifurcating centres:



In most organisms, NAD occurs mainly in the oxidized form, whereas in the case of NADP, the reduced form predominates; in aerobically glucose-fed exponentially growing *E. coli* cells the NAD<sup>+</sup>:NADH ratio is 31, and the NADP<sup>+</sup>:NADPH 0.018 (238). Hence, the effective reduction potential (Formula 3.8) of NAD inside the cell increases to  $E' = -280$  mV and that of NADP decreases to  $E' = -370$  mV from the standard  $E'^0 = -320$  mV (221). Therefore an additional donor with sufficiently low reduction potential (in this case Fd<sub>red</sub>) is necessary as a driver to reduce a significant amount of NADP<sup>+</sup> (confurcation direction).

The enzyme also operates in the bifurcation direction, where NADPH transfers one electron to the higher potential acceptor NAD<sup>+</sup> and the lower potential acceptor Fd<sub>ox</sub>. However, as NADH can only donate or receive (confurcation/bifurcation) concertedly two electrons, a more complex mechanism involving two flavins is present (Fig.3.21) similar to that of the one discussed in the example above.



**Figure 3.21** **A.** Overall structure and electron transfer pathway of *P. furiosus* NfnI (PDB: 5JCA). The blue arrows show the direction of electron transfer, and the black arrows indicate the reactions. The electron transfer pathway is presented in the direction of electron bifurcation although the physiological function of Nfn is found to reduce NADP<sup>+</sup>. Image taken from (239). **B.** Mechanism of FAD-based electron bifurcation proposed for the NfnAB complex. NADPH reduces β-FAD to β-FADH<sup>-</sup>, which bifurcates. One electron goes to α-FAD and the resulting semi-reduced β-FAD<sup>\*</sup> then transfers the second electron to ferredoxin (Fd) through an electron transport chain. Repetition of this process yields to the fully reduced α-FADH<sup>-</sup>, which reduces NAD<sup>+</sup> to NADH through concerted two-electron transfer. Movements of the NfnA subunit allow the control of the electron flow by increasing the distance of the [2Fe-2S] cluster. This process is represented in the bifurcation direction. Image taken from (221).

A more detailed electron flow scheme (bifurcation direction) is shown in Fig.3.21.B; the NfnAB enzyme uses classical bifurcation in which the β-FAD cofactor has an inverted potential and receives concertedly two electrons (hydride transfer) from NADPH becoming fully reduced (β-FADH<sup>-</sup>). β-FADH<sup>-</sup> then donates the first electron (“harder to donate”) to the nearest higher potential acceptor which is a [2Fe-2S] cofactor in the NfnA subunit which quickly transfers the electron to the α-FAD cofactor (which becomes semi-reduced, α-FADH<sup>-</sup>); β-FADH<sup>-</sup> then donates its second electron (“easier to donate”) to the lower potential acceptor which is the proximal [4Fe-4S] cluster. This second electron travels through the electron transport chain and reduces a bound Fd<sub>ox</sub>. The process is repeated after the binding of a second NADPH molecule which results in the reduction of a second Fd<sub>ox</sub> substrate and the reduction of α-FADH<sup>-</sup> which becomes fully reduced (α-FADH<sup>-</sup>). α-FADH<sup>-</sup> then transfer concertedly two electrons to NAD<sup>+</sup>.

Detailed steps that show how the movement of the NfnA subunit controls electron flow can be found in Ulrich Ermler and co-workers' paper (240).

### **3.1.14 *T. maritima* hydrogenase does not fit the classical FBEB model**

*T. maritima* hydrogenase (*TmHydABC*) was previously shown to contain flavin cofactors and therefore it was proposed that the enzyme utilizes a flavin-based bifurcation mechanism somewhat similar to the characterized ones in EtfAB- and NfNAB-containing complexes. However, so far, all the attempts to force a classical bifurcation in which the fully reduced flavin acts as a bifurcating centre (as described in the above examples and Chapter 3.1.9) were in disagreement with the here obtained 3D structure. In particular, the lack of a second flavin (or analogous two electron donor) as observed in EtfAB- and NfNAB-containing complexes makes the mechanistic modelling more challenging as no previous examples can be used as a guide.

Recently, three additional structures of bifurcating enzymes homologous to *TmHydABC* were published: HydABC from the acetogenic bacteria *Acetobacterium woodii* and *Thermoanaerobacter kivui* (207), and the NiFe-HydABC<sub>SL</sub> hydrogenase from *A. mobile* (211). In the first two cases, the authors attempted to model bifurcation using a questionable thermodynamic approach that I will show violates thermodynamic coupling. The third case (NiFe-HydABC<sub>SL</sub>) is presented generally and the mechanism remains vague. Also, the published mechanism that my co-workers proposed for *TmHydABC* (171) is not convincing.

I will here propose that *TmHydABC*, and all or most of the NAD(P)H dehydrogenases that have homology with the Nqo1 subunit of complex I may, in fact, not bifurcate, instead they may perform exclusively kinetic coupling (they split an electron pair but send both electrons downhill). This is a statement in stark contrast with more than a decade of published research (including ours), however, I will use a rigorous approach to support this hypothesis.

## 3.2 Aims of the chapter

In this result Chapter, the obtained Cryo-EM structure of *T. maritima* hydrogenase (*TmHydABC*) will be presented. The published paper and structures are available (171), however, here additional data and modelling results are presented.

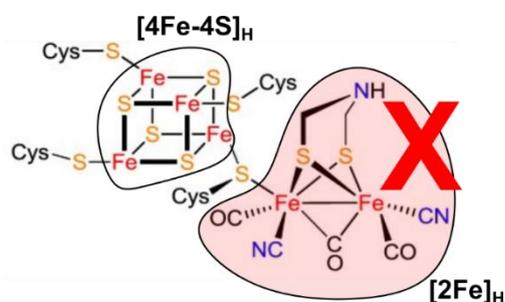
The chapter aims are:

- 1) To present the obtained structure of *TmHydABC* and its cofactor arrangement.
- 2) To present an unpublished model of *TmHydABC* with bound Fd.
- 3) To demonstrate that *TmHydABC* and other putatively bifurcating HydABC hydrogenases probably only perform kinetic but not thermodynamic coupling.
- 4) To model the kinetic coupling mechanism of HydABC enzymes.

## 3.3 Results

### 3.3.1 Screening the holoenzyme and apoenzyme preparation of *T. maritima* hydrogenase

Grids were prepared with the holo and apo *TmHydABC* enzyme minimizing oxygen exposure by operating as quickly as possible during grid preparation (Chapter 2.1.3). In the holoenzyme, the H-cluster was fully reconstructed as *E. coli* (the expression organism) could not synthesize it (150). In the apoenzyme, the H-cluster was missing the  $[2\text{Fe}]_{\text{H}}$  subcluster of the H-cluster (Fig.3.22).

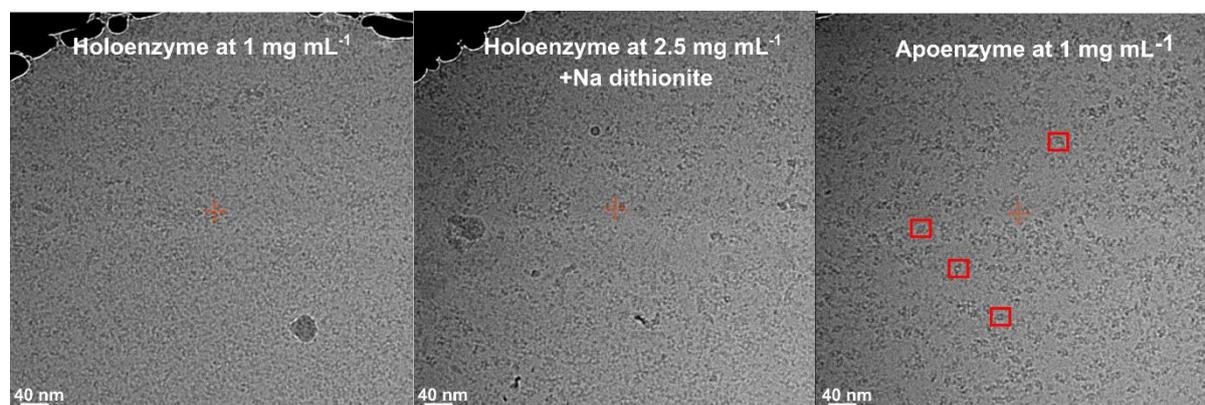


**Figure 3.22** Structure of the active site of  $[\text{FeFe}]$ -hydrogenases also called H-cluster. The  $[2\text{Fe}]_{\text{H}}$  subcluster was not present in the preparation which resulted in the published structure. Adapted from (184).

The grids were observed under the electron microscope and it was concluded that only the apoenzyme did not significantly degrade (Fig.3.23). As the holoenzyme was shown to be active by our collaborators, it was concluded that the presence of the  $[2\text{Fe}]_{\text{H}}$  subcluster significantly increases oxygen sensitivity to a level at which preparing grids in ambient air is severely detrimental to the sample even when a reducing agent (sodium dithionite) is present. The high oxygen sensitivity of the H-cluster in  $[\text{Fe-Fe}]$ -hydrogenases is well known and in particular the role of  $[2\text{Fe}]_{\text{H}}$  subcluster in initiating oxygen degradation has already been demonstrated (241).

It is interesting that the inclusion of the  $[2\text{Fe}]_{\text{H}}$  subcluster could disrupt the complex so thoroughly. It is now becoming clear that not only during grid preparation the sample is exposed to a large area of air-water interface but the air-water interface also generates  $\text{H}_2\text{O}_2$  (242). Even small quantities of Fe released (i.e. by oxygen disruption of the active site) can convert  $\text{H}_2\text{O}_2$  to various aggressive radical species through well-established Fenton chemistry (243).

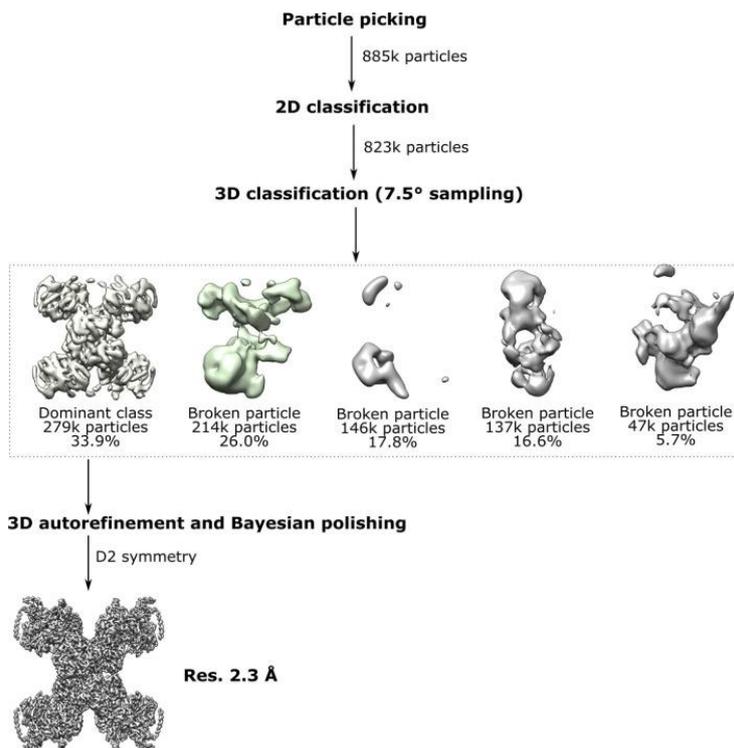
As our lab did not at that time have anaerobic gloveboxes that could fit a Vitrobot device for anaerobic grid preparation, it was decided that it was appropriate to collect data only on the apoenzyme. In support of our choice, previous studies have indicated that the incorporation of the [2Fe]<sub>H</sub> subcluster minimally affects the structure of [FeFe]-hydrogenases (154) (except for the enzyme from *Chlamydomonas reinhardtii* (244)).



**Fig.3.23** Images of grid holes containing holoenzyme (left), holoenzyme with sodium dithionite (centre), and apoenzyme (right). Only the apoenzyme preparation contained visibly intact particles (the red squares show some of them). Adding sodium dithionite to the holoenzyme solution before preparing grids only resulted in minimal improvement (very few intact particles) compared to the preparation without.

### 3.3.2 Data collection and data processing pipeline: apoenzyme

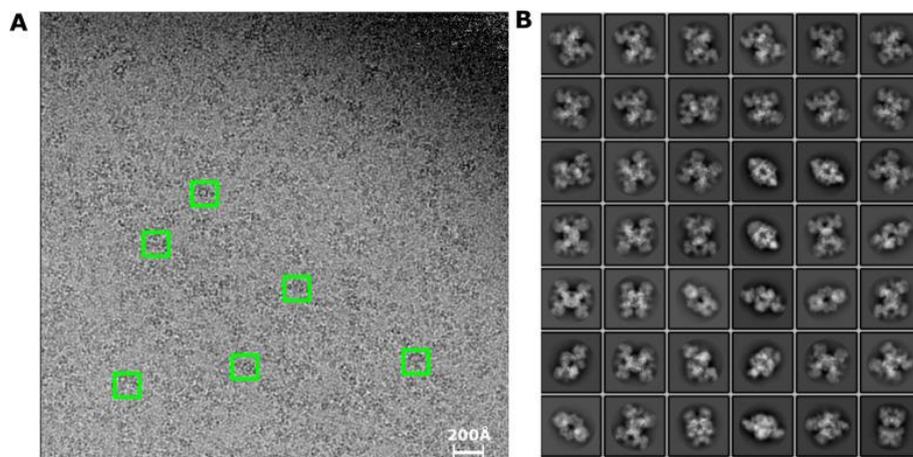
Data collection on an apoenzyme cryo-EM preparation using gold grids (see materials and methods) was attempted. Data processing involved multiple steps as described in the method section and more generally in Chapter 1.8.9. Particle picking from 4,790 movies, resulted in 885,000 particles. 2D and 3D classification was used to remove all the particles that corresponded to damaged proteins or noise/contaminants. The resulting 279,000 ‘good particles’ corresponding to views of intact *TmHydABC* were used to build a 3D model initially without symmetry imposed. D2 symmetry was then noted which was used to increase the resolution of the 3D map (effectively it’s like having 4 times the number of particles). The final 3D map with D2 symmetry imposed reached a resolution of 2.3 Å. The data processing pipeline used to obtain the published D2 symmetry map (EM map: EMD-13199) is summarized in Fig. 3.24.



**Figure 3.24** Classification and refinement of the cryo-EM density map for *TmHydABC* using the RELION pipeline. Following automatic particle picking and 2D and 3D classification to discard broken particles and contaminants, 3D refinement, and particle polishing were performed. Subsequently, the particles were classified using an angular sampling of 7.5° and pixel size of 3.4 Å, limiting the data to 6.8 Å (see Materials and methods). All five classes provided were populated. The dominant class refined to 2.3 Å. The remaining classes contained what appeared to be damaged particles or fragments of broken particles.

### 3.3.3 2D classification results

It can be seen from the 2D classes (Fig.3.25) that there is a good particle distribution with multiple views of *TmHydABC* contributing to the 3D reconstruction.

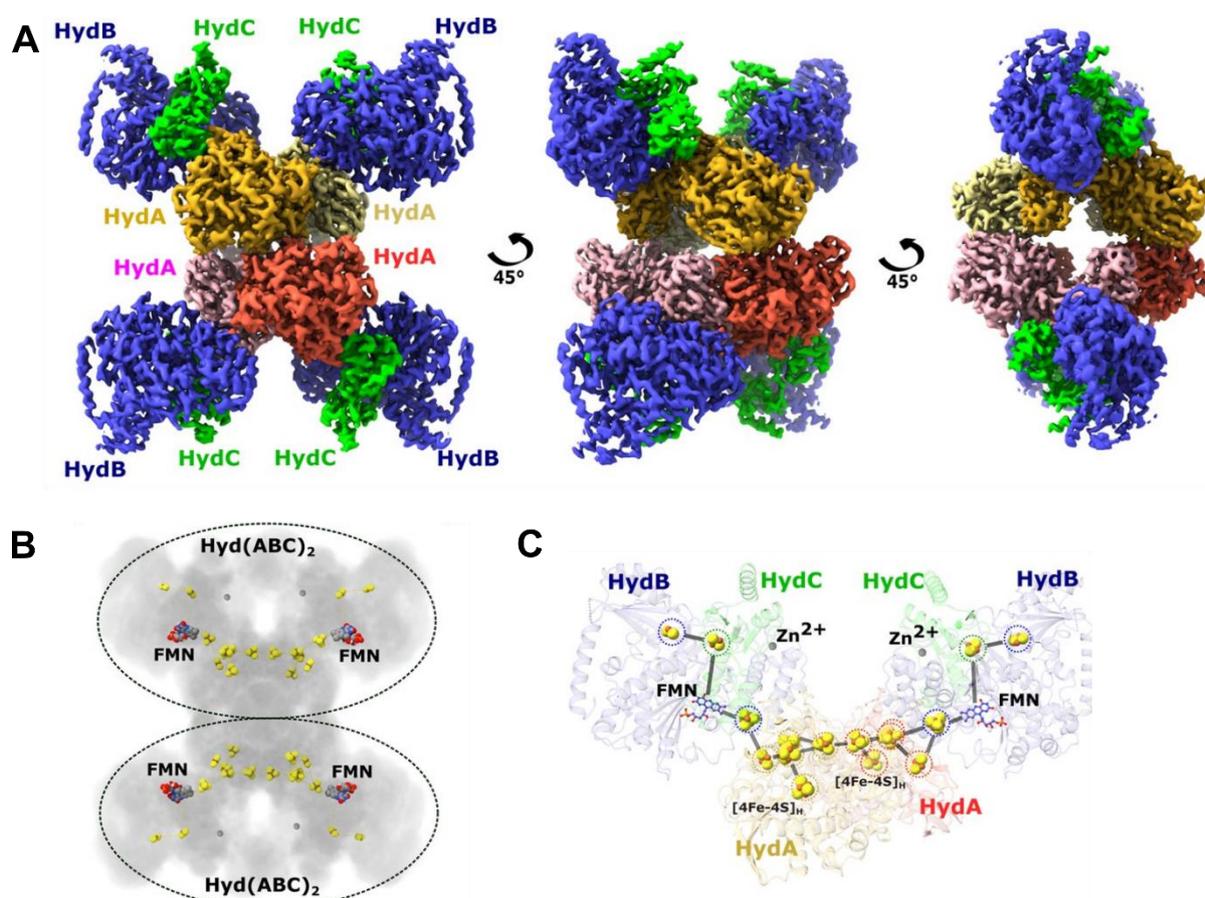


**Figure 3.25** Micrograph and 2D classes averages of *T. maritima* HydABC. **A.** Representative micrograph from the data collection used to resolve the structure of *TmHydABC*. A few different views of well-isolated particles are shown, highlighted with a green box. **B.** 2D class averages after cleaning the dataset with coarse 3D classification, only the full tetramer of trimers is observed.

### 3.3.4 3D map (D2 symmetry) and atomic model

We obtained a 2.3 Å electron cryo-microscopy map of apo-*T. maritima* hydrogenase (*TmHydABC*) by applying D2 symmetry (Fig.3.26.A). This map was used to build the atomic model (PDBID: 7P5H) as described in the methods section.

The structure shows a hetero-dodecamer composed of two independent ‘halves’ (Fig.3.26.B) each made of two strongly interacting HydABC heterotrimers connected via two His-ligated [4Fe–4S] clusters (Fig.3.26.C).

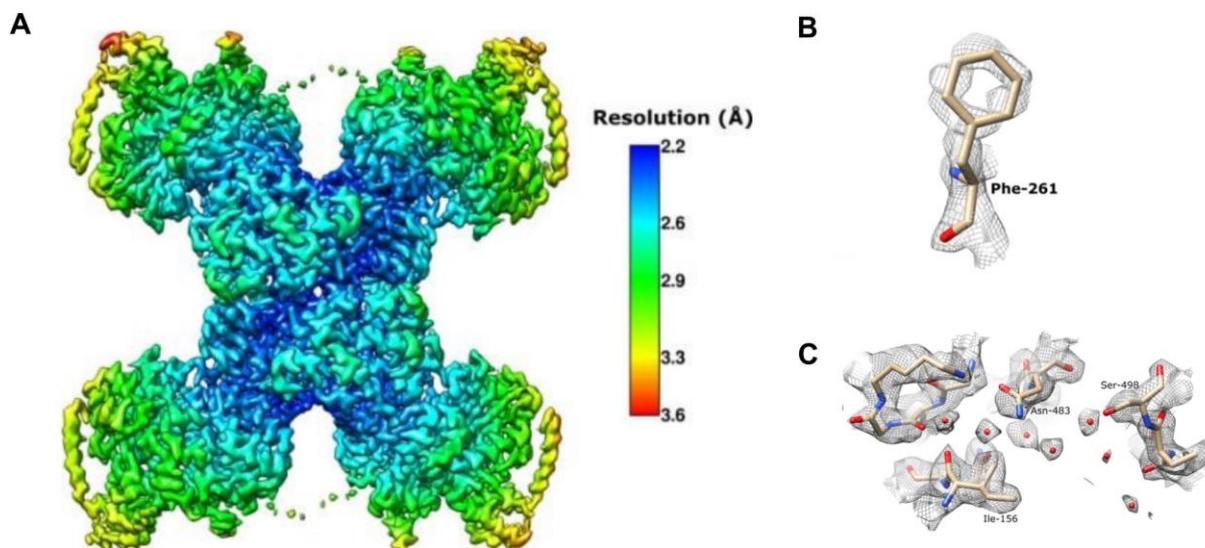


**Figure 3.26** Cryo-EM structure of the HydABC tetramer and arrangement of the redox cofactors. **A.** The unsharpened 2.3 Å map of Hyd(ABC)<sub>4</sub> with D2 symmetry enforced showing a tetramer of HydABC heterotrimers. All four copies of HydB and C are coloured blue and green, respectively. The four HydA copies that make up the core of the complex are in orange, yellow, pink, and red. The top and bottom halves of the complex are constituted by dimers of HydABC protomers (each HydABC unit is a protomer); the two protomers within the same dimer are strongly interacting, while a weaker interaction is present between the top and bottom dimers. **B.** The arrangement of redox cofactors within the protein complex, showing two independent identical redox networks (dashed circles); each redox network is composed of iron-sulfur clusters belonging to a Hyd(ABC)<sub>2</sub> unit composed of two strongly interacting HydABC protomers. **C.** HydABC dimer highlighting the iron-sulfur clusters and flavin mononucleotide (FMN) constituting the electron transfer network.

### 3.3.5 Local resolution of the 3D map

The obtained 3D map with D2 symmetry had an average resolution of 2.3 Å, however locally this varies substantially and decreases in more flexible regions of the protein (Fig.3.27.A). In the core part which displays higher resolution, water molecules (Fig.3.27.C) and the aromatic

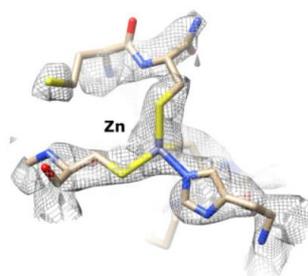
rings of residues were resolved (Fig.3.27.B). These features are unique to high-resolution cryo-EM structures.



**Figure 3.27** Cryo-EM structure of the HydABC tetramer and arrangement of the redox cofactors. **A.** Local resolution of the D2 map. Local resolution was estimated using the local resolution function in RELION with default parameters. The figure was created using UCSF Chimera. **B.** Density maps and model of Phenylalanine-261 in HydA (threshold 0.078) showing the resolved aromatic ring. **C.** Example of waters in HydA (threshold 0.038)

### 3.3.6 The controversy over the zinc site

The atomic model that was built in the D2 map (and later maps) showed a tetrahedrally coordinated site in HydB which from initial geometry observations appeared to be consistent with a zinc site (Fig.3.28).



**Figure 3.28.** Density maps and model Zn<sup>2+</sup> site (threshold 0.040) in HydB showing the tetrahedral coordination.

However, due to the lower local resolution in the region, we could not immediately exclude a [2Fe2S] cluster. Fitting of a [2Fe2S] cluster was attempted as well, however, this led to a poorer

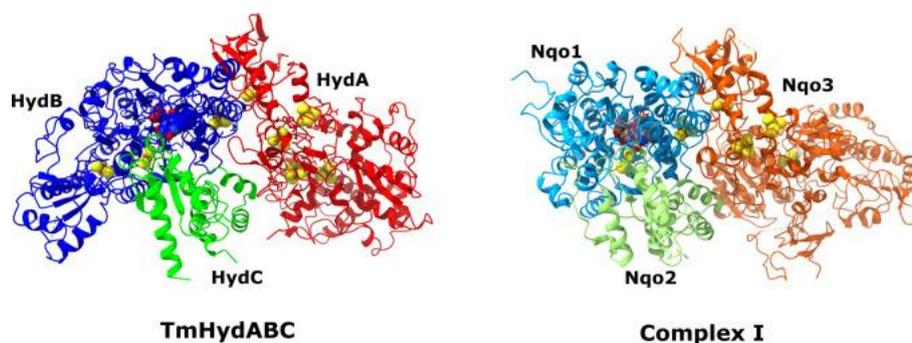
match of the density. Inductively coupled plasma mass spectrometry was then attempted on the separately produced and purified HydB subunit resulting in  $0.99 \pm 0.43$  Zn/protein and  $\approx 14.2 \pm 1.5$  Fe/protein. As the observed Fe content matches with the estimated Fe content of HydB, it was concluded that the additional site was indeed a zinc site. In addition, attempting to coordinate a [2Fe-2S] cluster resulted in a strained geometry.

In a related bifurcating [NiFe]-hydrogenase (HydABCSL) from *A. mobile* (PDBID: 7T2R (211)) it was proposed that instead of a zinc site, *AmHydB* contains an additional [2Fe-2S] cluster, however, the local resolution of the site in the HydABCSL map makes impossible to exclude one or the other (Fig.3.39.E).

### 3.3.7 Structural comparison of HydABC with homologous proteins

#### *Complex I*

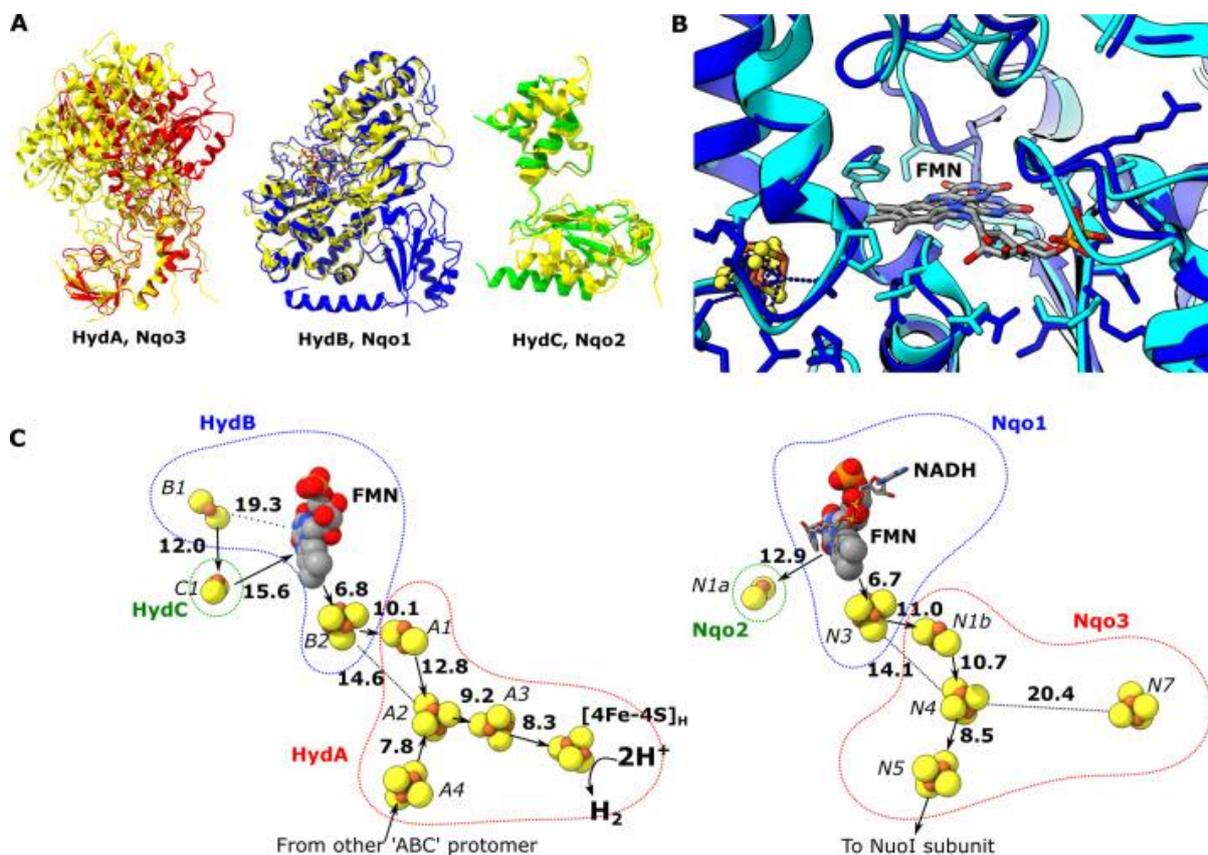
The spatial arrangement of subunits HydA, B, and C in the HydABC protomer is similar to that of subunits Nqo3, Nqo1, and Nqo2, respectively, in the NADH oxidation (N) module of *Thermus thermophilus* (Tt) respiratory complex I (Fig.3.29).



**Figure 3.29** HydABC protomer next to Nqo3, Nqo1, and Nqo2 subunits of complex I from *Thermus Thermophilus* (PDB ID: 6I1P) in their native arrangements [Gutiérrez-Fernández et al., 2020].

The individual subunits are also structurally highly similar and here we use RMSD (root-mean-square deviation between the  $C\alpha$  positions in homologous pairs of amino acids) as a quantitative measure of similarity between proteins. The highest similarity is between HydB and Nqo1 (RMSD of 1.040 Å) (245), followed by HydC and Nqo2 (RMSD 1.152 Å), and the lowest similarity between HydA and Nqo3 [RMSD 1.294 Å] (Fig.3.30). The remarkable structural similarities between HydB and Nqo1 subunits agree with their common evolutionary

origins (212) and suggest that NADH oxidation follows a similar mechanism in both enzymes (Fig.3.30.B).

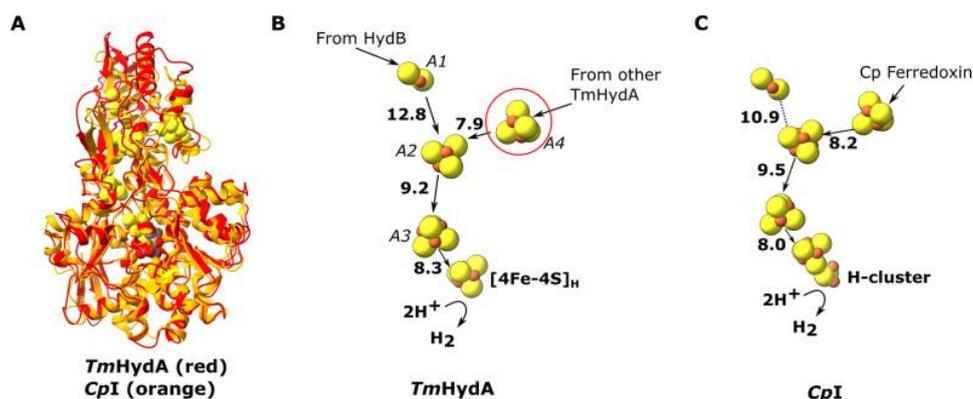


**Figure 3.30** Comparison of the HydA, B and C subunits of the electron bifurcating [FeFe] hydrogenase from *Thermotoga maritima* with the Nqo3, 1, and 2 subunits from respiratory complex I from *Thermus thermophilus*. **A.** Subunits HydA (red), HydB (blues), and HydC (green) overlaid with, respectively, Nqo3, Nqo1, and Nqo2 (all yellow) of complex I from *T. thermophilus* (PDB: 6ZIIY (245)). **B.** Comparison of the NADH-binding site of the Nqo1 subunit of complex I from *T. thermophilus* (light blue) with the flavin mononucleotide (FMN) site in HydB; the high similarity suggests NADH binds in the proximity of FMN in HydABC similar to complex I. **C.** Electron transfer network in HydABC compared to complex I from *T. thermophilus* with edge-to-edge distances indicated in bold. The red, blue, and green dotted lines indicate the cofactors present in the HydA (Nqo3), HydB (Nqo1), and HydC (Nqo2) subunits, respectively. Note that our structure is of the apo-HydABC and lacks the [2Fe]H subcluster of the H-cluster.

### *Clostridium pasteurianum*, CpI

The HydA subunit has close structural homology (35% sequence identity) to the well-characterized monomeric non-bifurcating [FeFe]-hydrogenase from *Clostridium pasteurianum*, CpI. In contrast to ‘electron-bifurcating’ [FeFe]-hydrogenases, non-bifurcating [FeFe]-hydrogenases use a single redox partner, typically ferredoxin. Aligning the two enzymes (using holo-CpI containing the [2Fe] subcluster) shows high similarity (rmsd 1.119 Å) and excellent conservation of the iron-sulfur clusters, including the A4 cluster, which

connects neighbouring HydA subunits in HydABC (Fig.3.31). However, in CpI, for which ferredoxin is the only redox partner, the cluster homologous to A4 is thought to lead to the ferredoxin-binding site (246), although a study on the related enzyme from *Clostridium acetobutylicum* (CaHydA) came to a different conclusion (247). The multimerization of HydA blocks this site, so the two enzymes must have different ferredoxin-binding sites. This rearrangement is an example of how closely related systems may have different electron transfer pathways formed by different multimerization of their subunits.

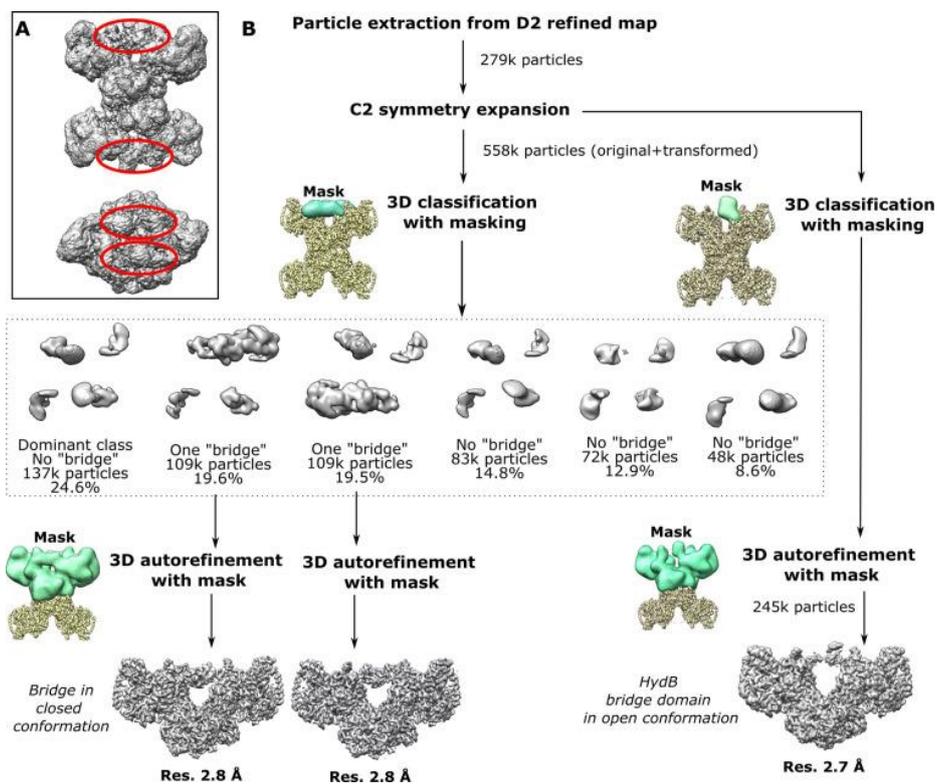


**Figure 3.31** Comparison of the HydA subunit of the electron bifurcating [FeFe]-hydrogenase from *Thermotoga maritima* with the [FeFe]-hydrogenase (CpI) from *Clostridium pasteurianum*. **A.** HydA from *Thermotoga maritima* (red) compared with CpI hydrogenase from *Clostridium pasteurianum* (orange) [Artz et al., 2020, PDB: 6N59]. **B.** Electron transfer network in HydA showing the iron–sulfur cluster that connects adjacent HydABC protomers (red circle). **(C)** Electron transfer network in CpI, with Cp ferredoxin predicted to bind closely to the [4Fe–4S] cluster on the right (246), although the [2Fe–2S] cluster has also been suggested (247). Edge-to-edge electron transfer distances are indicated in bold. The  $2\text{H}^+/\text{H}_2$  interconversion reaction in **(B)** illustrates the site at which this reaction occurs, but this will only occur in the fully assembled H-cluster including  $[2\text{Fe}]_{\text{H}}$ .

### 3.3.8 Investigating mobile domains: ‘The bridges’ domains

In the obtained map a blurry density was observed when a low-density threshold was applied (Fig.3.32.A). This indicated that mobile domains were present. To investigate it further a data processing technique called symmetry expansion (248) was used in combination with masks. Symmetry expansion is used where the symmetry of a macromolecule is broken for a reason such as a conformational change or a binding partner being bound. Due to this symmetry-breaking, imposing symmetry within Relion will not work to improve resolution and potentially relevant conformation changes will be missed. Symmetry expansion works by

expanding each particle into a number of sub-particles by applying a given symmetry operator; the asymmetric subunit is then separated by applying a focused mask. The asymmetric units can then be classified and explored individually. Given the D2 core of TmHydABC, we originally used D2-expansion, however, it is clear that the bridge domain (discussed below) disrupted the D2-symmetry so in the end C2 symmetry was used. The approach produced two new structures showing the mobile domains (called ‘bridge’ domains) in two positions: a ‘closed’ and an ‘open’ conformation. The data processing pipeline is summarized in Fig.3.32.B.

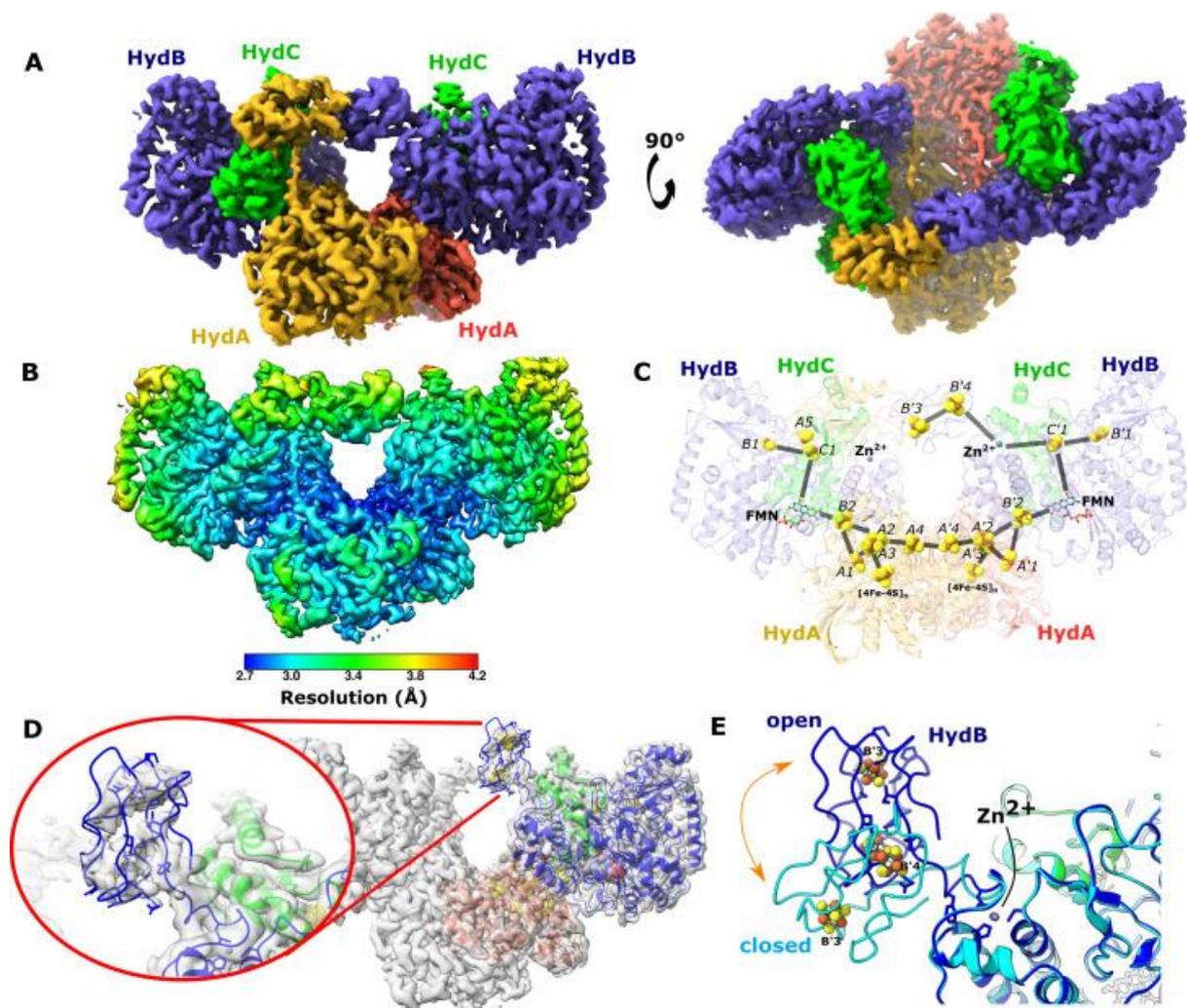


**Figure 3.32.** Classification and refinement of the symmetry-expanded cryo-EM density maps for TmHydABC using the RELION pipeline. **A.** Map at low-density threshold shows blurry regions between the HydBC lobes (red circle). **B.** Symmetry expansion, classification, and refinement using the extracted particles from the high-resolution D2 map using RELION pipeline. Following particle extraction, C2 symmetry expansion was applied, effectively doubling the dataset and allowing independent classification of the two subparticles top and bottom (Hyd(ABC)<sub>2</sub> units) in each Hyd(ABC)<sub>4</sub> complex. 3D classification of the expanded dataset with masking revealed two classes with a clear bridging density. Those two classes were refined separately revealing a bridge backward and forward conformation.

### **3.3.9 The mobile domains revealed**

A bridging domain formed by the flexible C-termini of the HydA and HydB subunits was revealed after data processing as described above (local resolution of these domains is  $\sim 3$  Å). Only two conformations of these domains were found: a ‘closed’ (Fig.3.33.A) and an ‘open’ (Fig.3.33.D) conformation.

These two bridging domains (CT of HydB and CT of HydA) are present on both sides, however, no structure was found with both bridges ‘closed’ and only one side at a time could be resolved.

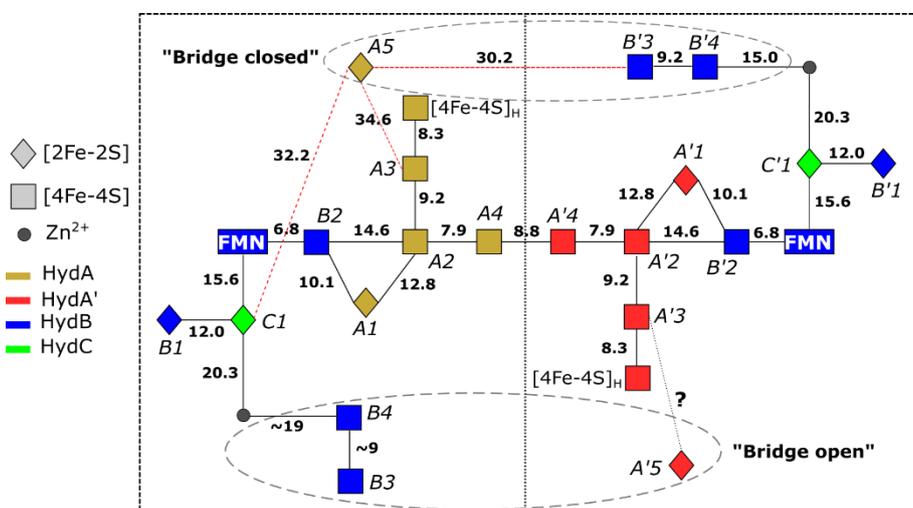


**Figure 3.33** Cryo-EM structures of the closed-bridge and open-bridge conformations of HydABC from *Thermotoga maritima*. **A.** The unsharpened 2.8 Å map of the bridge ‘closed’ conformation. The map shows only the Hyd(ABC)<sub>2</sub> unit as the two Hyd(ABC)<sub>2</sub> units constituting the Hyd(ABC)<sub>4</sub> complex were found to be independent after 3D classification. All four copies of HydB and C are coloured blue and green, respectively. The two HydA copies are in light brown and light red. **B.** Local resolutions were estimated using the local resolution function in RELION with default parameters. **C.** The atomic model that was built into the map density with the iron–sulfur electron transfer chain. **D.** Map showing the HydB bridge domain in the open position and its fitted model. **E.** Zn<sup>2+</sup> hinge region, showing the two possible conformations of the HydB bridge domain, open (blue) and closed (light blue).

### 3.3.10 The overall cofactor arrangement

The complete cofactor arrangement of a Hyd(ABC)<sub>2</sub> unit is shown in Fig.3.34. This figure also includes distances between cofactors. Based on the data collected (including both bridges ‘open’ and ‘closed’ conformation) it seems that iron-sulfur clusters B3/B4 in the mobile CT of HydB and A5 in the mobile CT of HydA are disconnected from the rest of the electron transport

chain. However, combined large domain movements of HydC and the CT of HydB and HydA may allow these ‘isolated’ clusters (B3/B4 and A5) to connect with the rest of the electron transfer chain (likely through C1).



**Figure 3.34** Schematic of the electron transfer network of one of the two identical Hyd(ABC)<sub>2</sub> units showing edge-to-edge distances (Å) between the components. Represented are the iron–sulfur clusters, [4Fe–4S]<sub>H</sub> subcluster of the H-cluster, flavin mononucleotide FMN, and Zn<sup>2+</sup> site; the bridge components and Zn site are enclosed in a dashed ellipse. Each of the two HydABC protomers constituting the Hyd(ABC)<sub>2</sub> unit is included within a dashed rectangle. Here, the top bridge is represented in its closed conformation, while the bottom one is in its open conformation. Note that our structure is of the apo-HydABC and lacks the [2Fe]<sub>H</sub> subcluster of the H-cluster.

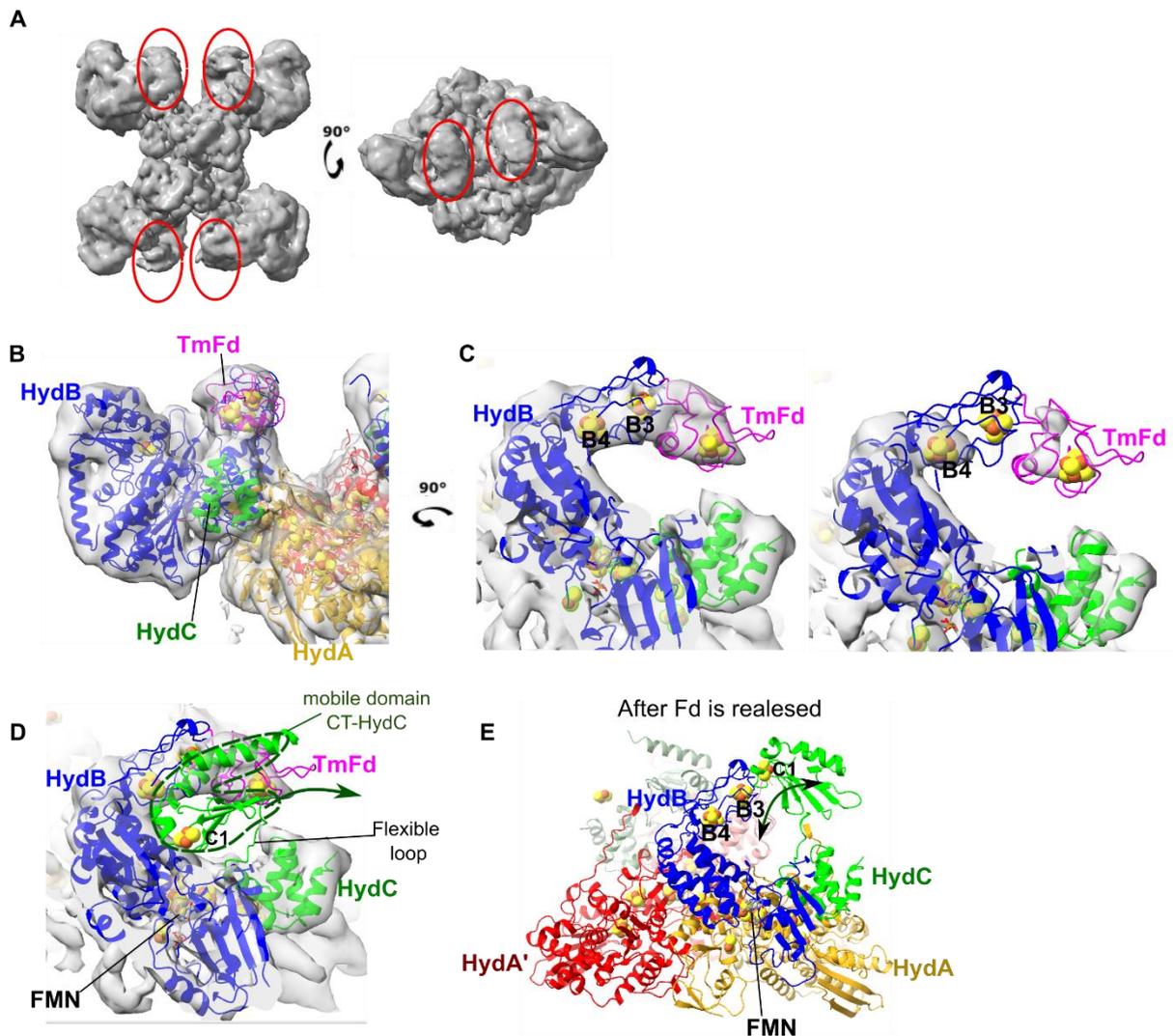
### 3.3.11 Exploring the ferredoxin binding site

From the homology between HydB and Nqo1 and the strong similarity between FMN sites (Fig.3.30.B), it is reasonable to assume that NADH binds in proximity to FMN similarly as in complex I. Furthermore, there are no other suitable binding sites for NADH other than this.

The H-cluster is instead the site of hydrogen reduction/oxidation. Remained to be clarified the site of ferredoxin binding. To do so, a solution containing the apo-enzyme was mixed with *T. marima* ferredoxin (see materials and methods) immediately before grid preparation. Then the grid was imaged. It was noted a blurry region in a different position compared to the previous results (Fig.3.35.A vs. Fig.3.32.A). A mask was then used to classify particles and resolve the bridge domains in this novel conformation (after C2 symmetry expansion was used to effectively double the number of particles). 53,000 particles were selected this way and were used for the final 3D reconstruction. The bridge domain region was poorly resolved (Figure 3.35.C) and therefore only approximate fitting with the already available PDB coordinates of

the mobile CT-HydB domain (also referred to as the ‘bridge domain’ of HydB) was attempted. The stronger signal coming from the iron-sulfur clusters was used as a guide. In addition, the PDB structure of *T. maritima* ferredoxin was fitted as well guided by the iron-sulfur cluster signal. From the cryo-EM map, it appears that when ferredoxin binds a portion of the HydC domain is displaced (Fig.3.35.D) blocking electron transfer between FMN and the [2F2-2S] cluster in HydC (C1). Then when Fd is released this mobile portion of HydC may be brought in proximity of the B3 or B4 cluster (Fig.3.35.E) allowing electron transfer between C1 and B3/B4 (for this movement there is no experimental evidence, however, geometry and the presence of the flexible loop in HydC are compatible with this hypothesis). However, general caution in interpreting this poorly resolved region should be used. Because of the very high flexibility of the region, it is also possible that this conformation is intrinsically not resolvable to a sufficient level even with additional efforts.

Note: the hypothesis that Fd binds in the proximity of the C-terminus of HydB in the mobile ‘bridge domain’ was also proposed in a homologous bifurcating hydrogenase (207).



**Figure 3.35** Cryo-EM structures of the ferredoxin-bound structure of HydABC from *Thermotoga maritima*. **A.** Map at low-density threshold shows blurry regions within the same HydBC lobe. **B.** *TmHydABC* PDB model was fitted on the Fd-bound Cryo-EM map. Part of HydC is missing as there was no density to account for it, indicating that it moved. **C.** Side views of the CT-HydB bound to *TmFd* at different density thresholds. The PDB model of *T. maritima* ferredoxin (*TmFd*) was obtained from PDBID: 1ROF (199). **D.** The mobile domain of HydC containing the C1 cluster must have moved as no density in the map is present where it was previously found. **E.** Proposed movement of the mobile domain of HydC containing the C1 cluster. After Fd is released the already displaced mobile domain of HydC comes in proximity with B3 and receives/donates an electron before returning to its stable position with C1 in the proximity of FMN.

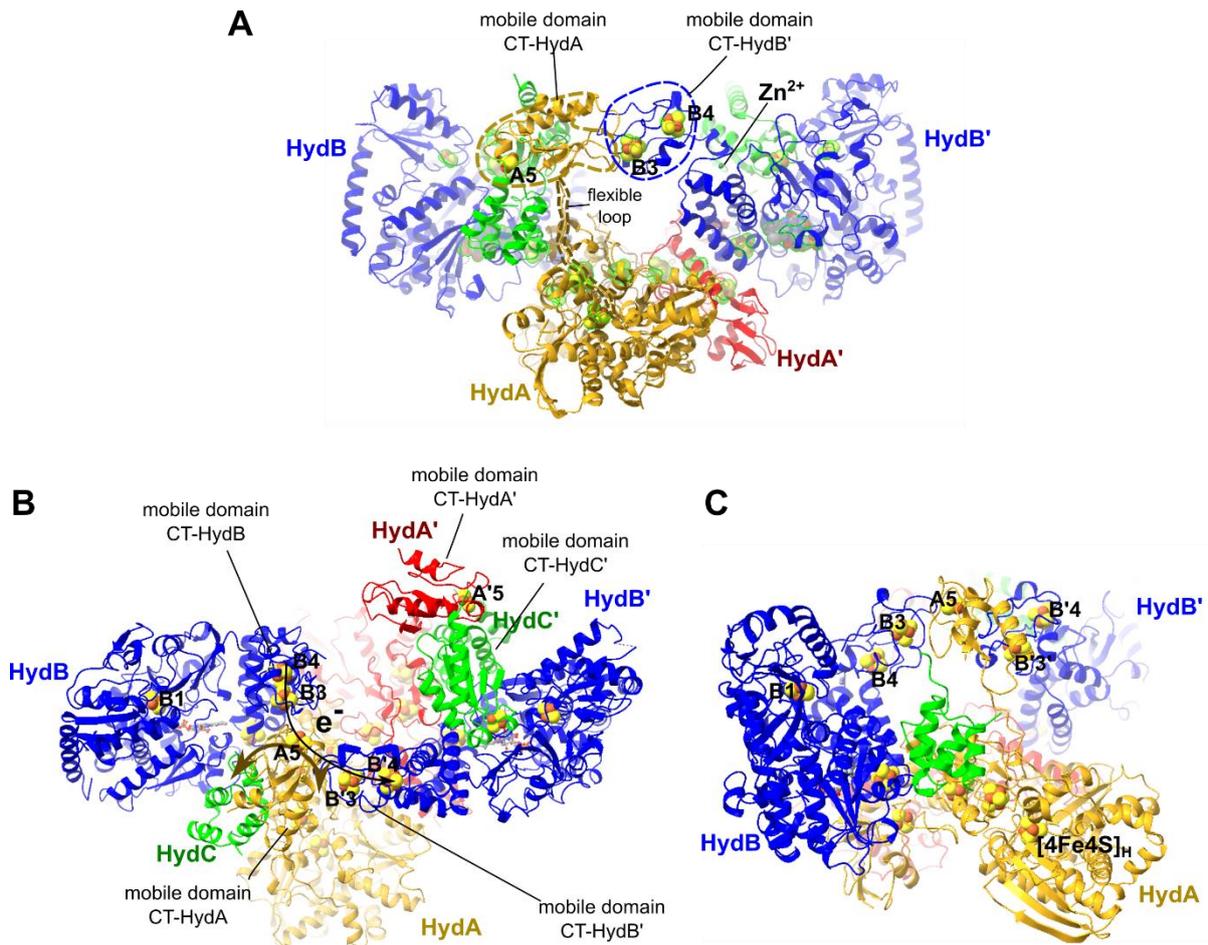
### 3.3.12 The function of the mobile C-terminus of HydA

The C-terminus (CT) domain of HydA is connected with the rest of the HydA through a long flexible loop (Fig.3.36.A) as can be seen in the resolved ‘bridge closed’ conformation (PDBID: 7P8N (171)). We could only see this mobile CT domain of HydA in the ‘bridge closed’ conformation, however, it is reasonable to imagine this domain moving toward the B3 cluster in HydB (HydB in a position similar to the one found in the Fd-bound structure). The CT domain in HydA can then receive or donate an electron from/to B3/B4, and finally move towards the opposite mobile CT domain of HydB’ in the opposite HydABC protomer and transfer electrons (Fig.3.36.B,C).

Therefore, probably, the CT domain of HydA serves as an ‘electrical’ connection between the two protomers in the Hyd(ABC)<sub>2</sub> unit to improve the catalytical rate; for example, if an Fd molecule reduces a B3 cluster but the HydABC protomer is still not ready to progress with the reaction (e.g. waiting for an NADH molecule to bind, etc.) or parts of HydABC protomer are damaged, then the opposite HydABC’ protomer can continue the reaction.

It is possible that the CT domain of HydA is the only effective functional connection between the two protomers in Hyd(ABC)<sub>2</sub> as the A4 and A4’ clusters (Fig.3.34) have a histidine ligand which significantly increases their reduction potential relative to the other nearby iron-sulfur clusters. A similar conclusion on the connection along the main chain was proposed previously in homologous supposedly bifurcating hydrogenases (207).

Furthermore, all biochemically characterized putative electron-bifurcating [FeFe]-hydrogenases, except for *TmHydABC*, possess the mobile CT domain of HydB but lack the mobile CT domain of HydA (249, 250) suggesting a less critical role for this latter domain. This is compatible with my hypothesis that the mobile CT domain of HydA serves purely as an ‘optional’ electron transfer connector between protomers. Most likely the HydABC protomer without the mobile CT domain of HydA is the minimal functional unit as it was already experimentally shown in a bifurcating hydrogenase from *Thermoanaerobacter kivui* (207), however, mutagenesis studies on *TmHydABC* are needed to claim this for certain.



**Figure 3.36** **A.** The atomic model that was built into the map density corresponding to the ‘bridge closed’ conformation (map in Fig.3.33.A, PDBID:7P8N (171)). The flexible loop connecting the mobile CT of HydA with the rest of the HydA subunit is highlighted. **B.** Top view of a constructed PDB model that shows how the mobile CT of HydA can potentially act as an electrical connection between the two HydABC protomers in the Hyd(ABC)<sub>2</sub> unit. The mobile CT-HydA domain moves and rotates facilitated by the high flexibility of its flexible loop. No experimental evidence is available for this model (only based on geometry and flexibility considerations). **C.** Side view of the same model represented in B. In both B and C, the mobile CT domain of HydC is not shown as it’s unclear where that may be positioned.

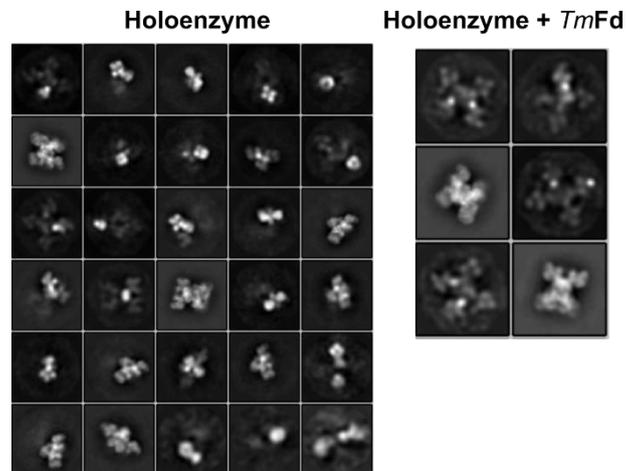
### 3.3.13 Holoenzyme structure

After anaerobic gloveboxes arrived in our lab, we attempted to make anaerobic grids with the fully reconstructed *Tm*HydABC (holoenzyme) with and without ferredoxin. The Vitrobot device was mounted inside the anaerobic glovebox and grids were made as described in the materials and methods chapter.

However, only ~9,000 intact particles and ~2,000 intact particles were found on the datasets without and with ferredoxin respectively with the vast majority of particles being degraded (2D

classes in Fig.3.37). With the available data, no sufficiently high-resolution structure could be obtained to make any conclusions.

The damage observed on the sample may have been caused by imperfect anaerobic handling or phenomena related to the high static observed during grid preparation.



**Figure 3.37** 2D classes of the holoenzyme TmhydABC with and without *T. maritima* ferredoxin (*TmFd*) showing significant protein degradation.

## 3.4 Critical Discussion

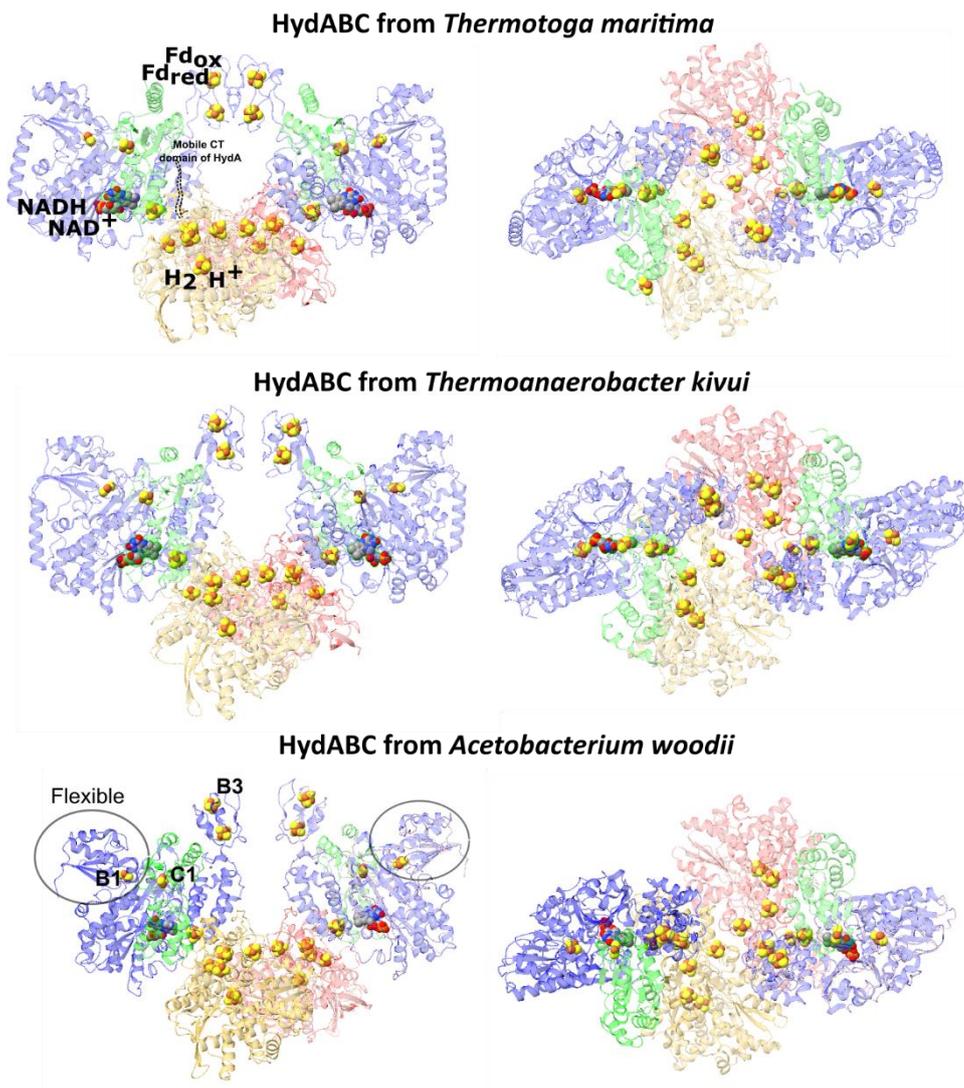
### 3.4.1 Mechanism premises

Based on the obtained PDB structures (PDBID: 7P5H, 7P8N, 7P91, 7P92 (171)) and observation of the ferredoxin-bound map (Chapter 3.3.11) it is possible to rationalize a more complete electron transfer network in which the movements of the mobile CT domains of HydA, HydB, and HydC are taken into account.

Exploratory manipulations were performed on ChimeraX by intuitively respecting degrees of flexibility and steric occupations of these mobile domains and the results have been already presented:

- In Fig.3.35.D, E it is shown how the mobile domain of HydC may move. This domain contains cluster C1 which acts as an electron mediator by dynamically connecting B3 with FMN and B1.
- In Fig.3.36.B, C it is shown how movements of the mobile CT domain of HydA can allow electron flow between two opposite mobile CT domains of HydB effectively dynamically ‘electrically’ connecting the two protomers in the Hyd(ABC)<sub>2</sub> unit.

However, in this discussion, the movements of the mobile domain of HydA are ignored as it is likely the enzyme would function without this domain (see Chapter 3.3.12). Furthermore, we will consider that a single HydABC protomer without the mobile HydA domain is the minimal functional unit. This is justified by mutagenesis results on homologous bifurcating [FeFe]-hydrogenases (207) and the structure of the homologous [NiFe]-HydABC<sub>2</sub>SL hydrogenase from *A.mobile* (211). Given the strong similarity and almost identical cofactor arrangement between this family of Nqo1 homologous HydABC hydrogenases (Fig.3.38), this model (presented in the next subchapters) can be easily generalized.



**Figure 3.38** HydABC PDB models of [FeFe]-hydrogenases from *T. maritima* (this work), *Thermoanaerobacter kivui* (PDBID:8BEW), and *Acetobacterium woodii* (PDBID:7Q4V). In *Acetobacterium woodii* HydABC an additional region in HydB was shown to have significant flexibility facilitating electron transfer between C1 and B3 (207) (most likely the dangling helix in the other structures reduces the flexibility of this domain). Only HydABC from *T. maritima* has a mobile CT domain in HydA (here not shown).

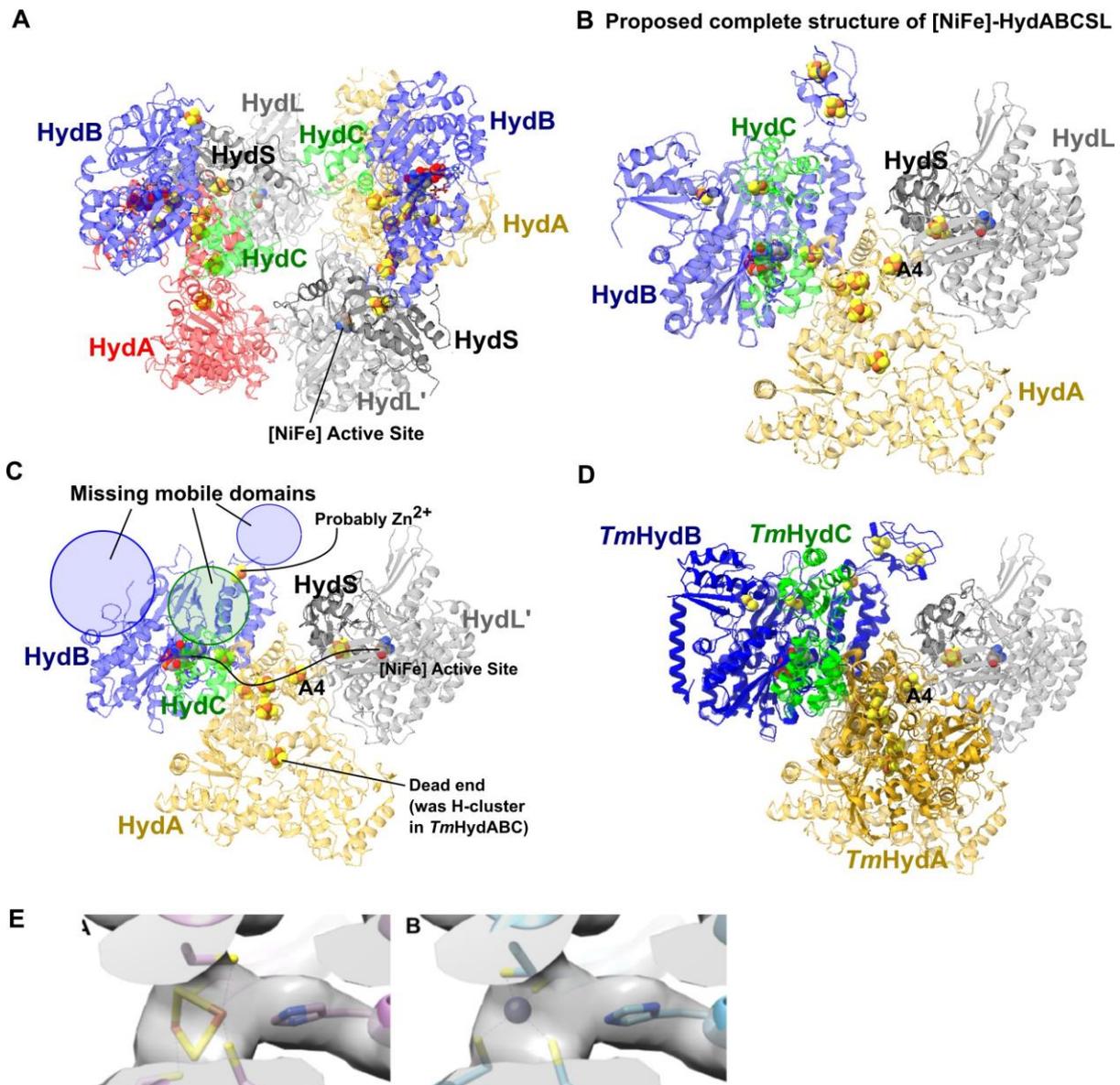
### 3.4.2 Some observations on NiFe-HydABC<sub>SL</sub> hydrogenase from *A. mobile*

[NiFe]-HydABC<sub>SL</sub> hydrogenase from *A. mobile* is a homologous hydrogenase to *TmHydABC* that catalyzes the reversible reduction of NAD<sup>+</sup> and Fd<sub>ox</sub> from H<sub>2</sub> (211). In [NiFe]-HydABC<sub>SL</sub> two independent HydABC<sub>SL</sub> protomers face each other (Fig.3.39.A) with the corresponding cluster to the ‘connector’ A4 cluster in *TmHydABC* repurposed to connect the chain with the [NiFe] active site in HydL (Fig.3.39.C).

The cofactor arrangement of a single protomer seems highly preserved (Fig.3.39.D) with the only difference being that H<sub>2</sub> oxidation/reduction happens in HydL instead of HydA. Due to the similarities with *TmHydABC* and other presented [FeFe]-HydABC hydrogenases (Fig.3.38), it is likely that the mechanism remains very similar.

Unfortunately, as shown in Fig.3.39.C, the PDB structures (PDBID: 7T30, 7T2R (211)) of this [NiFe]-hydrogenase are missing both the mobile CT domains of HydA and HydB (sequence predicted) and, in addition, the Zn site may have been exchanged for a [2Fe-2S] site (Fig.3.39.E). Also, the NT domain of HydB is missing (this domain was shown to be mobile also in *A. woodii* HydABC (207)).

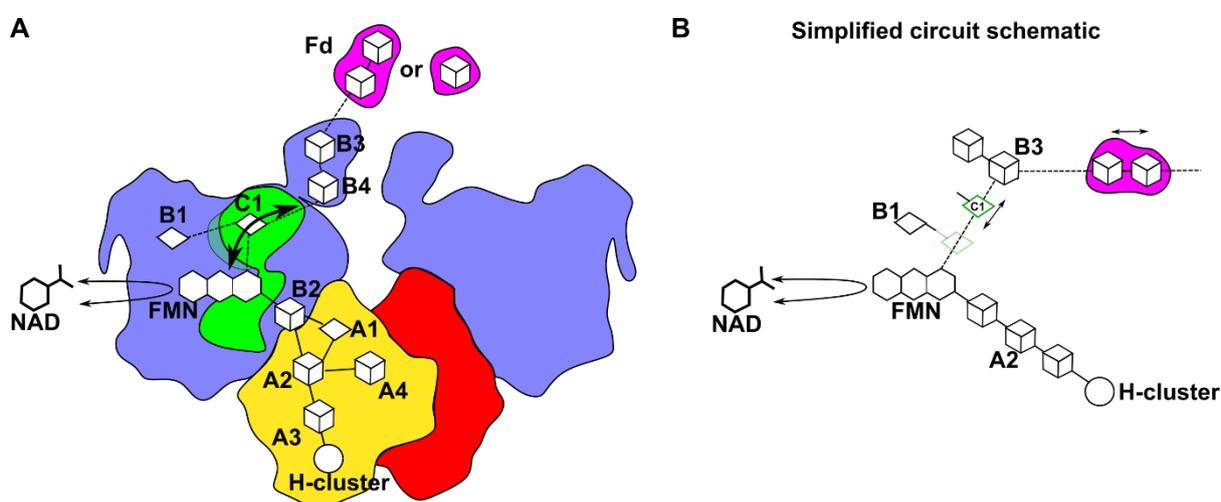
However, by fitting the PDB structures of HydABC SL with those of *TmHydABC* (Fig.3.39.D) or other [FeFe]-HydABC hydrogenases in combination with homology models (Phyre2 server (144)) a complete structure of [NiFe]-HydABC SL can be predicted (Fig.3.39.B). This data further supports the evidence that a single HydABC protomer is the minimal functional unit in homologous [FeFe]-hydrogenases (HydABC SL showed stoichiometric coupling activity even though A4 is repurposed).



**Figure 3.39** **A.** PDB models of [NiFe]-HydABC SL from *A. mobile* (PDBID: 7T30 (211)). **B.** artificially reconstructed model of [NiFe]-HydABC SL to account for missing parts. A combination of homology models (Phyre2) and available structures of [FeFe]-HydABC was used. **C.** PDB model of a single HydABC SL protomer from *A. mobile* (PDBID: 7T30) showing the missing mobile domains (circles). **D.** PDB model of a single HydABC SL (PDBID: 7T30) overlapped with a *TmHydABC* protomer (PDBID: 7P8N (171)) showing similar arrangements of the subunits HydA, HydB, and HydC. **E.** On the left, the cofactor centre in chain B (HydB) as published with a [2Fe–2S] cluster, PDBID: 7T30 (211). On the right, the same centre but built with a tetrahedral Zn; other tetrahedral metals with similar geometry will also provide a satisfactory fit, given the resolution ( $\approx 3$  Å). In both cases, the map (EMD-25647) is displayed identically, at a threshold of 0.138.

### 3.4.3 Generalized electron pathway model of the minimal functional unit

Based on the above premises, a generalized electron pathway model of the minimal functional unit of Nqo1 homologous HydABC hydrogenases can be constructed (Fig.3.40.A) as the electron transfer network is highly conserved among this class of enzymes. This pathway can be further minimalized and translated into an analogous schematic circuit (Fig.3.40.B). I will use these schematics to demonstrate that *TmHydABC* and homologous HydABC hydrogenases are potentially non-bifurcating, however, I will first demonstrate that the mechanism proposed by Jan M. Schuller *et al.* for this class of enzymes (207) is not thermodynamically coupled.



**Figure 3.40** **A.** General electron-transfer pathway in HydABC putatively bifurcating enzymes. The minimal functional unit is a HydABC protomer in a Hyd(ABC)<sub>2</sub> unit. [4Fe-4S] clusters are shown as cubes, while [2Fe-2S] clusters as rhombuses. **B.** The electron transfer pathway in A can be used to draw a simplified analogue circuit schematic. In the schematic, the H-cluster is continuously ‘electrically’ connected with FMN and can transfer/receive one electron at a time through the three [4Fe-4S] clusters. The FMN is intermittently ‘electrically’ connected with the B3 cluster through C1. C1 moves along the dashed line (like on a conveyor belt) assuming two positions: a position close to the B3 cluster where it can receive/donate an electron with B3 and a position close to FMN and B1 where it can receive/donate an electron with FMN or B1. B3 is intermittently ‘electrically’ connected with ferredoxin (Fd, magenta).

#### *Jan M. Schuller et al.’s proposed mechanism is NOT thermodynamically coupled*

Jan M. Schuller *et al.* proposed a mechanism in the bifurcating direction (Equation 3.20, electron donor H<sub>2</sub> in yellow) (207) that I argue is not thermodynamically coupled.



In summary, they propose that the bifurcating reaction occurs in the following three main substeps. Note that in the original paper, cluster B2 is cluster B1 of my model (they published

after us but did not consistently keep our nomenclature). To avoid confusion I will only use our nomenclature as in Fig.3.40.

- 1)  $H_2 \rightarrow FMN \rightarrow B1/C1$
- 2)  $H_2 \rightarrow FMN \rightarrow NAD(P)^+$  ( 3.21)
- 3)  $B1/C1 \rightarrow Fd$

In step 1,  $H_2$  donates two electrons sequentially to FMN (the electron transport chain only allows sequential one-electron transfer). The fully reduced FMN ( $FMNH^-$ ) then donates one electron to C1 and another electron to B1.

In step 2, a second molecule of  $H_2$  donates two electrons sequentially to FMN. Then FMN donates two electrons concertedly to  $NAD^+$  (hydride transfer), which becomes fully reduced (NADH). NADH is released.

In step 3, the reduced C1 and B1 clusters donate their electrons to a ferredoxin substrate (a single two-electron acceptor Fd or two one-electron acceptors Fd molecules). In detail, the transfer is:  $C1 \rightarrow B4 \rightarrow B3$ ,  $B1 \rightarrow C1 \rightarrow B4$ ,  $B4 \rightarrow Fd$ ,  $B3 \rightarrow B4 \rightarrow Fd$  (see Fig.3.40.A).

Based on the proposed reactions, the Equalities in Table 3.6 can be obtained at equilibrium.

**Table 3.6** Equilibrium conditions based on proposed reactions. FMN: fully oxidized (FMN), semi-reduced ( $FMN^{\bullet-}$ ), and fully reduced ( $FMNH^-$ ). Note B1 and C1 in Step 1 can be swapped without changing the results. Also, note that in Fd the reduction potential remains almost constant in both separate reductions (227).

Step 1	Step 2	Step 3
$E_{H_2/H^+} = E_{FMN/FMNH^-}$	$E_{H_2/H^+} = E_{FMN/FMNH^-}$	$E_{B1_{ox}/B1_{red}} = E_{C1_{ox}/C1_{red}} = E_{Fd_{ox}/Fd_{red}}$
$E_{FMN/FMN^{\bullet-}} = E_{B1_{ox}/B1_{red}}$	$E_{FMN/FMNH^-} = E_{NAD^+/NADH}$	
$E_{FMN^{\bullet-}/FMNH^-} = E_{C1_{ox}/C1_{red}}$		

However, we also know that:

$$E_{FMN/FMNH^-} = \frac{1}{2}(E_{FMN/FMN^{\bullet-}} + E_{FMN^{\bullet-}/FMNH^-}) \quad (3.22)$$

And therefore by combining Formula 3.22 with the equalities in Step 1 and Step 3:

$$E_{FMN/FMNH^-} = \frac{1}{2}(E_{B1_{ox}/B1_{red}} + E_{C1_{ox}/C1_{red}}) = E_{Fd_{ox}/Fd_{red}} \quad (3.23)$$

$E_{FMN/FMNH^-}$  is also equal to  $E_{H_2/H^+}$  in both Steps 1 and 2 and therefore the following equalities (Equalities 3.24) can be derived. Note that if the protein environment changes the reduction potential of FMN in Step 2 relative to Step 1 the following equalities remain true (easily provable by using FMN' in Step 2).

$$E_{H_2/H^+} = E_{NAD^+/NADH} = E_{Fd_{ox}/Fd_{red}} \quad (3.24)$$

Equalities 3.24 prove that this proposed mechanism is a case of kinetic coupling without thermodynamic coupling ('false bifurcation' presented in Chapter 3.1.11).

Note: the authors use an unnecessarily complicated method to model bifurcation (supplementary materials (207)) which is conceptually flawed.

### 3.4.4 The mechanism of HydABC: Can it be bifurcating?

I will now demonstrate that based on the experimental cofactor arrangements of HydABC-type enzymes, it is physically impossible to construct any bifurcating mechanism.

A bifurcating mechanism requires the presence of thermodynamic coupling as per the definition of bifurcation (Chapter 3.1.7). Mathematically, in HydABC-type enzymes, this can be expressed using the reduction potentials at equilibrium of the couples  $H_2/H^+$ ,  $NAD^+/NADH$ ,  $Fd_{ox}/Fd_{red}$  ( $E_{H_2/H^+}$ ,  $E_{NAD^+/NADH}$ ,  $E_{Fd_{ox}/Fd_{red}}$ ) as:

$$E_{H_2/H^+} = \frac{1}{2}(E_{NAD^+/NADH} + E_{Fd_{ox}/Fd_{red}}) \text{ with } E_{NAD^+/NADH} \neq E_{Fd_{ox}/Fd_{red}} \quad (3.25)$$

The above is a key equation in defining thermodynamic coupling and I will demonstrate that it is intrinsically impossible based on cofactor arrangements to satisfy it. I will reason in the bifurcating direction, however, the demonstration is readily extendible to the confurcation direction.

#### ***H-cluster***

The first consideration is on the H-cluster. After an  $H_2$  molecule binds to the H-cluster a series of reactions occur resulting in a sequential release of two electrons. The H-cluster goes from the fully reduced state ( $H_{sred}$ ) to the semi-reduced state ( $H_{red}$ ) and finally to a fully oxidized state ( $H_{ox}$ ) as described in Chapter 3.1.2. NB: inconsistency with nomenclature as per literature use, 'sred' here means super-reduced and not semi-reduced.

At equilibrium, the redox potential of the couple  $H_2/H^+$  must be equal to the redox potential of the  $H_{ox}/H_{sred}$  couple:

$$E_{H_2/H^+} = E_{H_{ox}/H_{sred}} = \frac{1}{2}(E_{H_{ox}/H_{red}} + E_{H_{red}/H_{sred}}) \quad (3.26)$$

Both electrons released from the H-cluster must travel sequentially through the A3-A2-B2 chain (Fig.3.40) as there are no other possible alternatives based on the PDB structures (even considering structural flexibility).

At equilibrium, the redox potentials of both the  $H_{ox}/H_{red}$  and  $H_{red}/H_{sred}$  couples must be equal to the redox potentials of the  $A3_{ox}/A3_{red}$ ,  $A2_{ox}/A2_{red}$ , and  $B2_{ox}/B2_{red}$  couples. Therefore:

$$E_{H_{ox}/H_{red}} = E_{H_{red}/H_{sred}} \quad (3.27)$$

Using the above equality in Equation 3.26:

$$E_{H_2/H^+} = E_{H_{ox}/H_{sred}} = E_{H_{ox}/H_{red}} = E_{H_{red}/H_{sred}} \quad (3.28)$$

Equation 3.26 had a similar structure to Equation 3.25 and could potentially lead to a bifurcation mechanism with the H-cluster as a bifurcating centre, however, I have just shown through the Equalities 3.28 that this is impossible due to the common electron transfer pathway.

### ***FMN (first option)***

The two electrons coming sequentially from the H-cluster must travel to FMN as no other options are available. The fully oxidized FMN (FMN) after receiving the first electron becomes semi-reduced ( $FMN^{\bullet-}$ ).  $FMN^{\bullet-}$  receives the second electron and becomes fully reduced ( $FMNH^-$ ).

Based on the just described sequence of reductions, at equilibrium, the following equalities must be satisfied (note that the next formulas do not change by swapping  $E_{FMN/FMN^{\bullet-}}$  with  $E_{FMN^{\bullet-}/FMNH^-}$ ):

$$\begin{aligned} E_{H_{red}/H_{sred}} &= E_{FMN/FMN^{\bullet-}} \\ E_{H_{ox}/H_{red}} &= E_{FMN^{\bullet-}/FMNH^-} \end{aligned} \quad (3.29)$$

Combining the above equalities with Equation 3.22:

$$E_{FMN/FMNH^-} = \frac{1}{2}(E_{FMN/FMN^{\bullet-}} + E_{FMN^{\bullet-}/FMNH^-}) = \frac{1}{2}(E_{H_{ox}/H_{red}} + E_{H_{red}/H_{sred}}) \quad (3.30)$$

However, as already demonstrated  $E_{H_{ox}/H_{red}} = E_{H_{red}/H_{sred}}$  (Equation 3.27) and therefore at equilibrium:

$$E_{FMN/FMNH^-} = E_{FMN/FMN^{\bullet-}} = E_{FMN^{\bullet-}/FMNH^-} = E_{H_{ox}/H_{red}} = E_{H_{red}/H_{sred}} = E_{H_2/H^+} \quad (3.31)$$

Formula 3.30 had a similar structure to Equation 3.25 and could potentially lead to a bifurcation mechanism with FMN acting as the bifurcating centre, however, I have just shown through Equalities 3.31 that this is impossible.

### ***FMN (second option)***

There is another option for the two electrons coming sequentially from the H-cluster: the first electron is transferred to the fully oxidized FMN (FMN) that becomes semi-reduced (FMNH<sup>•-</sup>). The FMNH<sup>•-</sup> immediately donates the electron to C1 or B1 returning to the fully oxidized state. The fully oxidized FMN then receives the second electron becoming again semi-reduced. The now semi-reduced FMNH<sup>•-</sup> has only two options left:

1. Donating its electron again to C1 or B1 returning to the fully oxidized state. However in this case NAD<sup>+</sup> can never be reduced to NADH through hydride transfer unless the next electrons coming from the H-cluster fully reduce FMN to FMNH<sup>-</sup> (as in Jan M. Schuller *et al.* proposal (207)) which falls into the ‘first option’ discussed above.
2. Receive another electron from the H-cluster becoming fully reduced (FMNH<sup>-</sup>) which again falls into the ‘first option’ discussed above.

As can be seen, this second option still requires that during the catalytic cycle, the FMN becomes fully reduced, and therefore the Equalities obtained in the ‘first option’ must be satisfied at equilibrium.

### ***From FMN to NAD***

NAD<sup>+</sup> can only receive two electrons concertedly (hydride transfer) from FMNH<sup>-</sup> and therefore (combined with Equalities 3.31):

$$E_{NAD^+/NADH} = E_{FMN/FMNH^-} = E_{H_2/H^+} \quad (3.32)$$

If the above equality is true, then substituting  $E_{H_2/H^+}$  with  $E_{NAD^+/NADH}$  in Equation 3.25:

$$E_{NAD^+/NADH} = \frac{1}{2}(E_{NAD^+/NADH} + E_{Fd_{ox}/Fd_{red}}) \quad (3.33)$$

As a consequence:

$$E_{NAD^+/NADH} = E_{Fd_{ox}/Fd_{red}} \quad (3.34)$$

The key requirement that at equilibrium  $E_{NAD^+/NADH} \neq E_{Fd_{ox}/Fd_{red}}$  (Equation 3.25), which is necessary for thermodynamic coupling, is not achievable. Q.E.D.

### ***Conclusions and summary***

I have demonstrated that with the most minimal postulates (summarized below), thermodynamic coupling (a necessary condition for bifurcation) is not achievable based on the structure and cofactor arrangement of HydABC-type enzymes.

Postulates:

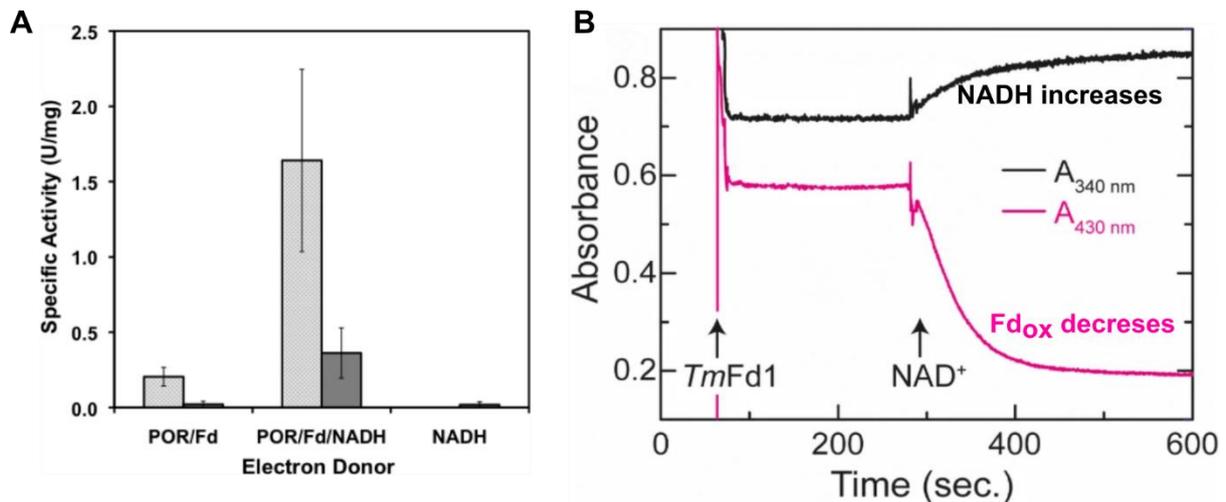
- 1) The protomer HydABC is the minimal functional unit (*demonstrated experimentally by Schuller et al. (207)*).
- 2) All electrons coming from the H-cluster are sequentially transferred through a common electron transport chain to FMN (*the PDB structures do not show any alternative way even if domain mobility is taken into account*).
- 3) It is not possible for NAD or Fd to receive/donate electrons from/to the H-cluster bypassing FMN (*the PDB structures do not show any alternative way even if domain mobility is taken into account*).
- 4) NAD can only donate/receive electrons concertedly through hydride transfer to/from FMN (*NAD is an obligate two-electron donor/acceptor and FMN is the only cofactor in the structure that can donate electrons concertedly*).

Unless at least one of postulates 2, 3, or 4 is proven to be false, it is not possible to imagine any bifurcating mechanism for HydABC-type enzymes. If the minimal functional unit is a Hyd(ABC)<sub>2</sub> (contrary to experimental evidence) thermodynamic coupling still remains unachievable as can be demonstrated by repeating similar reasoning to the above presented.

The only possible explanation left for the previously obtained enzyme assay results is that scientists may have mistaken kinetic coupling for thermodynamic coupling (Chapter 3.1.11, ‘false bifurcation’). I will now present the evidence to show that in reality, we only have solid proof of kinetic coupling while proof for thermodynamic coupling remains inconclusive.

### **3.4.5 Lack of evidence that *Tm*HydABC is bifurcating (and potential evidence against)**

There are only two published enzyme assays that were used to claim bifurcation and confurcation in *Tm*HydABC: an assay from Adams *et al.* in the confurcation direction (Fig.3.41.A) (200) and an assay from Chongdar *et al.* in the bifurcation direction (Fig.3.41.B) (150).



**Figure 3.41. A.** Substrate dependence of *TmHydABC*. The specific activity in the standard  $\text{H}_2$  production assay containing Fd/POR, NADH, or Fd/POR and NADH was determined using the cytoplasmic extract (dark bars) and the purified hydrogenase (light bars). Image taken from Adams *et al.* (200) **B.** Activity assay of *TmHydABC* using physiological partners. Simultaneous reduction of *TmFd* (decrease in  $A_{430\text{nm}}$ ) and  $\text{NAD}^+$  (increase in  $A_{340\text{nm}}$ ) by *TmHydABC* at  $70^\circ\text{C}$  under an atmosphere of 100%  $\text{H}_2$  was monitored by UV-Vis spectroscopy. To a 1 mL reaction mixture containing  $\approx 680$  ng *TmHydABC* and  $50\ \mu\text{M}$  FMN in 200 mM potassium phosphate (pH 8) buffer,  $\approx 35\ \mu\text{M}$  *TmFd* was added (indicated by the first arrow). To the same reaction mixture,  $0.5\ \text{mM}$   $\text{NAD}^+$  was added (indicated by the second arrow). Image taken from Chongdar *et al.* (150).

### Confurcation assay

In the  $\text{H}_2$  production assay (confurcation direction), a reaction mixture containing NADH and/or *T. maritima* Fd reduced enzymatically by its physiological partner POR was used to demonstrate that *TmHydABC* (both from cytoplasmic extracts and purified) required the presence of both NADH and reduced Fd to produce hydrogen (Fig.3.41.A). However, I will show that data are insufficient to assess whether in the used conditions the oxidation of NADH (the higher reduction potential donor) is endergonic or exergonic.

Based on the reported data from Adams *et al.* (200), to determine the source of the reductant for  $\text{H}_2$  production, known amounts of NADH were added to their standard assay (containing Fd), and the amount of  $\text{H}_2$  produced once the reaction reached completion was measured. The maximum amount of NADH added in a 0.5 mL reaction mixture was about 100 nmol (Fig.3.42) which resulted in a total of about 160 nmol of  $\text{H}_2$  evolved at completion. The authors reported using 8-mL stoppered serum vials and therefore it can be assumed that the 160 nmol of  $\text{H}_2$  accumulated in 7.5 mL headspace before being measured. Using the ideal gas formula ( $PV=nRT$ ) with  $T = 80^\circ\text{C}$  (as per reported conditions) the resulting partial pressure of  $\text{H}_2$  ( $P_{\text{H}_2}$ )

at equilibrium was  $6.3 \cdot 10^{-4}$  bar. At pH =7.5 (as per reported conditions) the reduction potential (in V) at equilibrium of the  $H^+/H_2$  couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.525 - 0.035 \log_{10} P_{H_2} = -0.413 V \quad (80^\circ\text{C}, \text{pH}=7.5) \quad (3.35)$$

Calculating the reduction potential of the  $NAD^+/NADH$  couple at  $80^\circ\text{C}$  requires knowledge of the standard potential of the couple at  $80^\circ\text{C}$ ,  $E_{NAD^+/NADH}^\circ(80^\circ\text{C})$ , which scales linearly with the temperature according to the temperature coefficient  $\left(\frac{dE^\circ}{dT}\right)_{298K}$  (251):

$$E_T^\circ = E_{298K}^\circ + (T - 298.15) \cdot \left(\frac{dE^\circ}{dT}\right)_{298K} \quad (3.36)$$

As I could not find the value of  $E_{NAD^+/NADH}^\circ(80^\circ\text{C})$  or that of the temperature coefficient I can only make approximate estimates. The temperature coefficient is typically in the order of  $\sim -1$  mV/K (251, 252), therefore by spanning  $\left(\frac{dE^\circ}{dT}\right)_{298K}$  in a reasonably broad range of  $-0.1$  to  $-1.2$  mV/K:

$$\begin{aligned} E_{NAD^+/NADH}^\circ(80^\circ\text{C}) &= -0.113 + 55 \cdot \left(\frac{dE^\circ}{dT}\right)_{298K} \\ \Rightarrow E_{NAD^+/NADH}^\circ(80^\circ\text{C}) &= -0.150 \pm 0.030 \end{aligned} \quad (3.37)$$

With  $E_{NAD^+/NADH}^\circ(25^\circ\text{C}) = -0.113V$  calculated from the tabulated standard value at pH=7.0 ( $-320$  mV).

Using the  $E_{NAD^+/NADH}^\circ(80^\circ\text{C})$  obtained from Formula 3.37 and substituting it into the Nernst Equation (Formula 3.7), it can be calculated that  $E_{NAD^+/NADH}^\circ(80^\circ\text{C}, \text{pH} = 7.5) = -0.413 \pm 0.030 V$ .

Using the calculated reduction potentials of the couples  $H^+/H_2$  and  $NAD^+/NADH$  it can be shown that it cannot be excluded that, at equilibrium ( $E_{H^+/H_2} = E_{NAD^+/NADH}$ ), most NADH can be oxidized (Equation 3.38) without thermodynamic coupling.

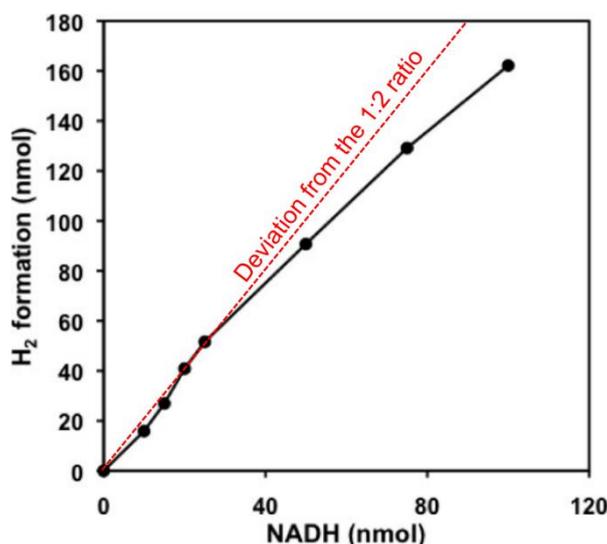
$$\begin{aligned} -0.413 &= -0.413 \pm 0.030 - 0.035 \log_{10} \frac{[NADH]}{[NAD^+]} \\ \Rightarrow \log_{10} \frac{[NADH]}{[NAD^+]} &= 0 \pm 0.86 \end{aligned} \quad (3.38)$$

I conclude that the confurcation assay performed by Adams *et al.* only provides evidence of kinetic coupling and cannot be used to claim thermodynamic coupling.

Furthermore, evidence in favour of a lack of thermodynamic coupling is the rapid deviation from the 1:2 ratio (Fig.3.42, red line) the more  $H_2$  accumulates in the headspace; if the reaction

were thermodynamically coupled, at equilibrium,  $E_{H_2/H^+} = \frac{1}{2}(E_{NAD^+/NADH} + E_{Fd_{ox}/Fd_{red}})$  and as  $Fd_{red}$  is continuously regenerated (Fd/POR system with a large excess of pyruvate) then adding 100 nmol of NADH would result in exactly 200 nmol of  $H_2$  (or a tiny bit less as enzymes are not perfectly efficient couplers) and such substantial increasing deviation from the expected ratio would not be observed.

Interesting note: if the enzyme is indeed only a kinetic coupler it can be derived from the plot in Fig.3.42 that for the  $NAD^+/NADH$  couple  $\left(\frac{dE^\circ}{dT}\right)_{298K} = -0.36 \text{ mV/K}$  (only 80 nmol of NADH out of 100 nm were oxidized).



**Figure 3.42.** Hydrogen production using NADH as the electron donor in the presence of the POR/Fd system. The amounts of NADH added to the standard  $H_2$  production assay and the amounts of  $H_2$  produced are indicated. In red it is shown how the reaction deviates from the predicted 1:2 ratio predicted by thermodynamic coupling the more  $H_2$  is produced. Image adapted from Adams *et al.* (200).

### ***Bifurcation assay***

In the  $H_2$  uptake assay (bifurcation direction), Chongdar *et al.* demonstrated that  $H_2$  oxidation starts only when  $NAD^+$  is added to the reaction mixture (150) claiming that the reduction of ferredoxin (the lower potential acceptor) is dependent on the presence of  $NAD^+$  that acts as a driver through thermodynamic coupling. However, it is easy to mathematically demonstrate that with the reported conditions  $TmFd$  would be completely (95%+) reduced even in the absence of  $NAD^+$ .

At  $pH = 8$ ,  $P_{H_2} = 1 \text{ atm}$  (1.013 bar),  $70^\circ\text{C}$  (as per reported conditions) the reduction potential (in V) of the  $H^+/H_2$  couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.544 \text{ V} \quad (70^\circ\text{C}, \text{pH}=8) \quad (3.39)$$

$E_{H^+/H_2}$  is very negative in the specified conditions. In particular, as reported previously, the midpoint reduction potential of the  $TmFd_{ox}/TmFd_{red}$  couple is -453 mV at 80°C (200). Using Equation 3.36 it is reasonable to assume that at 70°C this reduction potential would only decrease by a maximum of 15 mV remaining significantly higher than  $E_{H^+/H_2}$ . This indicates that even without a high potential ‘driver’ such as  $NAD^+$  the reduction of  $TmFd$  would proceed to completion. As in the previous case, only kinetic coupling is demonstrated but not thermodynamic coupling.

### 3.4.6 Lack of evidence that HydABC from *A. woodii*, *T. kivui*, and *R. albus* are bifurcating

#### *A. woodii*

Enzyme assays in the bifurcation direction were used by Schuchmann and Müller to claim that HydABC from *A. woodii* is bifurcating (253). However, based on reported conditions it is easy to show that also in this case the reduction of Ferredoxin would proceed to completion even in the absence of the higher potential ‘driver’  $NAD^+$ .

At pH = 8,  $P_{H_2} = 1.1$  bar, 30 °C (as per reported conditions) the reduction potential (in V) of the  $H^+/H_2$  couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.482 \text{ V} \quad (30^\circ\text{C}, \text{pH}=8) \quad (3.40)$$

$E_{H^+/H_2}$  is very negative in the specified conditions. In particular, the authors used *Clostridium pasteurianum* ferredoxin which has a midpoint reduction potential that depending on sources varies from -383 mV (25°C, pH =8) (254) to - 412 mV (room temperature, pH 6.3 to 10.0) (255). At 30 °C these potentials are unlikely to change much (max 15 mV).

Therefore, the Fd potential is significantly more positive than  $E_{H^+/H_2}$  indicating that even without a high potential ‘driver’ such as  $NAD^+$  the reduction of *C. pasteurianum* Fd would proceed to completion. As in the previous case, only kinetic coupling is demonstrated but not thermodynamic coupling.

Similarly, the assays that were recently performed by Shuller *et al.* (207) used the same conditions as above and cannot be used as proof of thermodynamic coupling.

On a note, Schuchmann and Müller (253) also argue that based on another study (256) the monomeric iron-only hydrogenase of *C. pasteurianum* (CpI) could only reduce 55% of the ferredoxin under the same reaction conditions. However, upon inspection of the reference used for the claim (256), I noted that the conditions of the CpI experiment were significantly different (reduction of Fd with 100% H<sub>2</sub> 1.2 bar, pH 7.0, 40 °C). With these conditions,  $E_{H^+/H_2}$  can be calculated to be significantly more positive ( $-0.437 V$  vs  $-0.482 V$ , Formula 3.41) making the comparison invalid.

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.437 V \quad (40^\circ\text{C}, \text{pH}=7.0) \quad (3.41)$$

(As at equilibrium, 55% of Fd was reduced it can be derived that  $E_{F_{d_{ox}}/F_{d_{red}}}^\circ(40^\circ\text{C}, \text{pH } 7.0) = -0.432 V$ )

### ***T. kivui***

Enzymatic assays (bifurcation direction) were performed also on *T. kivui* HydABC (257).

At pH = 8,  $P_{H_2} = 2$  bar, 66°C (as per reported conditions) the reduction potential (in V) of the H<sup>+</sup>/H<sub>2</sub> couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.548 V \quad (66^\circ\text{C}, \text{pH}=8) \quad (3.42)$$

$E_{H^+/H_2}$  is even lower than in the above case. The authors used *Clostridium pasteurianum* ferredoxin (as above) and therefore also here only kinetic coupling but not thermodynamic coupling was proven.

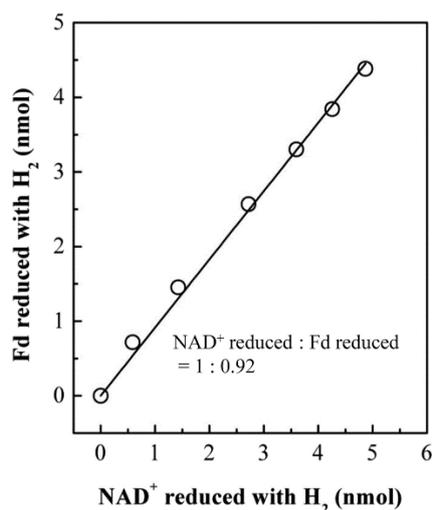
### ***R. albus***

Enzyme assays in the bifurcation direction (Fig.3.43) were performed at 37 °C, pH 7.5, and 100% H<sub>2</sub> (208). The H<sub>2</sub> pressure is not specified, thus I assumed a pressure of 1 bar.

Using these conditions the reduction potential of the H<sup>+</sup>/H<sub>2</sub> couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.461 V \quad (37^\circ\text{C}, \text{pH}=7.5) \quad (3.43)$$

$E_{H^+/H_2}$  is lower than the reduction potential of *Clostridium pasteurianum* ferredoxin (used by the authors) and therefore also here only kinetic coupling but not thermodynamic coupling was proven.



**Figure 3.43.** Stoichiometry of ferredoxin and NAD reduction by H<sub>2</sub> catalyzed by purified electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase from *R. albus*. Image taken from (208).

### 3.4.7 Potential evidence that HydABC from *M. thermoacetica* is bifurcating

*M. thermoactica* bifurcating hydrogenase was used for enzyme assays in both bifurcating and confurcating directions (209).

#### *Bifurcation direction (H<sub>2</sub> uptake)*

An enzymatic assay (Fig.3.44.A) was performed at 45°C, pH 7.5, and H<sub>2</sub> (100%) at 1.2 × 10<sup>5</sup> Pa.

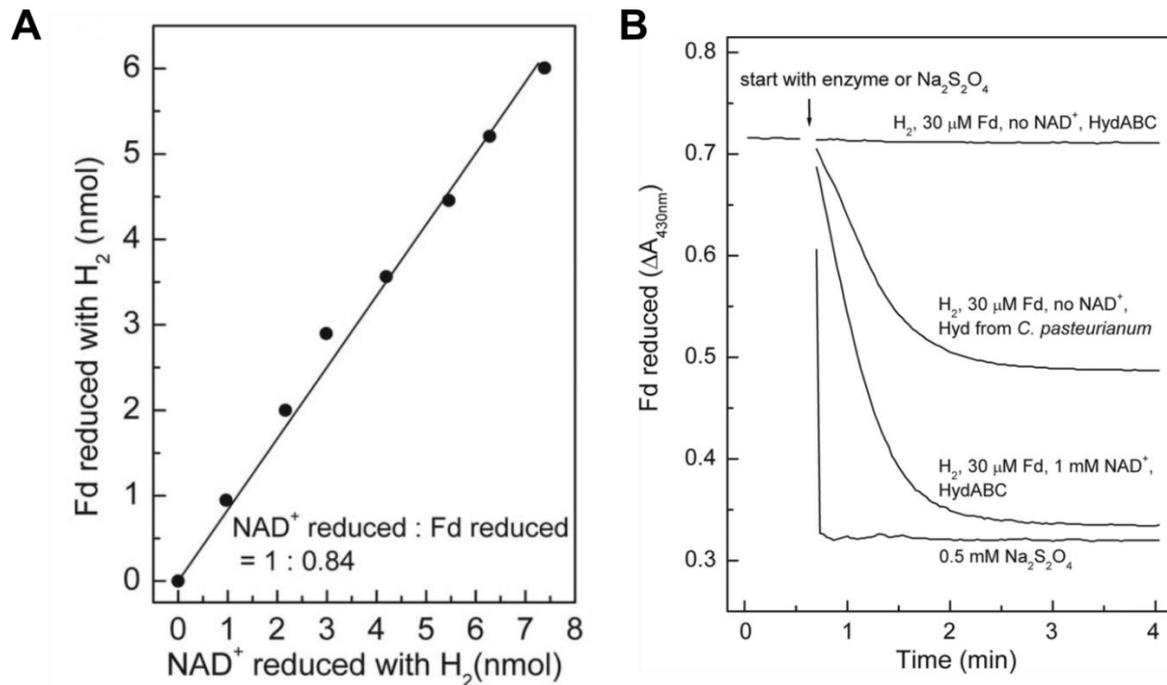
Using these conditions the reduction potential of the H<sup>+</sup>/H<sub>2</sub> couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.476 \text{ V} \quad (45^\circ\text{C}, \text{pH}=7.5) \quad (3.44)$$

$E_{H^+/H_2}$  is more negative than the reduction potential of *Clostridium pasteurianum* ferredoxin (used by the authors) and therefore here only kinetic coupling but not thermodynamic coupling was proven.

However, in a different enzymatic assay (Fig.3.44.B) there is a potential indication of thermodynamic coupling. In this particular assay, 1.5-ml anoxic cuvettes were filled with 0.8 ml of the reaction mixtures (pH 7.0) as shown in Fig.3.44.B. This experiment shows that using the same reaction condition only ~55% of ferredoxin can be reduced by the monomeric *C. pasteurianum* hydrogenase, while more than 90% is reduced with *M. thermoacetica* HydABC

in presence of both ferredoxin and  $\text{NAD}^+$ . However, this experiment seems to be derived from a single measurement and it can be shown that even a small change in measured pH can significantly affect the results, especially in the context of a comparison. In fact, at  $45^\circ\text{C}$  and  $P_{\text{H}_2} = 1.2$  bar, the reduction potential of the  $\text{H}^+/\text{H}_2$  couple changes from  $-450$  mV (pH = 7.1) to  $-438$  mV (pH=6.9), and therefore I expect a single unrepeated experiment to be unreliable. Multiple repetitions of this assay are therefore required.

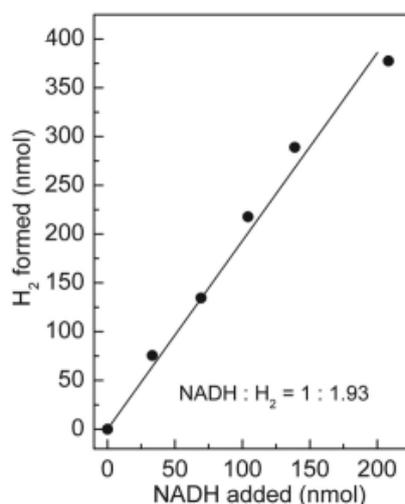


**Figure 3.44. A.** Stoichiometry of ferredoxin and NAD reduction by  $\text{H}_2$  catalyzed by electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase from *M. thermoacetica*. **B.** Fd reduction with  $\text{H}_2$  catalyzed by HydABC from *M. thermoacetica*. As controls, Fd reduction with 100%  $\text{H}_2$  catalyzed by the monomeric Fd-dependent [FeFe]-hydrogenase (Hyd) from *C. pasteurianum* as well as the spontaneous reduction of Fd with sodium dithionite were used. Images taken from (209).

### Confurcation direction ( $\text{H}_2$ formation)

Enzymatic assays were performed at  $45^\circ\text{C}$ , pH 7.0, and  $\text{N}_2$  (100%,  $1.2 \times 10^5$  Pa). A reduced ferredoxin-regenerating system was present. The reactions took place in 6.5-ml anoxic serum bottles sealed with rubber stoppers containing 0.8 ml of the reaction mixture. In this assay, different amounts of NADH are added in the presence of a reduced ferredoxin regenerating system and the amount of  $\text{H}_2$  produced was measured (Fig.3.45). As shown in the reported plot (Fig.3.45) the NADH: $\text{H}_2$  ratio remained similar also when  $\sim 200$  nmol of NADH was added and therefore it cannot be used to argue against thermodynamic coupling as done previously in

the case of *TmHydABC* (Fig.4.42). Nevertheless, enzymatic assay evidence either against or in favour of thermodynamic coupling remains scarce and has not been reproduced in other studies or measurements.



**Figure 3.45.** Stoichiometry of NADH-dependent H<sub>2</sub> formation in the presence of an excess amount of reduced ferredoxin. A reduced ferredoxin regeneration was present. Image taken from (209).

### 3.4.8 Lack of evidence that NiFe-HydABC SL hydrogenase from *A. mobile* is bifurcating

Enzyme assays (bifurcation and confurcation direction) were performed also on a homologous putatively bifurcating [NiFe]-hydrogenase from *A. mobile* (211).

#### *Hydrogen uptake (bifurcation assay)*

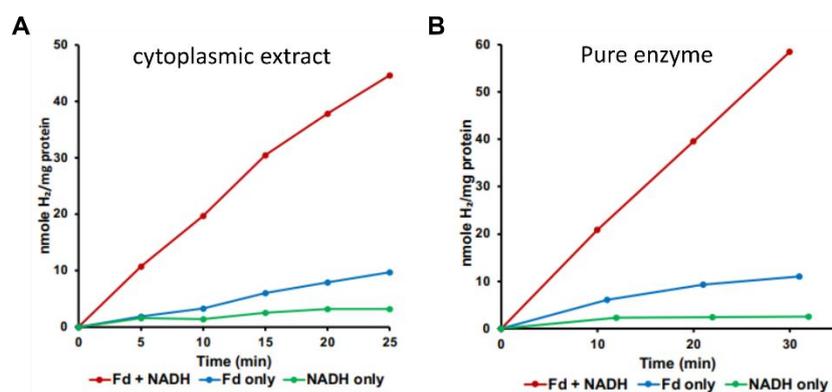
At pH = 7.5,  $P_{H_2} = 1$  atm (1.01 bar), 50°C (as per reported conditions) the reduction potential (in V) of the H<sup>+</sup>/H<sub>2</sub> couple is:

$$E_{H^+/H_2} = 0 - \frac{RT}{2F} \ln \frac{P_{H_2}}{[H^+]^2} = -0.480 \text{ V} \quad (50^\circ\text{C}, \text{pH} = 7.5) \quad (3.45)$$

$E_{H^+/H_2}$  is very negative in the specified conditions. In particular, the authors used *A. mobile* ferredoxin which is homologous to *T. maritima* hydrogenase. The reduction potential of *A. mobile* ferredoxin is not known, however, with  $E_{H^+/H_2} = -0.480 \text{ V}$  it is likely that most of the *A. mobile* ferredoxin would be reduced (the redox potential of Fd is typically around from -400 to -450 mV). The assay is inconclusive in proving proof for thermodynamic coupling.

### *Hydrogen evolution (confurcation assay)*

In the confurcation assay, the nmol of H<sub>2</sub> evolved was monitored over time in a solution containing 1 mmol NADH and 10 μM *A. mobile* Fd. A POR regenerating system was present to reduce Fd. This assay remains inconclusive and does not provide the required data to assess whether thermodynamic coupling is present. However, it clearly shows that both NADH and reduced ferredoxin are required at the same time.



**Figure 3.46.** Hydrogen production using NADH as the electron donor in the presence of the POR/Fd system using cytoplasmic extract (A) and pure enzyme (B). Image taken from (211).

### **3.4.9 HydABC-type enzymes may not be bifurcating but only perform kinetic coupling**

Based on the PDB structure I have demonstrated that with the most basic postulates HydABC-type enzymes cannot perform thermodynamic coupling, however, they can perform kinetic coupling (Chapter 3.4.4). I have also shown that the evidence available can only be used to confidently claim kinetic coupling (also referred to as ‘stoichiometric coupling’) while remaining inconclusive for thermodynamic coupling. In fact, the conditions used in most assays do not allow to prove thermodynamic coupling and those that do show conflicting results and are not reproduced.

In this case, one must wonder what the advantage of using exclusive kinetic coupling *in vivo* is. I think it probably has something to do with the amount of protein used (one enzyme instead of two) and the regulation of metabolic pathways. For example, physiologically, *T. maritima* HydABC is used in the confurcation direction where it oxidizes the NADH produced in the ED pathway and the Fd produced in the acetate fermentation steps (Fig.3.4.A). If we imagine overdriving the kinetic coupling, then we would end up speeding up the Fd oxidation while the

NADH oxidation rate would be similar or slightly slower. This would probably decrease the use of the lactate metabolism (predominant in the wild-type bacteria, see Chapter 3.1.3) by sequestering pyruvate due to a much faster Fd regeneration rate. A decrease in lactate metabolism (which uses NADH) would in turn cause a higher use of the pentose phosphate pathway resulting in altered growth (195).

It is also possible that in physiological conditions the oxidation of NADH without thermodynamic coupling is exergonic. In submarine hydrothermal vents (habitat of *T. maritima*) the temperature is high ( $T \sim 80^{\circ}\text{C}$ ) and a slightly acid pH is common  $\text{pH} \sim 5$  (258), however, cytoplasmatic pH remains unknown and can differ from the external one (259).

### 3.4.10 Proposed kinetic coupling mechanism of HydABC

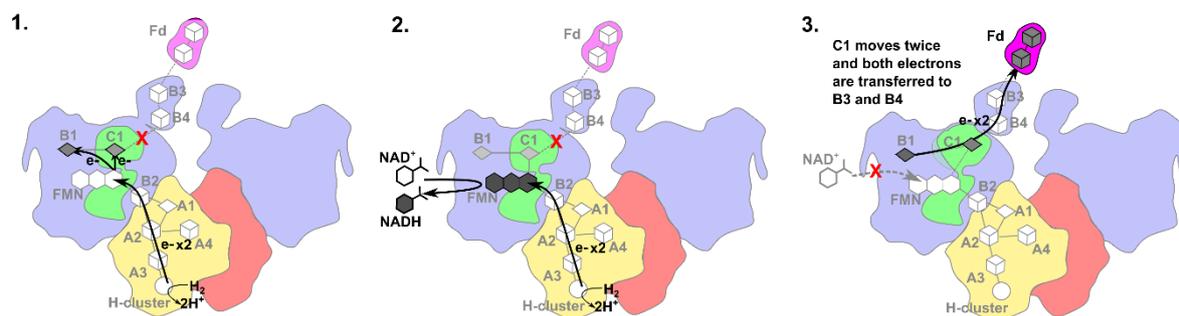
Three fundamental mechanistic requirements must be present to allow the kinetic coupling experimentally observed in HydABC enzymes:

- 1 **Electron transfer between NAD and Fd must be blocked**; otherwise, thermodynamic equilibrium concentrations that are identical to the ones in the absence of enzymes (or equivalently in the presence of two non-bifurcating enzymes such as those presented in Chapter 3.1.11) are reached.
- 2 **Alternating electron transfer** must be present:  $\text{H}_2$  reduces  $\text{NAD}^+$ , then  $\text{H}_2$  reduces  $\text{Fd}_{\text{ox}}$ , then  $\text{H}_2$  reduces  $\text{NAD}^+$ , then  $\text{H}_2$  reduces  $\text{Fd}_{\text{ox}}$ , etc., and similar in the opposite direction ( $\text{NADH}$  reduces  $\text{H}^+$ , then  $\text{Fd}_{\text{red}}$  reduces  $\text{H}^+$ , etc.).
- 3 **In the absence of either NAD or Fd, the reaction does not proceed** (kinetically blocked). Consecutive reductions of  $\text{NAD}^+$  by  $\text{H}_2$  and consecutive reductions of  $\text{Fd}_{\text{ox}}$  by  $\text{H}_2$  must be blocked. And similarly in the opposite directions (reductions of  $\text{H}^+$ ).

These three requirements are probably met by a combination of dynamic movements of domains and modulation of binding affinity for NAD(H) as proposed by Jan M. Schuller *et al.* (207). In particular, movements of the mobile CT domain of HydC are likely to control electron transfer between FMN and B3/B4 while the reduction of clusters B1 has been shown to modulate the binding affinity of  $\text{NAD(P)}^+$  (207).

The catalytical cycle proposed by M. Schuller *et al.* (207), which is composed of three main steps as in Fig.3.47, seems to provide a reasonable interpretation of a purely kinetic coupling

mechanism. However, as proved in Chapter 3.4.3, this mechanism was mistakenly used by the authors to describe thermodynamic coupling.



**Figure 3.47.** the three main steps of the catalytic cycle. **1.** A molecule of hydrogen binds to the H-cluster. The two electrons are sequentially transferred to the fully oxidized FMN cofactor which becomes fully reduced (FMNH<sup>-</sup>). FMNH<sup>-</sup> then transfers sequentially two electrons to C1 and B1 (most likely FMN → C1 → B1). **2.** A second molecule of hydrogen binds to the H-cluster. The two electrons are sequentially transferred to the fully oxidized FMN cofactor which becomes fully reduced (FMNH<sup>-</sup>). NAD<sup>+</sup> binds on top of FMNH<sup>-</sup> (increased affinity when B2 is reduced) and is reduced through hydride transfer. **3.** The release of NADH may trigger conformational changes and the reduced C1 (from step 1) is brought in proximity of B3 allowing electron transfer (C1 → B3 → B4). HydC moves again and the oxidized C1 is brought in proximity to B1. Electron transfer from B1 to C1 may trigger conformational changes that trigger another movement of the mobile domain of HydC towards B3. This leads to another electron transfer event to B3/B4 resulting in the reduction of Fd.

### 3.5 Conclusions and future work

The structure of *Thermotoga maritima* HydABC has been presented. This hydrogenase was shown to be oligomerized into two Hyd(ABC)<sub>2</sub> units. Each Hyd(ABC)<sub>2</sub> unit is composed of two HydABC protomers connected through a His-ligated [4Fe-4S] cluster.

It was shown that each subunit (HydA, HydB, HydC) contains a CT mobile domain with at least one iron-sulfur cluster. Most likely the mobile domain in HydA ‘electrically connects’ the two HydABC protomers in a Hyd(ABC)<sub>2</sub> unit while the mobile domain in HydB is the binding site for Fd. As only the mobile HydB subunit is conserved in homologous enzymes, it was proposed that only HydB is necessary for coupling. We also found a Zn site in HydB that may act as a rigid hinge.

The structure of *Tm*HydABC was followed by more recent publications of structures of homologous HydABC hydrogenases showing a highly conserved electron transfer network. However, no satisfactory mechanism was proposed that could explain thermodynamic coupling. I here showed that, based only on 4 basic observations, thermodynamic coupling seems to not be compatible with HydABC-type enzymes and that previous enzymatic assays were poorly interpreted leading to the unsupported claim that HydABC-type enzymes are thermodynamically bifurcating. I propose that most or all HydABC-type enzymes, including *Tm*HydABC, may only perform kinetic but not thermodynamic coupling.

#### *Next steps*

Rigorous enzymatic assays should be performed to confirm there is no thermodynamic coupling (e.g. by sufficiently decreasing pH in the H<sub>2</sub> uptake direction). Experimentally proving or disproving thermodynamic coupling should be prioritized over additional structural studies to avoid misguided interpretations.

# 4 Tagging Hyd-1 using CRISPR-Cas9 genome editing

## 4.1 Introduction

In this chapter, an introduction to the genome editing techniques in *E. coli* is provided followed by the experimental results obtained. *E. coli* was mutated in order to add a strep tag to the N-terminus of the HyaB subunit of Hydrogenase-1. The tag type and position were selected to improve the protein preparation (details in Chapter 5) and obtain Hyd-1 in the biologically relevant cytochrome-bound structure.

### 4.1.1 Genome editing techniques in *E. coli*

*Escherichia coli* is a Gram-negative, rod-shaped bacterium used as a model microorganism and a routine host for molecular biology and biotechnology applications. In *E. coli*, genome editing (*def.* inducing site-specific chromosome modification), is used for many biotechnological purposes, including metabolic engineering, strain optimization, and general molecular biology research.

*E. coli* has an active endogenous homologous recombination repair system, therefore, chromosomal integration can be achieved simply by donating DNAs with long (~1,000 bp) homology arms to recombine at the target insertion site. However, this method is inefficient, limiting its use (260). Fortunately, phage-derived proteins can be exploited to assist microbial genome editing, dramatically increasing the efficiency of chromosomal integration (261). However, traditional genome editing methods using the phage-derived lambda red system remain laborious and require the screening of several colonies (165).

More recently it was proposed that combining phage-derived lambda red proteins with the CRISPR/Cas selection could greatly simplify the genome editing process in bacteria (262). In

this case, CRISPR-associated (Cas) endonuclease (e.g. Cas9) induces Double Strand Breaks (DSBs) that kill those cells that fail to recombine with the donor DNA, avoiding the need for selectable markers.

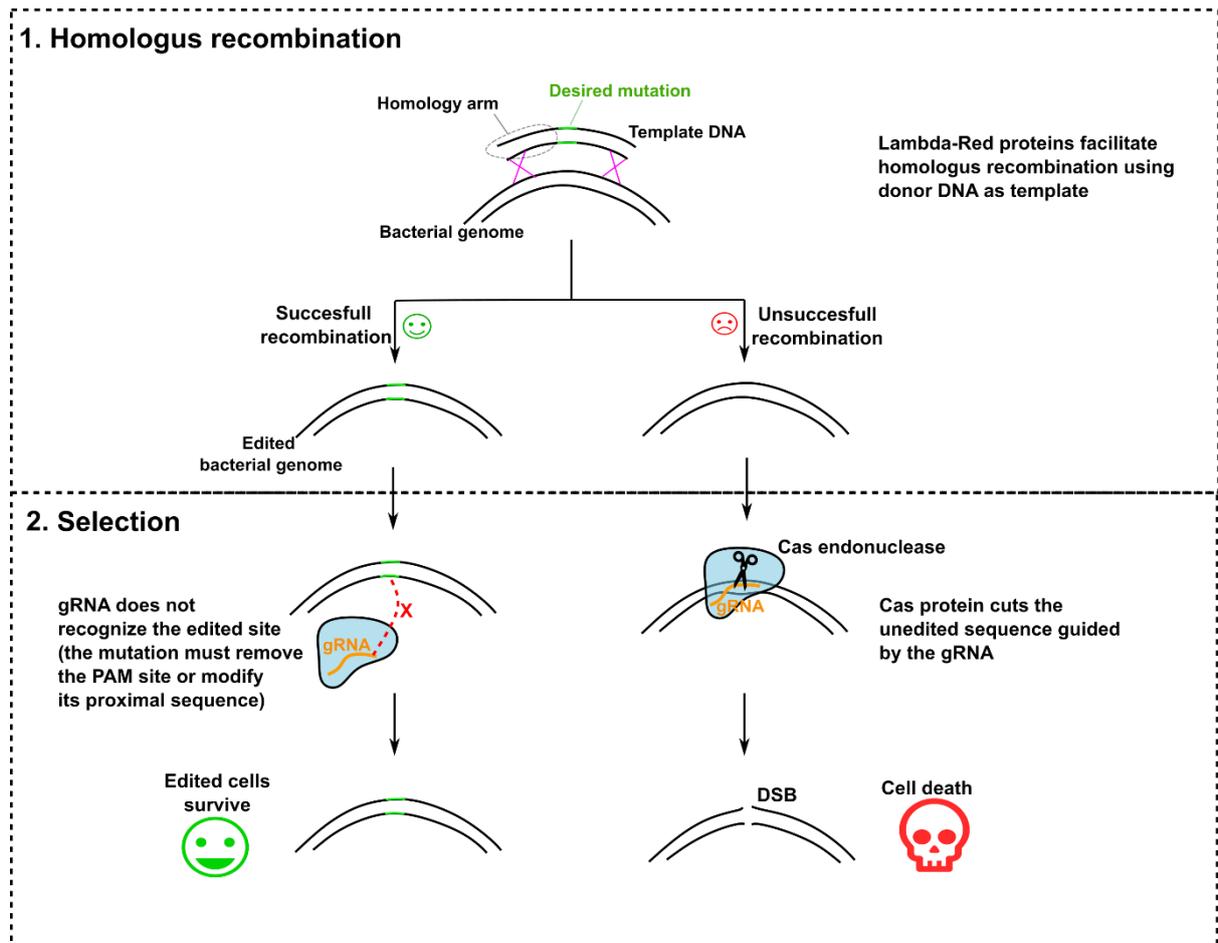
#### **4.1.2 CRISPR/Cas-assisted bacterial genome editing**

CRISPR/Cas-assisted genome editing provides a simpler and quicker method for markerless and scarless gene editing in bacteria. Generally, this method is significantly more efficient and amenable to multiplexing than traditional methods based purely on the Lambda-red recombination systems (263).

In general terms, CRISPR/Cas-assisted genome editing is constituted of two main phases (Fig.4.1):

1. The phage-derived lambda red protein assists homologous recombination with a template DNA sequence containing homology arms. Successful recombination leads to the desired chromosomal mutation.
2. CRISPR-associated (Cas) endonuclease is guided by a gRNA to the unedited DNA sequence, where it creates a double-stranded break (DSB). The created DSB is fatal to bacteria as they lack DSB repair mechanisms and therefore only cells that successfully recombined survive the selection.

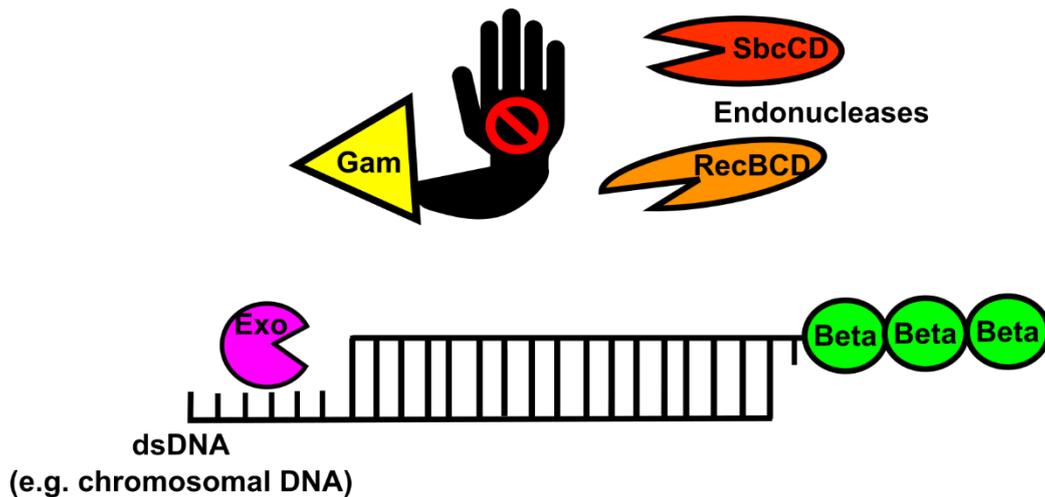
Among the various proposed CRISPR/Cas-assisted bacteria genome editing techniques, I have chosen to use a recently developed one by Sheng Yang *et al.* (in 2020) (163). This technology seemed the most optimized and simplified for genome editing in Enterobacteriaceae species, and, based on my experience, it showed very high efficiency and simple plasmid curing.



**Figure 4.1** General schematic of CRISPR/Cas-assisted genome editing in bacteria. The method is constituted of two main phases: homologous recombination facilitated by the lambda-red proteins and selection using a Cas endonuclease.

### 4.1.3 Lambda-Red homologous recombination

The first step in CRISPR/Cas genome editing is classical homologous recombination facilitated by the phage-derived Lambda-Red proteins (Exo, Beta, and Gamma). These three proteins have different functions in the recombination process (Fig.4.2): Gam prevents endogenous RecBCD and SbcCD nucleases from digesting linear DNA introduced into *E. coli*. Exo is an exonuclease that degrades linear dsDNA starting from the 5' end. Beta protects the ssDNA created by Exo and promotes its annealing to a complementary ssDNA target in the cell (264).



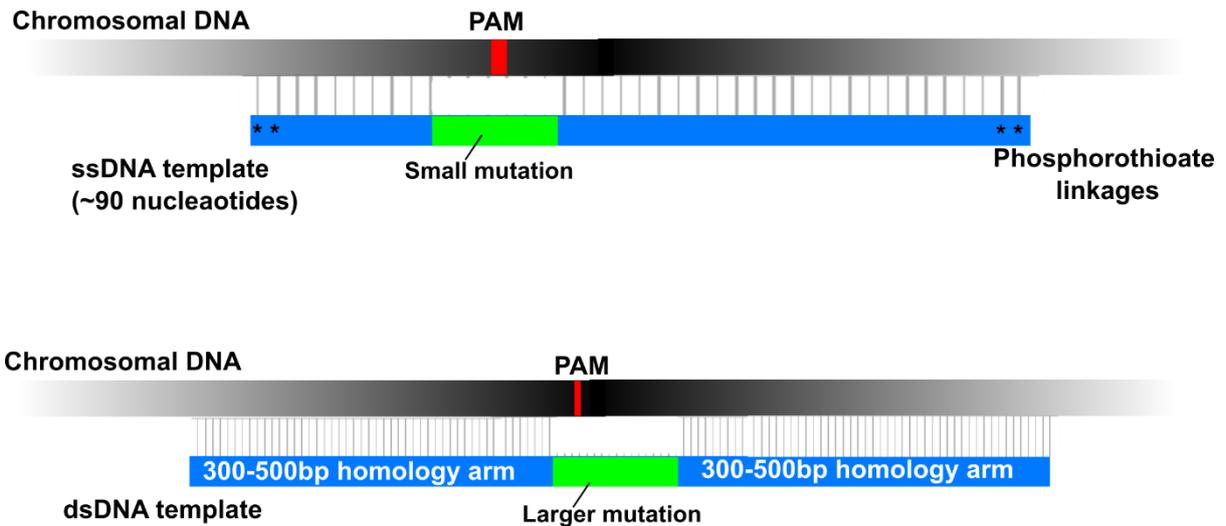
**Figure 4.2** Components of the Lambda Red Recombineering System. Figure inspired by (264).

All three proteins are required for recombineering with a dsDNA template; however, only Beta is essential when generating a modification with an ssDNA template (265).

#### 4.1.4 Templates used for homologous recombination

Homologous recombination requires a DNA template (Fig.4.1). The template DNA can be single-stranded (ssDNA) or double-stranded (dsDNA). ssDNA templates are only used for small deletions (<1 kb) or replacements (<20 bp), while dsDNA templates are used for insertions, and larger deletions (>1 kb) or replacements (>20 bp) (266). Cas9-mediated Lambda-red recombineering has been shown to successfully delete regions up to 12 kb (266, 267).

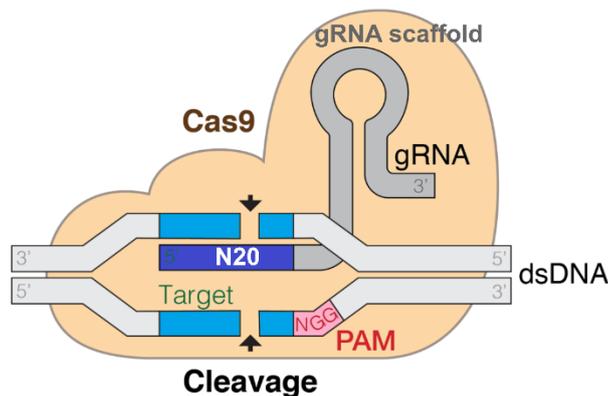
Both ssDNA and dsDNA templates (Fig. 4.3) have homology arms; typically ssDNA templates consist of ~90 nucleotides (including the desired mutation) with phosphorothioate linkages to the 5' and 3' termini to protect the ssDNA from nuclease degradation upon transformation. dsDNA templates have 300–500 bp homology arms on each side. In both cases, the resulting mutation must remove or mutate the PAM sequence or the protospacer sequence proximal to the PAM sequence, to prevent Cas9 from cutting (268) in the selection step (Fig.4.1).



**Figure 4.3** Schematic of ssDNA and dsDNA templates.

### 4.1.5 gRNA and Cas9 complex

A fundamental step in CRISPR/Cas genome editing is the design of the N20 portion of the gRNA sequence. The gRNA binds to the Cas protein and guides it toward the desired target sequence. The gRNA then binds to the target sequence causing the Cas protein to cut the dsDNA forming a double-stranded DNA break (DSB) as shown in Fig.4.4.



**Figure 4.4** Schematic of the Cas9 protein bound to a gRNA that cuts a dsDNA. Image adapted from Wikipedia (269). The entire gRNA is composed of a 20-nucleotide sequence (N20) and a gRNA scaffold sequence.

The target sequence must be adjacent to a protospacer adjacent motif (PAM) sequence which is essential for cleavage by the Cas nuclease (270). Each type of Cas nuclease recognizes a different PAM sequence (e.g. Cas9 only recognizes NGG sequences, where N is any

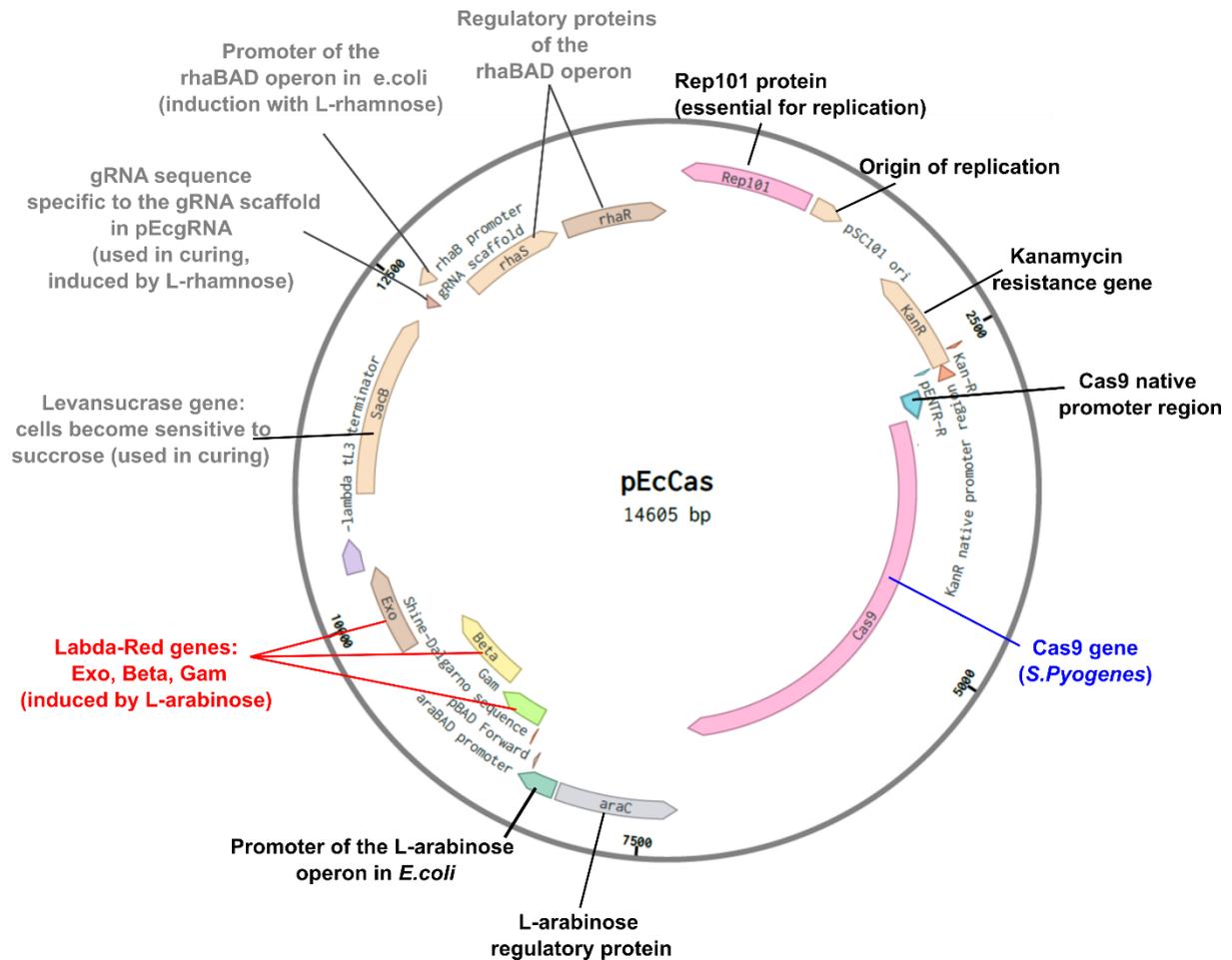
nucleotide). To design the N20 gRNA sequence multiple online tools can be used which identify all the PAM sites or, alternatively, these can be manually identified.

#### **4.1.6 pEcCas and pEcgRNA plasmids**

The CRISPR/Cas-assisted bacteria genome editing technique developed by Sheng Yang *et al.* (163) requires a template DNA sequence and two plasmids: pEcCas and pEcgRNA. The gRNA, which sequence is inserted into the pEcgRNA plasmid, and template DNA need to be custom-designed based on the desired chromosomal mutation.

##### ***pEcCas***

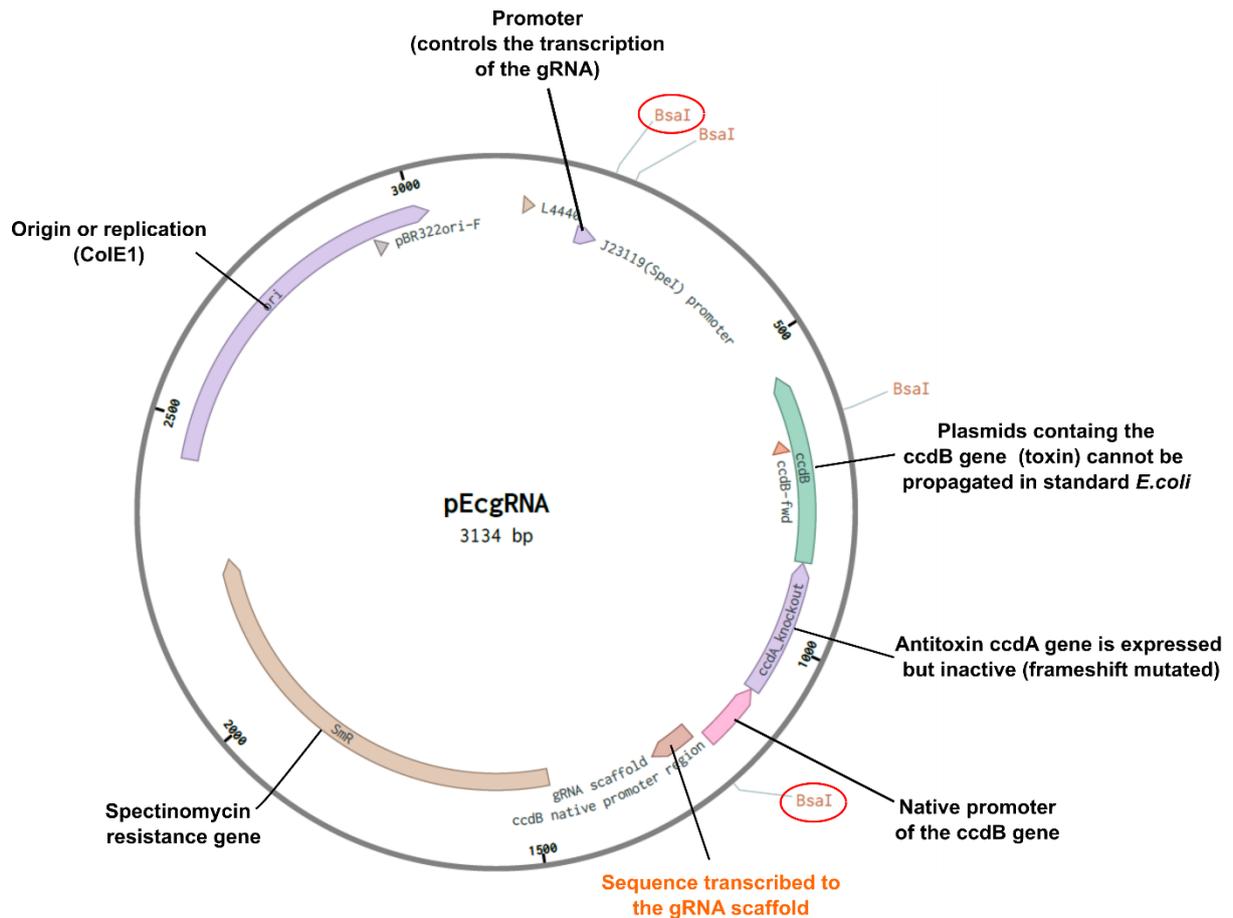
The pEcCas plasmid is a ~14,600 bp plasmid that contains a low-copy temperature-insensitive origin of replication (pSC101) (163). It also contains the genes encoding the three proteins of the lambda red recombineering system (Exo, Beta, and Gam) under the control of the araBAD promoter (induction with L-arabinose). pEcCas also contains the Cas9 gene from *Streptococcus pyogenes* under its native regulatory sequence (likely a continuous low transcription of the Cas9 gene is achieved in transformed *E. coli* cells) and other genes essential for plasmid curing. An overview of all the components of the pEcCas plasmid is provided in Fig4.5.



**Figure 4.5** pEcCas plasmid and its components. In red are the genes encoding for the lambda red proteins, in blue is the gene encoding for the Cas9 protein, and in grey are all the genes exclusively used for curing. Image obtained from importing Plasmid #73227 (Addgene collection) into Benchling online visualization software (<https://www.benchling.com/>). All genes and promoters not reported by the authors were analyzed and assigned using BLAST.

### *pEcgRNA*

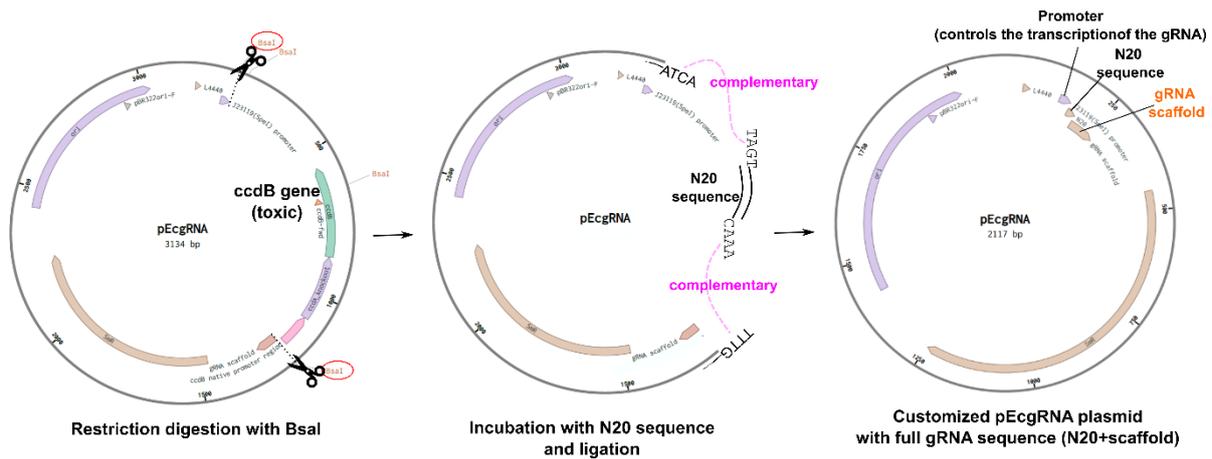
The pEcgRNA plasmid is a ~3,100 bp plasmid that contains a high-copy number origin of replication. It also contains the gRNA scaffold sequence and BsaI restriction sites. The BsaI restriction sites are used to modify the plasmid so that a custom gRNA sequence can be inserted. The gRNA transcribed from the customized pEgRNA plasmid combine with the Cas9 protein encoded in the pEcCas plasmid. Details on the components of the pEcgRNA plasmid are provided in Fig.4.6.



**Figure 4.6** pEcgRNA plasmid and its components. In orange the sequence that is transcribed into the gRNA scaffold. The BsaI sites that are used to insert the remaining customized gRNA sequence are circled in red. Image obtained by importing the Plasmid #166581 (Addgene collection) into Benchling online visualization software (<https://www.benchling.com/>). All genes and promoters not reported by the authors were analyzed and assigned using BLAST.

### Customization of pEcgRNA

As mentioned above, the pEcgRNA plasmid needs to be customized with the desired gRNA sequence before being used to transform cells. Two complementary oligos with sticky ends are annealed and then ligated to the BsaI-digested plasmid (BsaI leaves overhangs that are complementary to the sticky ends of the annealed oligos). The plasmid mixture is used to transform *E. coli* cells sensitive to the ccdB toxin (e.g. DH5 $\alpha$ ), therefore, only cells that are transformed with the modified plasmid (not containing the ccdB gene) survive. This plasmid design significantly simplifies the customization as no separation and extraction of plasmid fragments from agarose gel is required. A schematic of the customization steps is provided in Fig.4.7.

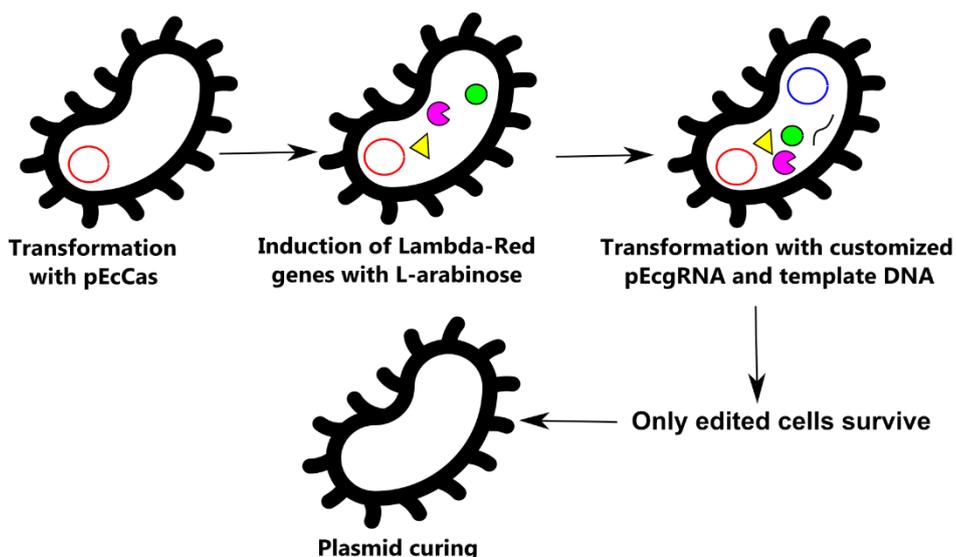


**Figure 4.7** The pEcgRNA plasmid needs to be modified with the custom gRNA sequence. First, the pEcgRNA plasmid is digested with the BsaI restriction enzyme leaving overhangs of four nucleotides on each side. Then the custom gRNA sequence 20bp (N20) with sticky ends is incubated and then ligated with the digestion products. The ligation mixture is transformed into *E. coli* cells for selection: only cells containing the customized plasmid, which does not have the ccdB gene will survive.

#### 4.1.7 Editing with the pEcCas and pEcgRNA plasmids

The main steps in the CRISPR/Cas-assisted bacteria genome editing technique developed by Sheng Yang *et al.* (163) are shown in Fig.4.8.

In the editing process, first cells are transformed with the pEcCas plasmid, and then the overexpression of lambda-red proteins is induced using L-arabinose. After induction, the cells are made electrocompetent again and used for the second and final transformation round. The second transformation is with both the customized pEcgRNA and template DNA; the cells, which already contain an overabundance of lambda-red proteins, immediately start the process of homologous recombination after receiving the template DNA. In the meantime, the Cas9-gRNA complex starts to form (Cas9 expressed from pEcCas and gRNA expressed from the customized pEcgRNA). The Cas9-gRNA complex acts as a selector by killing all the cells that were not successful in performing homologous recombination (unedited) as shown in Fig.4.1.



**Figure 4.8** Simplified schematic of the main steps in CRISPR/Cas-assisted bacteria genome editing using the pEcCas and pEcgRNA plasmids. The pEcCas and pEcgRNA are represented as a red and blue circle respectively. Lambda red proteins are represented by three different shapes (a green circle, a purple circular shape, and a yellow triangle).

#### 4.1.8 Natively expressing proteins from the genome

The maintenance of a plasmid in bacteria requires constant selection pressure which is provided by the use of antibiotics. Using antibiotics not only increases protein production costs but also can have negative effects on the environment. An alternative strategy for the production of proteins is the expression or overexpression from the genome which does not require any antibiotic usage. However, customizing and using expression plasmids is more routine and less time-consuming than genome editing, and therefore it is uncommon for research laboratories to express the desired proteins from the genome.

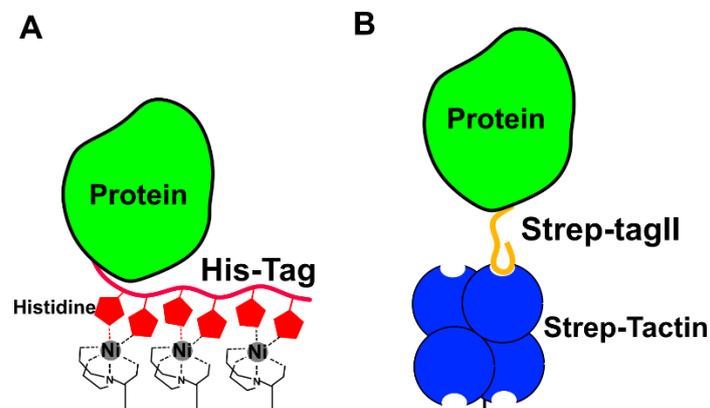
In this project, we decided to pursue a native expression (from the genome under its native regulation) of the tagged membrane-bound hydrogenase-1 from *E. coli*. This can be a sensible choice in structural studies of delicate membrane proteins as overexpression can lead to significant loss of native conformation (the protein production machinery is “stressed”). For proteins with complex cofactors, the cofactor maturation machinery can also be overwhelmed, leading to substoichiometric incorporation. These problems may be why, historically, when Hyd-1 is reported to have been successfully overexpressed using plasmids, the majority of protein was produced as inclusion bodies with low levels of hydrogenase-1 successfully translocated to the membrane (271–273).

In this chapter, genome editing was used to insert a strep tag to the N-terminus of the HyaB domain of Hyd-1 to allow native expression and purification.

#### 4.1.9 Affinity tag types

There are two commonly used affinity tags used for large-scale affinity-based protein purification (Fig.4.9):

- The polyhistidine tag (**His-tag**) consists of a chain of 6-12 histidines placed at the C- or N-terminus of a protein subunit. This tag binds tightly to the immobilized Nickel ions in an affinity column and is released using imidazole as an eluant or, less commonly, by lowering the pH or by using chelating agents.
- The **Strep-tag** is a synthetic polypeptide chain that binds specifically to engineered streptavidin proteins. There are different Strep-tags specific to their respective streptavidin variant with the most common one being the Strep-tag®II (binds to Strep-Tactin®) consisting of a chain of eight amino acids (Trp-Ser-His-Pro-Gln-Phe-Glu-Lys) placed at the C- or N-terminus of a protein subunit. To elute the protein desthiobiotin is used.



**Figure 4.9** Schematic of a His-Tagged (A) and Strep-tag®II (B) protein bound to Nickel ions and Strep-Tactin® respectively.

## 4.2 Aims of the chapter

In this result chapter, the strategy to tag Hyd-1 from *E. coli* is presented.

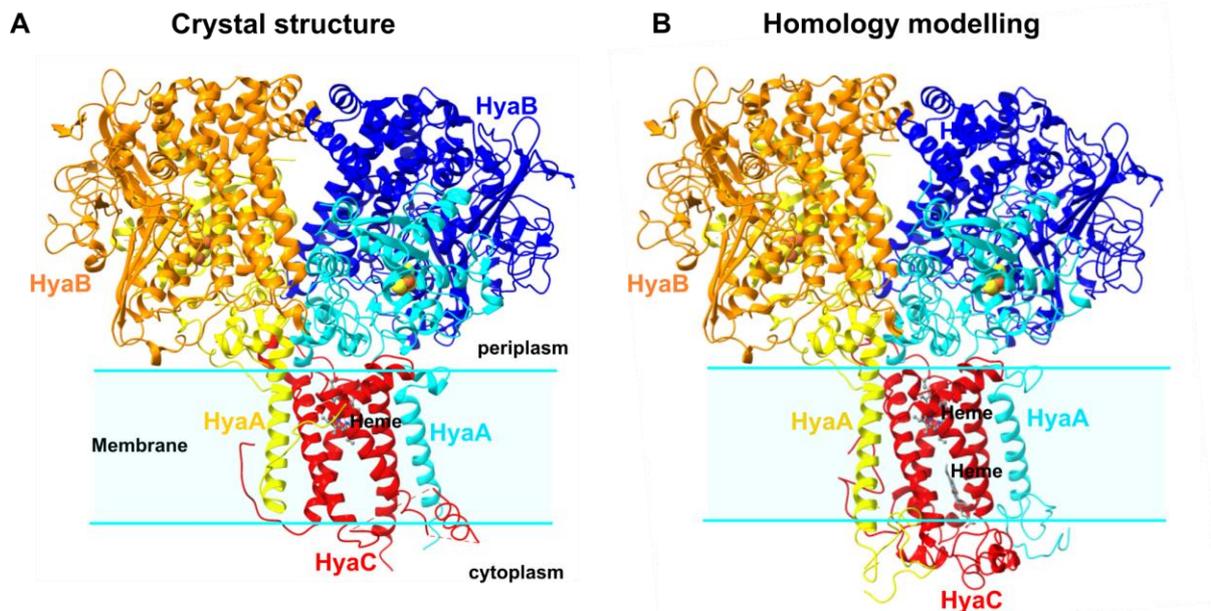
The chapter aims are:

- 1) To test the use of CRISPR-Cas assisted genome editing
- 2) To incorporate a strep-tag into the HyaB N-terminus using the above method

## 4.3 Results

### 4.3.1 Analysis of the structure of hydrogenase-1 (Hyd-1) from *E. coli*

Hydrogenase-1 (Hyd-1) is a membrane-bound hydrogenase in *E. coli* composed of three subunits (HyaA, HyaB, and HyaC). This enzyme remains anchored to the membrane through its cytochrome subunit (HyaC) and the  $\alpha$ -helices at the C-terminus of the HyaA subunits (Fig.4.10.A). Two crystal structures of Hyd-1 bound to the cytochrome are available (PDBID: 4GD3, 6G94), and both are obtained from purified Hyd-1 having a His-tag positioned at the C-terminus of HyaA (274, 275). The crystal structures (Fig.4.10.A) were incomplete: in both cases, the cytochrome subunit (HyaC) was missing ~50 amino acids and only possessed one of the two predicted hemes. In addition, only ~50% of the  $\alpha$ -helix at the C-terminus of the HyaA subunits was resolved. This suggests that the membrane domain was destabilized and partially denatured. It was likely that this was caused by the strong interaction between the hemes of the cytochrome and the imidazole (276) used to elute the protein, possibly exacerbated by the tag location as discussed below. To obtain an approximation of the complete Hyd-1 structure, homology modelling was used (Fig.4.10.B). This model was used to guide the affinity tag placement.



**Figure 4.10** **A.** Crystal structure of Hyd-1 from *E. coli* showing the subunit arrangements. PDBID: 4GD3 (274). **B.** Approximation of the complete Hyd-1 structure. Obtained by combining information from the available crystallographic structures and homology modelling of the subunits (Phyre2 server). The position of the second heme was proposed based on the homologous cytochrome subunit of the formate dehydrogenase from *E. coli* (PDBID: 1KQF (4)).

### 4.3.2 Choosing the affinity tag type and location

#### *Affinity tag type*

I have decided to use a Strep-tag, specifically the Strep-tag®II. This choice is based on the fact that the His-tag requires the use of imidazole as an eluant which is known to strongly interact with the hemes of the cytochrome (276). In addition, previous purification results of an already available His-tagged Hyd-1 (strain LAF-003 (165)) indicated a partial loss of the cytochrome subunit (details in Chapter 5). Eluting the available His-tagged Hyd-1 by changing pH was not attempted as this method is less desirable as it works by stripping off the Ni<sup>2+</sup> ions of the affinity column. In addition, I did not agree with the tag position (Fig.4.11.B) so I decided to start over by making my own strain instead of spending significant time tweaking elution conditions.

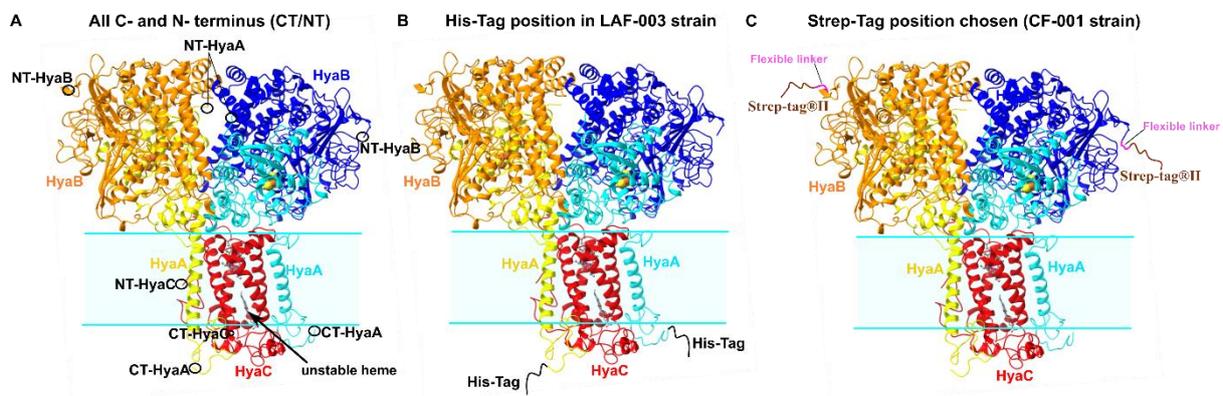
#### *Affinity tag position*

The affinity tag can be placed at the N- or C- terminus of a protein subunit, however, careful considerations should be applied:

- 1) The chosen N- or C- terminus must be ‘free’ (not buried inside the protein) allowing for the tag to stick out and interact with the ligands in the affinity column.
- 2) The tag position must not disrupt the protein maturation and transportation process.
- 3) The tag should not be placed in proximity of less stable protein regions, as this may destabilize them further (i.e. steric and electrostatic interactions).

In this case, I used the already available crystal structure of Hyd-1 (Fig.4.10.A) and the homology-based approximation of the complete structure (Fig.4.10.B) to predict the best position for the affinity tag (only homology models can be used if no experimental structure is available). All the C- and N-terminus available in Hyd-1 are shown in Fig.11.A.

Among all the possibilities (Table 4.1) I have chosen to place the affinity tag at the N-terminus of HyaB (Fig. 4.11.C). This position is the only one that satisfies all the three conditions described above.



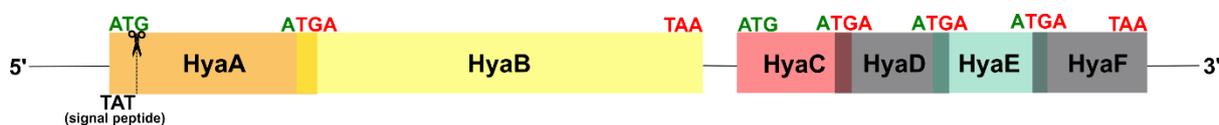
**Figure 4.11. A.** Complete Hyd-1 structure (as in Fig. 4.10.B) showing all the C- and N- termini of all the subunits. The C-terminus of HyaA, and the C- and N- N-terminus of HyaC are not experimentally validated and should be taken as an approximation (based on homology modelling). The C-terminus of HyaB is not visible as it is buried inside the protein. **B.** Position of the His-Tag in the strain LAF-003 (C-terminus of HyaA) (165). **C.** The tag position that was chosen (N-terminus of HyaB) for this project.

**Table 4.1** All potential positions for an affinity tag in Hyd-1.

Position	Comments
N-terminus HyaA	Bad position: must be left free for the TAT signal peptide (see chapter 4.3.3)
C-terminus HyaA	Potentially a destabilizing position: too close to the cytochrome and membrane domain (electrostatic and steric interaction between the tag, the membrane domain, and the detergent used for purification)
N-terminus HyaB	<b>Good position:</b> not buried and far from the delicate membrane domain
C-terminus HyaB	Bad position: this terminus is buried and undergoes cleavage during maturation by the endopeptidase HyaD (277)
N-terminus HyaC	Potentially a destabilizing position: the delicate membrane domain should not have any tag interference to avoid potential destabilization
C-terminus HyaC	Potentially a destabilizing position: the delicate membrane domain should not have any tag interference to avoid potential destabilization

### 4.3.3 Analysis of the *Hya* operon

The *HyaABCDEF* operon contains genes for the three Hyd-1 structural subunits (*HyaA*, *HyaB*, *HyaC*) and three additional maturation genes (*HyaD*, *HyaE*, *HyaF*) (272). The analysis of the operon sequence (Fig.4.12) reveals an overlap between the sequences corresponding to the start (AUG) and the stop codons (UGA) of adjacent genes (the overlap: 5'... ATGA...3'). The overlap of genes has been proposed to be involved in the regulation of gene expression by coupling the translation of genes (278).



**Figure 4.12.** *HyaABCDEF* operon showing the start (green) and stop (red) codons. The cleavable TAT signal peptide is shown as well.

In addition, the sequence encoding for an N-terminus twin-arginine translocation (TAT (279)) peptide is reported in *HyaA* (280) (note: the presence of the signal peptide can also be found by using online prediction servers). This signal peptide is essential for the transmembrane

transport of folded proteins as it targets proteins through the membrane-embedded twin-arginine translocation (Tat) complex.

The DNA sequence of the operon is essential to the design of the donor DNA. The full operon sequence from *E. coli* K12 MG1655 is shown in the appendix (Sequences section).

#### 4.3.4 N-terminus Strep-tag®II HyaB sequence

I decided to opt for Strep-tag®II connected with a flexible linker (stretches of Gly and Ser residues (281)) to the N-terminus of the HyaB subunit (Fig.4.11.C) designed in such a way that gene expression coupling is preserved. The amino acid sequences of the Strep-tag®II (WSHPQFEK (282)) and the flexible linker (GSSG) have been translated into their respective DNA sequence using codon optimization. The resulting codon-optimized DNA sequence of the Strep-tag®II (brown) connected to the flexible linker (purple) is:

5' - **TGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGT** - 3'

The above sequence needs to be inserted into the correct position and to do so the portion of DNA between the end of *HyaA* (translated to the C-terminus of HyaA) and the beginning of the *HyaB* (translated to the N-terminus of HyaB) genes is required (shown below).

The portion of the native sequence showing the overlap where *HyaA* stops (**TGA**) and *HyaB* starts (**ATG**):

5' ...TCAGCCAGGCAATGAGGATAAACAGGC**ATG**AGCACTCAGTACGAAACTCAGGGATACACCATCAATAAT...3'

As I wanted to preserve gene expression coupling, I did not remove the start/stop codon overlaps when inserting the Strep-tag®II-linker sequence. The resulting sequence that is translated into a tagged HyaB subunit is:

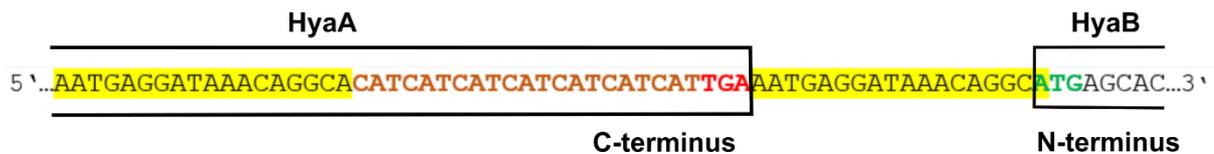
5' ...GAGGATAAACAGGC**ATGAGCTGGT**CACACCCCAATTTGAAAAAGGCTCCAGCGGTACTCAGTACGAAA...3'

Note that the 'GC' (bold) in the sequence above was left to avoid a frameshift in the translation. Based on the above sequence the HyaB subunit is translated to: **MSWSHPQFEKGSSGTQYE...**

Likely, the extra two amino acids (M and S) before the Strep-tag®II do not influence its ability to bind to Strep-Tactin® as can be deduced from analysis of PDB structures (PDBID: 1RSU (283)).

### 4.3.5 Comparison with the C-terminus His-tagged *HyaA* sequence

The already available LAF-003 strain, previously obtained by using the Red®/ET® recombination was prepared by Dr. Lindsey Flanagan using a different approach than mine (165). This strain has a polyhistidine tag at the C-terminus of *HyaC*, does not have a flexible linker, and does not preserve gene expression coupling. The edited sequence between the *HyaA* and *HyaB* genes (extrapolated from Fig.3.7 in Lindsey's thesis (165)) is shown below:



**Figure 4.13.** Edited sequence in LAF-003 strain. This shows a seven polyhistidine tag (brown) at the CT of *HyaA*. The 20 bp duplicated from the terminus of *HyaA* is highlighted in yellow.

In the above approach, seven histidines (**CAT** => His) immediately follow the C-terminus sequence of *HyaA* and the transcription terminates with a **TGA** codon after the polyhistidine tag. The **TGA** codon is immediately followed by 20 bp duplicated from the terminus of *HyaA* (highlighted in yellow) which prevents gene expression coupling of *HyaA* and *HyaB*. Based on Dr. Lindsey's thesis it appears that when the 20 bp duplicate is not used there is negligible Hyd-1 expression, however, as no sequence is provided it is not possible to understand the exact approach used in the unsuccessful attempt (i.e. ...**CATTGAATG**AGCA... or ...**CATATGA**AGCA... etc.). It is however expected that the upstream sequence to the start codon has an important role as a ribosome binding site, therefore explaining the negligible Hyd-1 expression.

### 4.3.6 Design of the gRNA

In Chapter 4.3.4 I have shown what the desired edited sequence should look like. This is replicated below (the sequences highlighted in cyan and yellow correspond to the chosen PAM site and its N20 recognition sequence):

```
5' ...TATAACGCCTGTTCTCCACACGCTGGAATGATGGCGTTTCTTTCCCAATCCAGTCTGGTCACGGCTGCCTG
GGCTGTGCGGAAAATGGTTTCTGGGATCGCGGTTCTGTTCTACAGCCGCGTGGTTCGATATTTCCGCAAATGGGTACT
CATTCCACCGCCGATACCGTTCGGTTTAAACCGCGCTTGGCGTGGTGGCAGCGGCTGTTGGTGTGCACGCAGTCGCC
AGCGCCGTTGACCAGCGCAGACGTCATAACCAGCAACCTACAGAAAACCGAACATCAGCCAGGCAATGAGGATAAAA
CAGGCATGAGCTGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGTACTCAGTACGAAACTCAGGGATACACCA
TCAATAATGCCGGACGCCGCTGGTGGTTCGACCCGATTACCGCATCGAAGGCCACATGCGCTGCCAAGTGAATA
TTAACGATCAGAATGTGATCACCAATGCCGTCTCCTGCGGCACCATCTTTTCGCGGGCTGGAGATCATCTACAAG
```

GGCGCGACCCGCGCGATGCGCGGGCGTTTCGTTGAACGTATCTGCGGCGTCTGTACTGGCGTACACGCCCTGGCTT  
CGGTTTACGCCATCGAAGATGCTATCGGTATTAAAGTGCCGGACAAC...3'

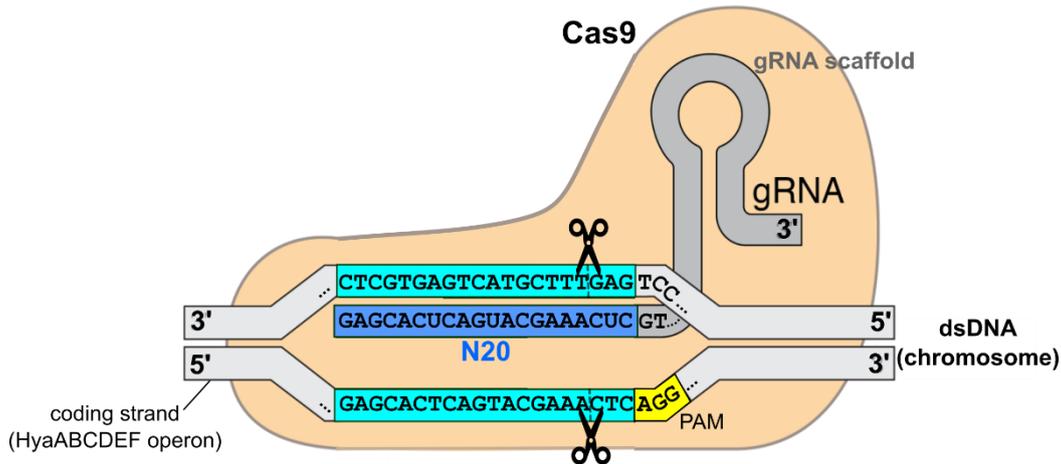
For comparison, the unedited sequence is:

5' ...TATAACGCCTGTTCTCCACACGCTGGAATGATGGCGTTTCTTTCCCAATCCAGTCTGGTCACGGCTGCCTG  
GGCTGTGCGGAAAATGGTTTCTGGGATCGCGGTTTCGTTCTACAGCCGCGTGGTCGATATTCCGCAAATGGGTACT  
CATTCCACCGCCGATAACCGTCGGTTTAACCGCGCTTGGCGTGGTGGCAGCGGCTGTTGGTGTGCACGCAGTCGCC  
AGCGCCGTTGACCAGCGCAGACGTCATAACCAGCAACCTACAGAAAACCGAACATCAGCCAGGCAATGAGGATAAA  
CAGGCATGAGCACTCAGTACGAAACTCAGG GATACACCATCAATAATGCCGGACGCCGCTGGTGGTCGACCCGA  
TTACGCGCATCGAAGGCCACATGCGCTGCGAAGTGAATATTAACGATCAGAAATGTGATCACCAATGCCGTCTCCT  
GCGGCACCATCTTTTCGCGGGCTGGAGATCATCCTACAAGGGCGCGACCCGCGCGATGCGCGGGCGTTTCGTTGAAC  
GTATCTGCGGCGTCTGTACTGGCGTACACGCCCTGGCTTCGGTTTACGCCATCGAAGATGCTATCGGTATTAAAG  
TGCCGGACAAC...3'

The above sequences show the ~300 nucleotides before and after the strep tag-linker insertion. Within these sequences, there are many potential PAM sites specific to the Cas9 protein (any NGG, with N = A, T, C, or G). Among the numerous PAM sites, I have decided to choose\* the one in yellow (AGG) as it is close to the insertion which on its own modifies the affinity for the Cas9-gRNA complex in that site; in fact, the insertion changes the distal nucleotides (far from the PAM sequence) of the N20 recognition sequence (cyan). However, distal nucleotide modification is not always sufficient to ensure that the Cas9-gRNA complex only recognizes the unedited sequence and the most reliable approach is that of removing the PAM sequence or changing the nucleotides in its proximity. This means that the edited sequence should be slightly modified (this will be covered in the next subchapter).

Based on my PAM site choice, the corresponding N20 sequence (to understand how it works see Fig.4.14) is:

5'-GAGCACUCAGUACGAAACUC-3'



**Figure 4.14.** Schematic of the Cas9 protein bound to the designed gRNA that cuts near the selected PAM site (AGG, yellow) in HyaB. The dsDNA sequence here represented corresponds to that of the unedited strain. The dark blue sequence is the N20 gRNA sequence.

In order to produce the gRNA with the desired N20 sequence, the pEcgRNA plasmid needs to be modified as in Fig.4.7. Therefore the ordered primer sequences are:

5' -TAGTGAGCACTCAGTACGAAACTC-3'  
 5' -AAACGAGTTTCGTACTGAGTGCTC-3'

Both primer sequences did not have any strong secondary structure (important to check before ordering) and when annealed they form the following dsDNA fragment with sticky ends (**bold**):

**TAGT**GAGCACTCAGTACGAAACTC  
 CTCGTGAGTCATGCTTTGAG**CAAA**

As shown in Fig.4.7, the above dsDNA fragment is ligated with the linearized pEgRNA plasmid which already contains the sequence that is transcribed into the gRNA scaffold.

*\* Note: many other choices of PAM sites are equally valid (for example PAM: CGG has a corresponding N20 sequence: GGATACACCATCAATAATGC, etc.).*

### 4.3.7 Design of the template DNA

In this project, a 36-nucleotide sequence (**TGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGT**) needs to be inserted into the genome. For insertions, a dsDNA template is recommended (267).

The dsDNA template must contain the insertion and eliminate the PAM sequence (in alternative modify its proximal sequence) to make it unrecognizable to the Cas9-gRNA complex (Fig.4.15). This ensures that only edited cells survive the Cas9 selection.

In this case, I have decided to both eliminate the PAM sequence (by nucleotide mutations) and mutate two other nucleotides in the N20 sequence (mutations: **bold underlined**). Note that the extra single nucleotide mutations in the N20 sequence are not necessary as eliminating the PAM sequence is sufficient on its own (can be avoided next time). The resulting chosen edited sequence is:

```
5' ...TATAACGCCTGTTCCCTCCACACGCTGGAATGATGGCGTTTCTTTCCCAATCCAGTCTGGTCACGG
CTGCCTGGGCTGTGCGGAAAATGGTTTCTGGGATCGCGGTTTCGTTCTACAGCCGCGTGGTTCGATATTC
CGCAAATGGGTACTCATTCCACCGCCGATAACCGTTCGGTTTAAACCGCGCTTGGCGTGGTGGCAGCGGCT
GTTGGTGTGCACGCAGTCGCCAGCGCCGTTGACCAGCGCAGACGTCATAACCAGCAACCTACAGAAAC
CGAACATCAGCCAGGCAATGAGGATAAACAGGCATGAGCTGGTCACACCCCAATTTGAAAAAGGCTC
CAGCGTACTCAGTATGAAACCCAAAGTTACACCATCAATAATGCCGGACGCCGCTGGTGGTTCGACC
CGATTACGCGCATCGAAGGCCACATGCGCTGCGAAGTGAATATTAACGATCAGAATGTGATCACCAAT
GCCGTCTCCTGCGGCACCATCTTTCGCGGGCTGGAGATCATCCTACAAGGGCGCGACCCGCGCGATGC
GCGGGCGTTTCGTTGAACGTATCTGCGGCGTCTGTACTGGCGTACACGCCCTGGCTTCGGTTTACGCCA
TCGAAGATGCTATCGGTATTAAGTGCCGGACAAC...3'
```

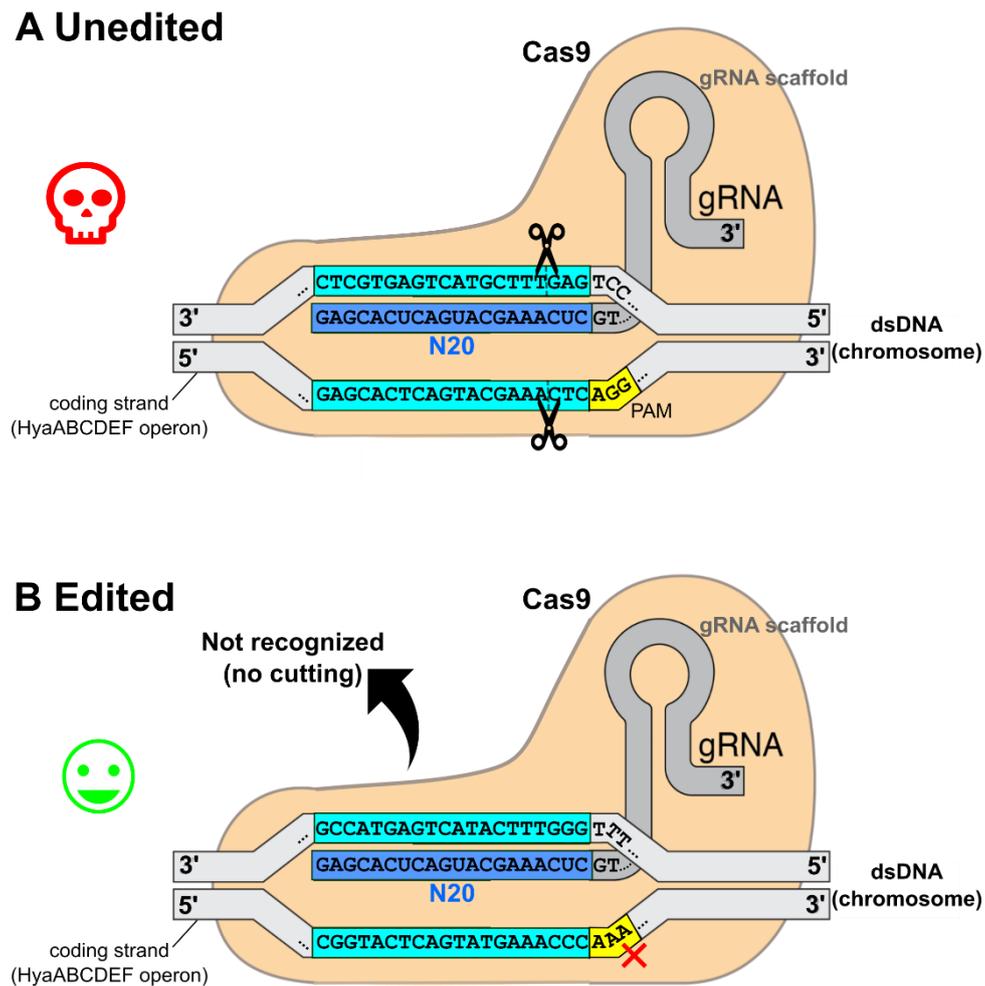
To produce this edited sequence a dsDNA with 300-500 bp homology arms is required. As the insertion is relatively small (36 bp) I proceeded with ~300 bp homology arms. The dsDNA template ordered (as Gene-block) was:

```
TATAACGCCTGTTCCCTCCACACGCTGGAATGATGGCGTTTCTTTCCCAATCCAGTCTGGTCACGGCTG
CCTGGGCTGTGCGGAAAATGGTTTCTGGGATCGCGGTTTCGTTCTACAGCCGCGTGGTTCGATATTCGCG
AAATGGGTACTCATTCCACCGCCGATAACCGTTCGGTTTAAACCGCGCTTGGCGTGGTGGCAGCGGCTGTT
GGTGTGCACGCAGTCGCCAGCGCCGTTGACCAGCGCAGACGTCATAACCAGCAACCTACAGAAACCGA
ACATCAGCCAGGCAATGAGGATAAACAGGCATGAGCTGGTCACACCCCAATTTGAAAAAGGCTCCAG
CGGTACTCAGTATGAAACCCAAAGTTACACCATCAATAATGCCGGACGCCGCTGGTGGTTCGACCCGA
TTACGCGCATCGAAGGCCACATGCGCTGCGAAGTGAATATTAACGATCAGAATGTGATCACCAATGCC
GTCTCCTGCGGCACCATCTTTCGCGGGCTGGAGATCATCCTACAAGGGCGCGACCCGCGCGATGCGCG
GGCGTTTCGTTGAACGTATCTGCGGCGTCTGTACTGGCGTACACGCCCTGGCTTCGGTTTACGCCATCG
AAGATGCTATCGGTATTAAGTGCCGGACAAC
```

Using this dsDNA template the resulting translated HydB subunit is:

```
Native HyaB    MSTQYETQGYT...
Edited HyaB    MSWSHPQFEKGSSGTQYETQSYT...
```

Other than the addition of the tag and linker there is a mutation where a Gly (**G**) has been swapped with a Serine (**S**). This mutation was inevitable for the elimination of the chosen PAM site and should not cause any change in protein function and structure (both are small amino acids at flexible 'sticking out' ends of the HyaB subunit).

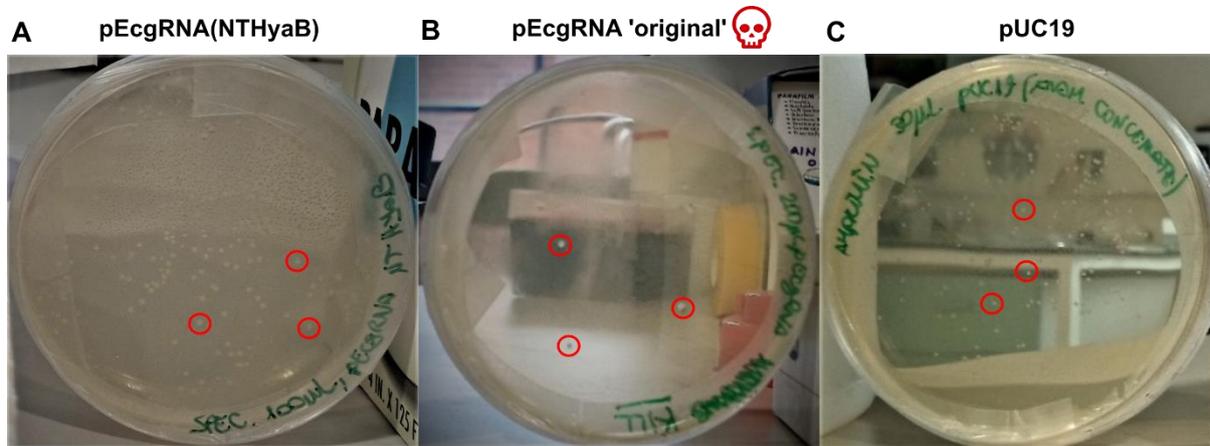


**Figure 4.15. A.** Schematic of the Cas9 protein bound to the designed gRNA that cuts near the selected PAM site (AGG, yellow) in *HyaB*. The dsDNA sequence here represented corresponds to that of the unedited strain. The DSB created kills the unedited cells. **B.** The same gRNA-Cas9 complex is unable to recognize the corresponding region in the edited strain (the PAM site AGG is muted to AAA) and therefore no cutting can occur. The edited strain survives.

#### 4.3.8 Transformation of *E. coli* DH5 $\alpha$ with pEcgRNA(NTHyaB)

The pEcgRNA was customized with the N20 sequence as described in Chapter 2.2 (Methods) and Fig.4.7. The customized pEcgRNA will be here called pEcgRNA(NTHyaB) and is characterized by not having the *ccdB* gene which protein product is toxic to most *E. coli* lab strains, including DH5 $\alpha$  (here used). This is a useful property as it ensures that only *E. coli* DH5 $\alpha$  cells that are transformed with the pEcgRNA(NTHyaB) plasmid survive while cells that receive the remaining plasmids that did not incorporate the N20 sequence do not survive.

*E. coli* DH5 $\alpha$  cells were separately transformed with the pEcgRNA(NTHyaB) plasmid solution, the original pEcgRNA plasmid (control experiment), and the pUC19 plasmid (control experiment). Then the cells were plated (Fig.4.16) and the colonies were counted to determine transformation efficiency (Table 4.2).



**Figure 4.16.** Pictures of the selection plates prepared with the revived cell suspension from the three transformation reactions. Some colonies are shown inside red circles (illustrative purposes) **A.** Spectinomycin plate prepared with 100  $\mu$ L of resuspended cells from the pEcgRNA(NTHyaB) transformation **B.** Spectinomycin plate prepared with 200  $\mu$ L of resuspended cells from the ‘original’ pEcgRNA transformation. **C.** Ampicillin plate prepared with 30  $\mu$ L of resuspended cells from the pUC19 transformation.

**Table 4.2** Transformation efficiency (TE) of each transformation reaction. The TE is calculated as the number of colonies produced divided by the  $\mu$ g of plasmid used for the transformation (#colonies/ $\mu$ g plasmid). Note that the amount of plasmid is approximated with the DNA concentration measured with a nanodrop spectrophotometer.

Transformation plasmid	Amount plated	#colonies	Transformation efficiency (TE)	Average TE
pEcgRNA(NTHyaB)	20 $\mu$ L from reaction	21	$2.2 \cdot 10^3$	$3.6 \cdot 10^3$ CFU/ $\mu$ g
	50 $\mu$ L from reaction	103	$4.3 \cdot 10^3$	
	100 $\mu$ L from reaction	~200	$4.2 \cdot 10^3$	
pEcgRNA	50 $\mu$ L from reaction	4	169	111 CFU/ $\mu$ g (Survival)
	200 $\mu$ L from reaction	5	53	
pUC19	30 $\mu$ L from a 1:10 dilution	7	$0.7 \cdot 10^8$	$1.3 \cdot 10^8$ CFU/ $\mu$ g
	30 $\mu$ L from reaction	190	$1.9 \cdot 10^8$	

As expected a large number of cells were successfully transformed with the pEcgRNA(NTHyaB) plasmid, however, the transformation efficiency remained significantly lower than that of the pUC19 plasmid ( $3.6 \cdot 10^3$  vs  $1.3 \cdot 10^8$  CFU/ $\mu$ g). The lower transformation efficiency compared to the control experiment with the pUC19 plasmid is not too surprising

given only a fraction of the reaction mixture contains correctly ligated pEcgRNA(NTHyaB) plasmids.

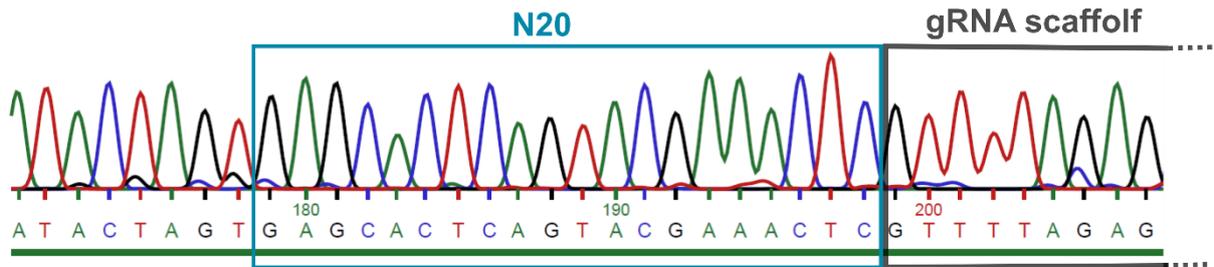
In addition, as expected, almost no cell survived the selection plate when transformed with the ‘original’ pEcgRNA plasmid (Fig.4.16.B). Based on the number of colonies that survived, it can be estimated that less than 3% of the ‘original’ pEcgRNA plasmid carried a mutation that made the *ccdB* gene ineffective (assuming the transformation efficiency for the pEcgRNA is similar to or higher than that of the pEcgRNA(NTHyaB):  $100 \cdot 111 / (3.6 \cdot 10^3) = 3\%$ ). The plasmid in one of the surviving colonies was sequenced out of curiosity.

### 4.3.9 Sequencing the pEcgRNA plasmids

Six colonies from the pEcgRNA(NTHyaB) selection plate had a portion of their plasmid DNA sequence determined. In addition, one colony that survived the ‘original’ pEcgRNA selection plate was sequenced as well to investigate the reasons for survival. The results are summarized in Table 4.3 which shows that at least three colonies from the pEcgRNA(NTHyaB) selection plate contained the pEcgRNA(NTHyaB) plasmid with the desired N20 sequence. Only the glycerol stock from colony #4 was kept and used for the next steps (equivalently colony #3, or #5 could be used). The sequencing results for colony #4 are shown in Fig.4.17.

**Table 4.3** Sequencing results from each colony picked.

Colony #	Plasmid concentration as purified	Sequencing results	Sequencing quality
pEcgRNA(NTHyaB) #1	98 ng/ $\mu$ l	TAGTGAGCACTCAGTACGAAACTC	Not great
pEcgRNA(NTHyaB) #2	111.5 ng/ $\mu$ l	n.a	Failed
pEcgRNA(NTHyaB) #3	131 ng/ $\mu$ l	TAGTGAGCACTCAGTACGAAACTC	Good
pEcgRNA(NTHyaB) #4	115.1 ng/ $\mu$ l	TAGTGAGCACTCAGTACGAAACTC	Good
pEcgRNA(NTHyaB) #5	84.6 ng/ $\mu$ l	TAGTGAGCACTCAGTACGAAACTC	Good
pEcgRNA(NTHyaB) #6	113.1 ng/ $\mu$ l	n.a	Failed
pEcgRNA ‘original’	161.3 ng/ $\mu$ l	<i>ccdB</i> NT disrupted by insertion of genomic DNA	Good



**Figure 4.17.** Snapshot of the portion of the sequencing results showing the correct incorporation of the N20 sequence. The data shown are from colony #4 from the pEcgRNA(NTHyaB) selection plate.

**Notes on the sequencing results from a colony that survived the transformation with the ‘original’ pEcgRNA plasmid.**

A few colonies from the ‘original’ pEcgRNA plasmid transformation survived. To understand how that was possible, one of the colonies was selected for plasmid extraction and sequencing. The sequencing results showed that an insertion occurred that disrupted the *ccdB* gene (Fig.4.18). The DNA inserted did not correspond to any portion of the pEcgRNA plasmid and appeared to be a portion of a putative IS1A family transposase gene (Blast search). It is possible that a portion of genomic DNA integrated into the plasmid or that the plasmid batch purchased was not pure. The sequencing results of this strange pEcgRNA plasmid can be found in the appendix.

**A. pEcgRNA**



**B. pEcgRNA with non-functional *ccdB* gene**



**Figure 4.18. A.** Schematic of the portion of the pEcgRNA plasmid containing the *ccdB* gene and the non-functional (frameshift mutated) *ccdA* gene. **B.** Schematic of the pEcgRNA plasmid with a truncated *ccdB* gene (non-functional) that was found in the cells of one of the few colonies that survived.

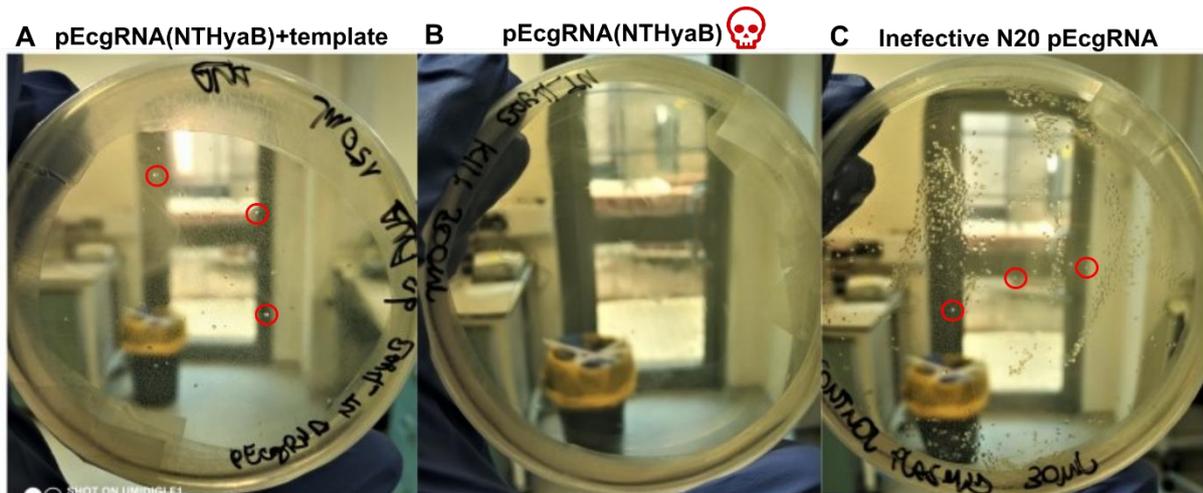
#### **4.3.10 Transformation of *E. coli* K12 MG1665 with pEcgRNA(NTHyaB) and template dsDNA**

Previously prepared electrocompetent *E. coli* K12 MG1665 cells harbouring pEcCas (with lambda-red proteins already induced) were transformed with the pEcgRNA(HyaB) plasmid and template dsDNA. The purified pEcgRNA(NTHyaB) was obtained from colony #4 (Chapter 4.3.9) while the dsDNA was purchased ready to be used as a gene block (Chapter 4.3.7).

Only edited cells (successful homologous recombination with the template DNA) can survive selection as the Cas9-gRNA complex cuts the unedited DNA in the target region forming a deadly DSB (Chapters 4.1.2 and 4.1.7, and Fig.4.15).

In addition to the main editing transformation reaction (pEcgRNA(HyaB) + dsDNA), another two control transformations were performed: in one control reaction, an aliquote of electrocompetent *E. coli* K12 MG1665 cells harbouring the pEcCas plasmid were transformed only with the pEcgRNA(HyaB) plasmid while in the other reaction, the cells were transformed only with a pEcgRNA plasmid customized with an ineffective N20 sequence (tagtGAGTTTCGTACTIONGAGTGCTC). Then the cells were plated (Fig.4.19) and the colonies were counted to determine transformation efficiency (Table 4.4).

As a note, *E. coli* K12 MG1665 cells harbouring the pEcCas plasmid that did not undergo transformation were streaked onto a Spectinomycin plate showing no growth as expected.



**Figure 4.19.** Pictures of the selection plates prepared with the revived cell suspension from the three transformation reactions. Some colonies are shown inside red circles (illustrative purposes) **A.** Spectinomycin + kanamycin plate prepared with 150  $\mu\text{L}$  of resuspended cells from the pEcgrRNA(NTHyaB) + template dsDNA transformation **B.** Spectinomycin + kanamycin plate prepared with 200  $\mu\text{L}$  of resuspended cells from the pEcgrRNA(NTHyaB) transformation. **C.** Spectinomycin + Kanamycin plate prepared with 30  $\mu\text{L}$  of resuspended cells from the transformation using the pEcgrRNA plasmid customized with an ineffective N20 sequence.

**Table 4.4** Transformation efficiency (TE) of each transformation reaction. The TE is calculated as the number of colonies produced divided by the  $\mu\text{g}$  of plasmid used for the transformation (#colonies/ $\mu\text{g}$  plasmid). Note: the average transformation efficiency with respect to the dsDNA template added is 850 CFU/ $\mu\text{g}$ .

Transformation plasmid	Amount plated	#colonies	Transformation efficiency (TE)	Average TE
pEcgrRNA(NTHyaB) + template dsDNA	20 $\mu\text{L}$ from reaction	1	303	1.0 $\cdot$ 10 <sup>3</sup> CFU/ $\mu\text{g}$
	50 $\mu\text{L}$ from reaction	9	1.1 $\cdot$ 10 <sup>3</sup>	
	150 $\mu\text{L}$ from reaction	42	1.7 $\cdot$ 10 <sup>3</sup>	
pEcgrRNA(NTHyaB)	30 $\mu\text{L}$ from reaction	1	202	77 CFU/ $\mu\text{g}$ (survival)
	100 $\mu\text{L}$ from reaction	0	0	
	200 $\mu\text{L}$ from reaction	1	30	
Ineffective pEcgrRNA	30 $\mu\text{L}$ from 1:10 dilution	215	4.9 $\cdot$ 10 <sup>5</sup>	3 $\cdot$ 10 <sup>5</sup> CFU/ $\mu\text{g}$
	30 $\mu\text{L}$ from reaction	~400	~1 $\cdot$ 10 <sup>5</sup>	

Analysis of the plate containing the cells transformed with both pEcgrRNA(NTHyaB) and template dsDNA (Fig.4.19.A) revealed several survived colonies indicating successful editing.

As expected, almost no cell survived the selection plate when transformed only with the pEcgrRNA(NTHyaB) plasmid (Fig.4.19.B) indicating that the gRNA-Cas9 complex did its job in killing the unedited cells.

In addition, cells that were transformed with the ineffective pEcgRNA survived as expected (Fig.4.19.C). This control transformation indicates that electroporation was two orders of magnitude more effective than heat shock transformation ( $3 \cdot 10^5$  vs  $3.6 \cdot 10^3$  CFU/ $\mu$ g) and the transformation with the template dsDNA ( $3 \cdot 10^5$  vs  $1.0 \cdot 10^3$  CFU/ $\mu$ g). However, the latter case only reflects the lower probability of a cell receiving both the plasmid and the template DNA\*.

*\*Here a total of 0.1 and 0.5 pmol of pEcgRNA(NTHyaB) and dsDNA, respectively, were used in the editing reaction. A change in their ratio may significantly change the transformation efficiency (the probability of receiving both plasmid and template dsDNA changes).*

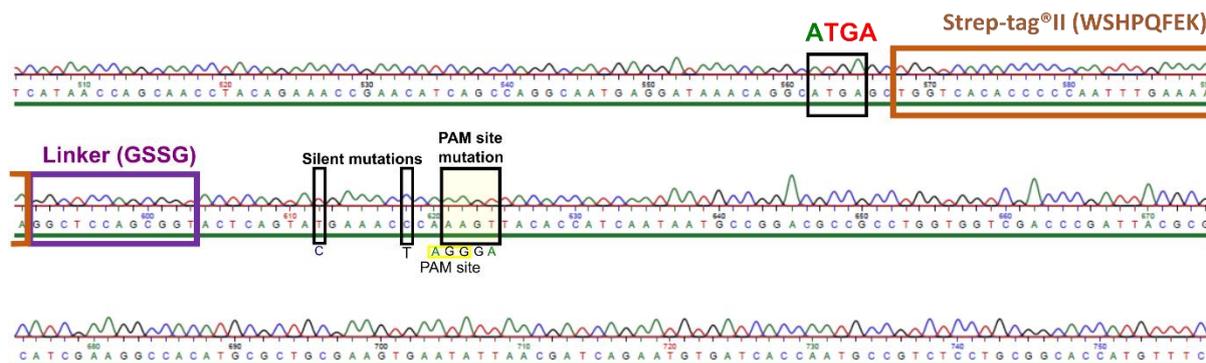
### 4.3.11 Sequencing the edited *E. coli* K12 MG1665 cells

Four colonies from the pEcgRNA(NTHyaB) + template dsDNA selection plate were sequenced. In addition, one colony of the unedited *E. coli* K12 MG1665 strain was sequenced as well (same setting and method) to confirm that its genome around the region of interest did not differ from the published database.

The results are summarized in Table 4.5 and Fig.4.21 which shows that all four colonies from the pEcgRNA(NTHyaB) + template dsDNA selection plate were correctly edited with their sequence being identical to the expected one based on the template DNA chosen. Colony #2 was kept (any other choice is equivalent) and a snapshot of its corresponding sequencing results is provided in Fig.4.20.

**Table 4.5** Sequencing results from each colony picked.

Colony #	Plasmid concentration as purified	Sequencing results	Sequencing quality
pEcgRNA(NTHyaB) + template dsDNA #1	10.4 ng/ $\mu$ l	...CTGGTCACACCCCAATTTGAAAAAGGCT CCAGCGGTACTCAGTATGAAACCCAAGT...	Good
pEcgRNA(NTHyaB) + template dsDNA #2	99.1 ng/ $\mu$ l	...CTGGTCACACCCCAATTTGAAAAAGGCT CCAGCGGTACTCAGTATGAAACCCAAGT...	Good
pEcgRNA(NTHyaB) + template dsDNA #3	109 ng/ $\mu$ l	...CTGGTCACACCCCAATTTGAAAAAGGCT CCAGCGGTACTCAGTATGAAACCCAAGT...	Good
pEcgRNA(NTHyaB) + template dsDNA #4	89 ng/ $\mu$ l	...CTGGTCACACCCCAATTTGAAAAAGGCT CCAGCGGTACTCAGTATGAAACCCAAGT...	Good
Unedited <i>E. coli</i> K12 MG1665	10.4 ng/ $\mu$ l	...CACTCAGTACGAAACTCAGGGA...	Good



**Figure 4.20.** Snapshot of the portion of the sequencing results showing the correct editing of the N-terminus of HyaB. The data shown are from colony #2 from the pEcgRNA(NTHyaB) + template dsDNA selection plate.

<i>E. coli</i> K12 MG1655 database	CGTCATAACCAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAACAGGCA	540
<i>E. coli</i> K12 MG1655 experimental	CGTCATAACCAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAACAGGCA	540
<i>E. coli</i> K12 MG1655 edited #1	CGTCATAACCAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAACAGGCA	540
<i>E. coli</i> K12 MG1655 edited #2	CGTCATAACCAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAACAGGCA	540
<i>E. coli</i> K12 MG1655 edited #3	CGTCATAACCAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAACAGGCA	540
<i>E. coli</i> K12 MG1655 edited #4	CGTCATAACCAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAACAGGCA	540
	*****	
<i>E. coli</i> K12 MG1655 database	TGAGC-----ACTCAGTACGAAACTCAGG	564
<i>E. coli</i> K12 MG1655 experimental	TGAGC-----ACTCAGTACGAAACTCAGG	564
<i>E. coli</i> K12 MG1655 edited #1	TGAGCTGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGTACTCAGTATGAAACCCAAA	600
<i>E. coli</i> K12 MG1655 edited #2	TGAGCTGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGTACTCAGTATGAAACCCAAA	600
<i>E. coli</i> K12 MG1655 edited #3	TGAGCTGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGTACTCAGTATGAAACCCAAA	600
<i>E. coli</i> K12 MG1655 edited #4	TGAGCTGGTCACACCCCAATTTGAAAAAGGCTCCAGCGGTACTCAGTATGAAACCCAAA	600
	*****	
<i>E. coli</i> K12 MG1655 database	GATACACCATCAATAATGCCGGACGCCGCCCTGGTGGTCGACCCGATTACGCGCATCGAAG	624
<i>E. coli</i> K12 MG1655 experimental	GATACACCATCAATAATGCCGGACGCCGCCCTGGTGGTCGACCCGATTACGCGCATCGAAG	624
<i>E. coli</i> K12 MG1655 edited #1	GTTACACCATCAATAATGCCGGACGCCGCCCTGGTGGTCGACCCGATTACGCGCATCGAAG	660
<i>E. coli</i> K12 MG1655 edited #2	GTTACACCATCAATAATGCCGGACGCCGCCCTGGTGGTCGACCCGATTACGCGCATCGAAG	660
<i>E. coli</i> K12 MG1655 edited #3	GTTACACCATCAATAATGCCGGACGCCGCCCTGGTGGTCGACCCGATTACGCGCATCGAAG	660
<i>E. coli</i> K12 MG1655 edited #4	GTTACACCATCAATAATGCCGGACGCCGCCCTGGTGGTCGACCCGATTACGCGCATCGAAG	660
	* *****	

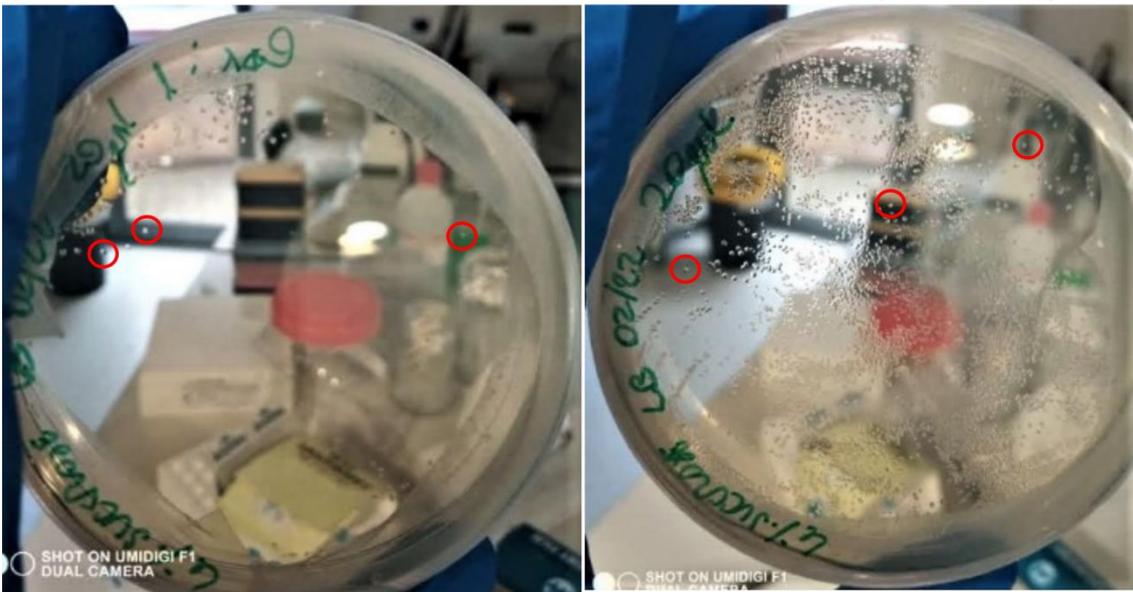
**Figure 4.21.** Sequence alignment (using Clustal Omega server) of the unedited and edited strains.

### 4.3.12 Plasmid curing

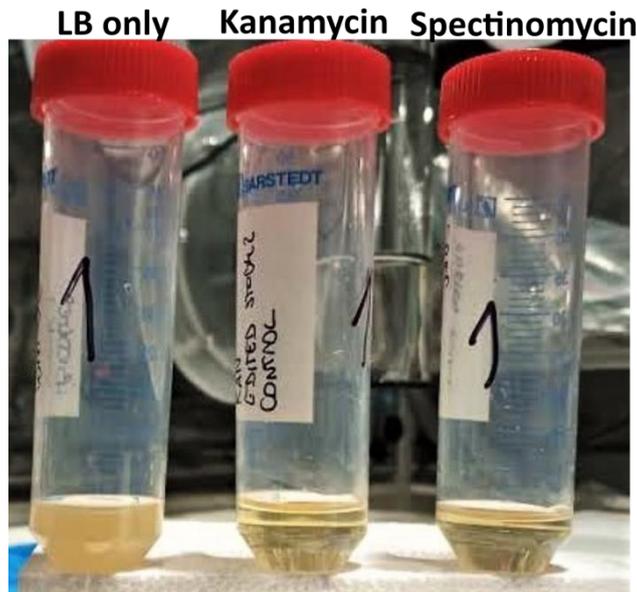
The edited colony #2 was selected for plasmid curing. The plasmid curing process consists of two steps: first L-rhamnose is used to induce the transcription of the gRNA encoded in the pEcCas plasmid which is specific to a sequence in the origin region of the pEcgRNA plasmid. This ‘curing gRNA’ combines with the Cas9 protein and creates a DSB in the pEcgRNA plasmid eliminating it with high efficiency (no selection plate used). The second step involves overnight incubation without any selection pressure (no antibiotics) during which some cells

spontaneously lose the low copy number pEcCas plasmid. 4% sucrose plates (Fig.4.22) are then used to select those cells that have eliminated the pEcCas plasmid (pEcCas contains the sacB gene which product converts sucrose to the toxin levan (284)). To confirm the elimination of both plasmids the cells derived from a single colony obtained from the sucrose plate are incubated with spectinomycin and kanamycin to check that they cannot grow (Fig.4.23). A glycerol stock of the cured colony (strain CF-001) is then stored and later used for Hyd-1 purification.

**A Sucrose, 20  $\mu$ L from 1:100 dilution    B Sucrose, 20  $\mu$ L (no dilution)**



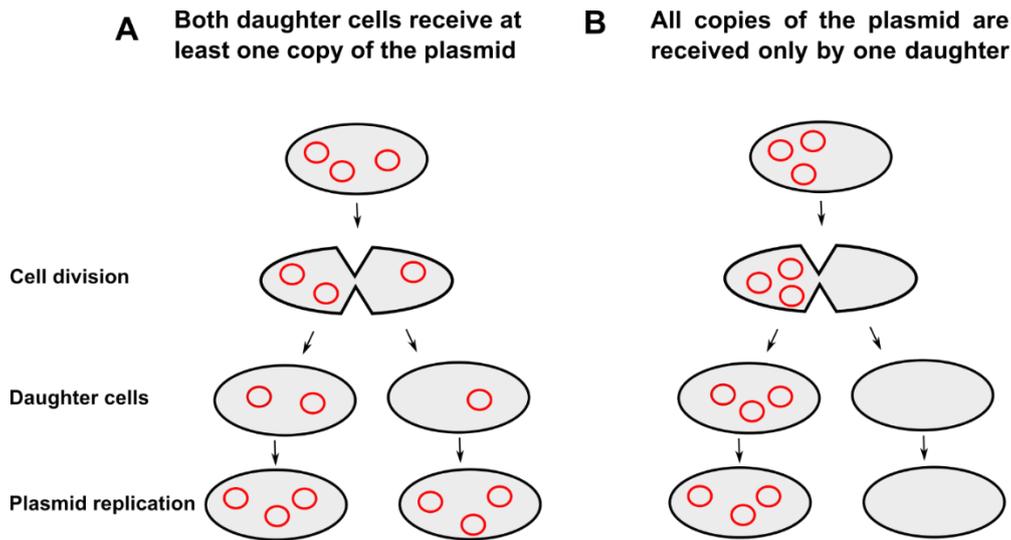
**Figure 4.22.** Pictures of the 4% sucrose selection plates prepared with the cell suspension obtained after overnight incubation in the absence of selection pressure (no antibiotics). Some colonies are shown inside red circles (illustrative purposes) **A.** Plate obtained by streaking 20  $\mu$ L from a 1:100 dilution of the cell suspension **B.** Plate obtained by streaking 20  $\mu$ L of the cell suspension.



**Figure 4.23.** Pictures of the centrifuge tubes used to confirm plasmid curing. Cells from a single colony from the 4% sucrose selection plate were incubated in LB-only medium (left) and in LB medium with kanamycin (middle) and spectinomycin (right). The cells only grew in LB-only medium (turbid) indicating successful curing of both pEcCas and pEcgRNA plasmids.

#### 4.3.13 Estimating the probability of plasmid loss

Plasmids are randomly distributed between the two daughter cells depending on their position at the time of division. However, when there are only a few copies of the plasmid (low copy number plasmid) inside each parent cell, the probability that all the copies of the plasmid are distributed only to one daughter cell becomes statistically significant (Fig.4.24.B). This random event is fundamental in many cases of plasmid curing, including the low copy number pEcCas plasmid.



**Figure 4.24.** **A** Schematic showing a parent cell dividing and distributing at least one plasmid (red) to both daughter cells. The plasmids then replicate (regulated by plasmid concentration (285)) inside each daughter cell reaching their original copy number. **B.** However, when all the plasmids inside the parent cell are distributed only around one replication pole (random distribution event) it results in one plasmid-free daughter cell.

I decided to use the data obtained from the pEcCas curing step to estimate the probability that a division event of a plasmid-containing parent cell results in a plasmid-free daughter cell ( $p$ ). Or, in equivalent terms, the average number of division events of plasmid-containing parent cells needed to produce one plasmid-free cell ( $p^{-1}$ ).

### **Estimation of $p$**

The experimental data needed to estimate  $p$  are the OD600 measured just before streaking the cells onto the sucrose selection plate and the number of colonies counted in each sucrose plate.

The measured OD600 was 0.620 (after 6h) which is approximately equivalent to  $5 \cdot 10^8$  cells/mL. Therefore 20  $\mu$ L from a 1:100 dilution of cell suspension corresponds to a total of  $10^5$  cells streaked onto each plate. Of these  $10^5$  cells an average of only ~220 cells survived (one colony is formed by a single surviving cell) meaning that ~0.22% ( $220/10^5 \cdot 100$ ) of cells lost the plasmid during the 6 h incubation period (Table 4.6).

**Table 4.6** Colonies counted in each sucrose selection plate and the respective % of plasmid loss.

4% sucrose plate selection		
Dilution	Colonies counted	% plasmid loss (based on initial OD)
20 $\mu$ L from 1:100 dilution	~140	0.14%
20 $\mu$ L from 1:100 dilution	~300	0.30%

In 6h of incubation, there is ~1h of lag phase and ~5h of exponential growth. During the exponential phase, *E. coli* cells divide every ~20 min (286), and therefore there are ~15 generations ( $n$ ) of cells (plotting the actual growth curve will give more precise estimates).

Based on this data it can be estimated (Equation 4.1, with % plasmid loss = 0.22% and  $n = 15$ ) that the probability ( $p$ ) that a duplication event results in a daughter cell not having the pEcCas plasmid is  $\sim 2.9 \cdot 10^{-4}$  (equivalent to 0.029%). This means that  $\sim 3'400$  ( $p^{-1}$ ) duplications of plasmid-containing cells are needed to produce a single plasmid-free cell.

$$\% \text{ plasmid loss} = \frac{n \cdot p}{2} \cdot 100 \quad (4.1)$$

pEcCas contains a pSC101 origin of replication which means that there are ~3-5 copies/cell (287). I have therefore shown that even with such a low copy number, the probability of all copies being distributed to only one daughter cell remains very slim.

#### ***Derivation of Equation 4.1***

Suppose that I inoculate a number of plasmid-containing cells ( $Ntot_0$ ) in fresh medium. During the log phase, the cells will duplicate at a constant rate meaning that the number of generations ( $n$ ) can be estimated. After  $n$  generations, the number of cells in the culture will have exponentially increased and its total number ( $Ntot_n = 2^n \cdot Ntot_0$ ) can be experimentally approximated by measuring the OD600. Among those cells, some will not have any plasmid, and their number ( $Y_n$ ) can be estimated by counting the colonies on a selection plate.

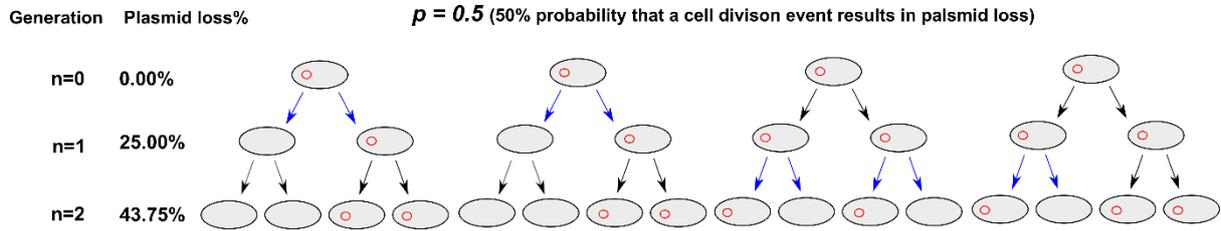
Therefore, the % of plasmid loss, which corresponds to the total number of cells that end up without a plasmid after  $n$  generations can also be expressed as:

$$\% \text{ plasmid loss} = \frac{Y_n}{Ntot_n} \cdot 100 \quad (4.2)$$

Where  $n$  is the number of generations,  $Y_n$  is the number of cells without plasmid at generation  $n$ , and  $Ntot_n$  is the total number of cells at generation  $n$ .

$Y_n$  depends on the probability that a division event of plasmid-containing cells results in a plasmid-free daughter cell ( $p$ ) and the initial number of inoculated cells ( $Ntot_0$ ). For example, suppose that I inoculate 4 plasmid-containing cells characterized by having a probability of 50% that a cell division results in a plasmid-free daughter cell (Fig.4.25); then after the first round of divisions (generation  $n = 1$ ), I expect that 2 out of 8 cells are plasmid-free ( $Y_1 = 2$ );

after another round of divisions (generation  $n = 1$ ), I expect that 7 out of 16 cells will be plasmid-free ( $Y_2 = 7$ ), and so on...



**Figure 4.25** Schematic showing how the plasmid loss % varies across generations when  $p = 0.5$ . Plasmids are represented as red circles. Division events that result in one plasmid-free daughter cell are shown as a pair of blue arrows.

In more general terms  $Y_n$  can be expressed as the sum of the cells generated from the division of the plasmid-free cells ( $2 \cdot Y_{n-1}$ ) and the cells generated from the division of plasmid-containing cells ( $p \cdot (Ntot_{n-1} - Y_{n-1})$ ):

$$Y_n = 2 \cdot Y_{n-1} + p \cdot (Ntot_{n-1} - Y_{n-1}) \quad (4.3)$$

However, when  $Y_{n-1}$  is significantly smaller than  $Ntot_{n-1}$  ( $Y_{n-1} \ll Ntot_{n-1}$ ) (expected considering that after 15 generations only 0.22% of cells lost the plasmid) it is possible to approximate Formula 4.3 as:

$$Y_n = 2 \cdot Y_{n-1} + p \cdot Ntot_{n-1} \quad (4.4)$$

By substituting  $Ntot_{n-1}$  with  $2^{n-1} \cdot Ntot_0$ , Formula 4.4 becomes:

$$Y_n = 2 \cdot Y_{n-1} + p \cdot Ntot_{n-1} = 2 \cdot Y_{n-1} + p \cdot 2^{n-1} \cdot Ntot_0 \quad (4.5)$$

By continuing the series, it can be noticed that:

$$\begin{aligned} Y_n &= 2 \cdot Y_{n-1} + p \cdot 2^{n-1} \cdot Ntot_0 = \\ &= 2 \cdot (2 \cdot Y_{n-2} + p \cdot 2^{n-2} \cdot Ntot_0) + p \cdot 2^{n-1} \cdot Ntot_0 = \\ &= 2^2 Y_{n-2} + 2 \cdot p \cdot 2^{n-1} \cdot Ntot_0 = \dots = 2^n Y_{n-n} + n \cdot p \cdot 2^{n-1} \cdot Ntot_0 \end{aligned} \quad (4.6)$$

However  $Y_{n-n} = Y_0 = 0$  (at the beginning there are no plasmid-free cells) and therefore:

$$Y_n = n \cdot p \cdot 2^{n-1} \cdot Ntot_0 \quad (4.7)$$

Substituting  $Y_n$  in Formula 4.2:

$$\% \text{ plasmid loss} = \frac{Y_n}{N_{tot_n}} \cdot 100 = \frac{n \cdot p \cdot 2^{n-1} \cdot N_{tot_0}}{2^n \cdot N_{tot_0}} \cdot 100 = \frac{n \cdot p}{2} \cdot 100 \quad (4.8)$$

Q.E.D

## 4.4 Conclusions

CRISPR-Cas assisted genome editing has been used for the first time to incorporate a Strep-tag into the *E. coli* chromosome for the production of Hydrogenase-1. This method resulted in a high-efficiency selection with all four colonies screened being successfully edited (Chapter 4.2.11) and should be considered superior to other genome editing methodologies due to its higher efficiency and reduced experimental steps.

The protocol developed can also be used to perform other mutations in the *E. coli* genome in a much shorter time; for example, a single DNA template can be used to induce several single-point mutations (within a region of max ~1000 bp) in one single editing round. Furthermore, by skipping the curing of the pEcCas plasmid, it is possible to immediately transform the edited cells with a new pEcgRNA plasmid and template DNA allowing for back-to-back editing.

This project also demonstrated for the first time that it is possible to preserve gene expression coupling in Hyd-1 by editing the genome in such a way that the overlap of the start and stop codons is preserved as demonstrated by the good purification yield of active protein obtained (details in the next chapter).

# 5 Cryo-EM studies on Hydrogenase-1 from *E. coli*

## 5.1 Introduction

This introduction focuses on the membrane-bound hydrogenase-1 (Hyd-1) from *E. coli* and is followed by the experimental results here obtained which include purification and cryo-EM data.

Some of the results have been published (170).

### 5.1.1 Hydrogenases from *E. coli*

*E. coli* grows on a wide variety of substrates in both aerobic and anaerobic conditions. In the absence of oxygen, *E. coli* produces acetate, ethanol, lactate, formate, succinate, carbon dioxide, and hydrogen (288). The concentration of external hydrogen needs to be low for anaerobic metabolism to be sustained, therefore *E. coli* often lives together with hydrogen-consuming organisms, such as methanogens or sulfate-reducing bacteria (289).

*E. coli* encodes up to four [NiFe]-hydrogenases, each of which is known, or predicted, to contribute to different aspects of cellular metabolism (290). Hydrogenase-1 (Hyd-1) and -2 (Hyd-2) are H<sub>2</sub> uptake enzymes, while Hydrogenase-3 (Hyd-3) and -4 (Hyd-4) are H<sub>2</sub>-producing enzymes, however, it has been shown that Hyd-2 and Hyd-3 revert the direction of function during glycerol fermentation to consume H<sub>2</sub> (291, 292).

This project focuses on the membrane-bound, periplasmically oriented, Hyd-1 which has been used for understanding the mechanism of the relatively rare oxygen-tolerant hydrogenases (290). Hyd-1 is likely to have a role in the anaerobic to aerobic transition (274, 293), however, its exact function still remains subject of debate.

### 5.1.2 Anaerobic regulation of the *Hya* operon

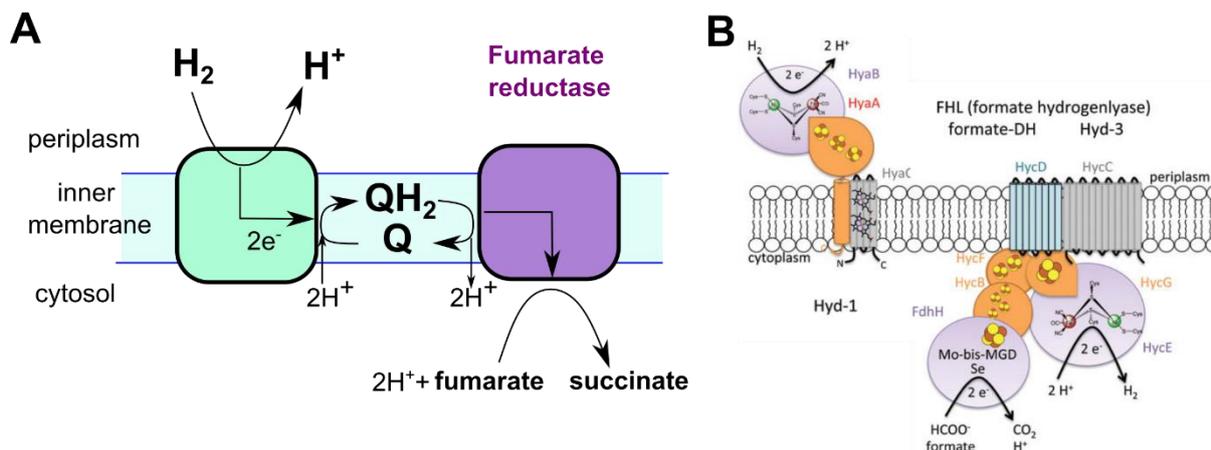
The *Hya* operon contains the gene of the three subunits of Hyd-1 and of three maturation proteins (Chapter 4.3.3). The transcription of the *Hya* operon was found to be induced by a shift from aerobic to anaerobic conditions (294) and low pH (295). This induction was regulated by the transcription factors ArcA and AppY (294); both these transcription factors are global regulators, with ArcA switching on the expression of fermentation genes and repressing the aerobic pathways (296) and AppY regulating the metabolism under several conditions such as carbon starvation, phosphate starvation, low pH, or anaerobiosis (297).

It was also found that the *Hya* operon transcription was induced by formate and repressed by nitrate. The highest level of anaerobic expression of the *Hya* operon was obtained in glucose medium supplemented with formate and in glycerol medium supplemented with fumarate (Table 5.1) (294).

**Table 5.1** Effects of different anaerobic growth conditions on the expression of the *Hya* operon (minimal medium with supplementations). Units are based on the beta-galactosidase activity. Data taken from (294).

Medium	Aerobic activity	Anaerobic activity
Glucose	0.04	1.1
<b>Glucose + Formate</b>	<b>0.05</b>	<b>5.0</b>
Glucose + Fumarate	0.04	1.4
Glucose + Nitrate	0.03	0.12
<b>Glucose + Formate + Fumarate</b>	<b>0.03</b>	<b>5.7</b>
Glucose + Nitrate + formate	0.04	0.12
<b>Glycerol + Fumarate</b>	<b>0.04</b>	<b>8.3</b>
Glycerol + Nitrate	0.06	1.4
<b>Glycerol + Fumarate + Formate</b>	<b>0.08</b>	<b>10.8</b>
Glycerol + Nitrate + Formate	0.06	3.7

Fumarate is a terminal electron acceptor in anaerobic metabolism (Fig.5.1.A) (298), while formate is oxidized by the formate hydrogenlyase (FHL) complex to produce hydrogen (Fig.5.1.B) (299).



**Figure 5.1** **A.** Fumarate is a terminal acceptor during anaerobic metabolism. Quinone (Q) is reduced to quinol (QH<sub>2</sub>) by membrane-bound enzymes, such as Hyd-1, and is used to convert fumarate to succinate. **B.** Formate is a key intermediate in the anaerobic metabolism and is used in the production of hydrogen by the FHL complex. Image adapted from (300).

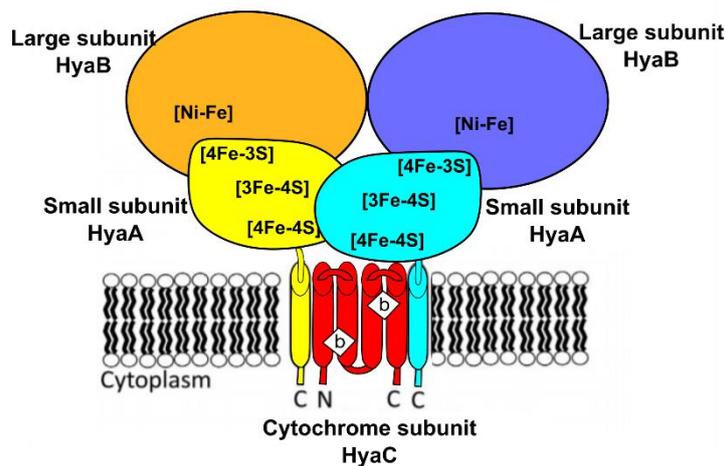
### 5.1.3 Structures of Hydrogenase-1

*E. coli* hydrogenase-1 (Hyd-1) is composed of three subunits HyaA, HyaB, and HyaC (Table 5.2) arranged in a 2:2:1 ratio (274). The presence of only one cytochrome subunit per Hya(AB)<sub>2</sub> dimer is consistent with the absence of translational coupling between the HyaB and HyaC genes (Chapter 4.3.3, Fig.4.12).

**Table 5.2.** Molecular weights in kDa of each subunit of Hyd-1 (including cofactors).

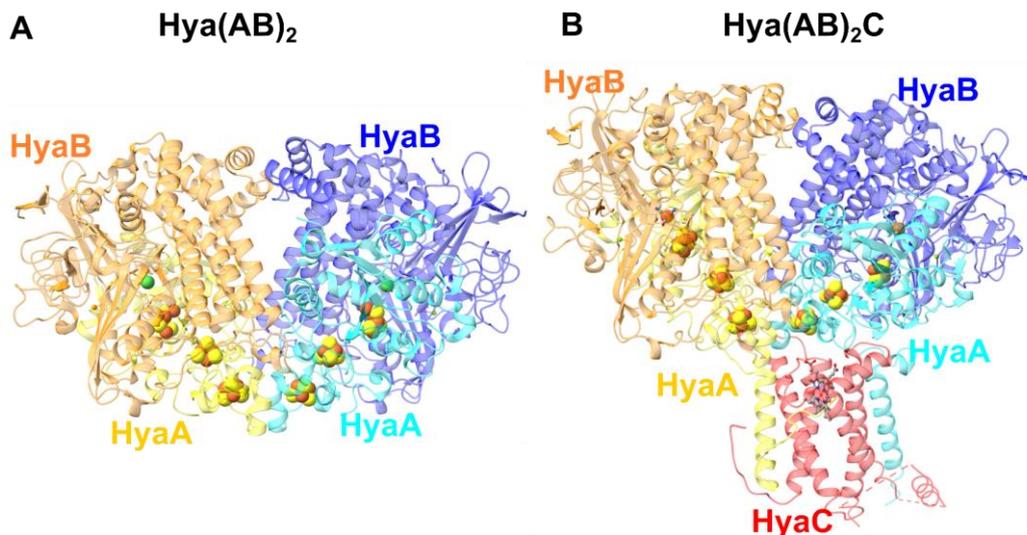
Subunit	Subunit	Molecular Weight (kDa)
HyaA	small subunit	36.7
HyaB	large subunit	66.3
HyaC	cytochrome subunit	28.8

Hyd-1 is anchored to the periplasmic side of the cytoplasmic membrane by the cytochrome subunit (HyaC) and the C-terminal transmembrane  $\alpha$  helices of the HyaA subunits (Fig. 5.2) (274).



**Figure 5.2.** Schematic of *E. coli* Hyd-1 showing the subunits arrangements. The enzyme comprises a large subunit (HyaB), containing a [NiFe] cofactor, tightly bound to a small subunit (HyaA), containing three FeS clusters. The dimer Hya(AB)<sub>2</sub> is anchored to the lipid bilayer by a single transmembrane domain located at the C-terminus of HyaA and a cytochrome subunit (hyaC). Image adapted from (290).

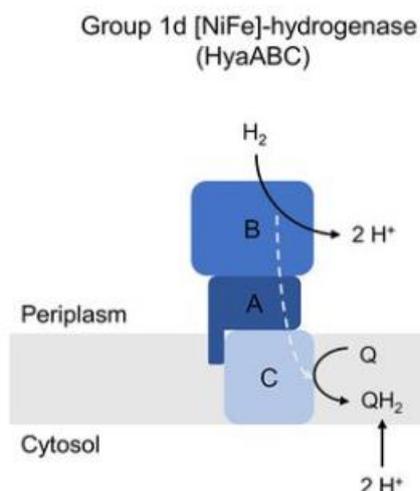
Currently, there are 12 X-ray structures of the Hyd-1 having various residue mutations, reduced states, and pH (274, 275, 301–304). All of them, except for two, are arranged as dimers of A and B subunits (Hya(AB)<sub>2</sub>) lacking the cytochrome and the membrane C-terminal domain of the HyaA subunit (Fig.5.3.A). The membrane domain could not be resolved due to being disordered in the crystal (305). Only two X-ray structures are available for Hyd-1 in complex with the cytochrome (PDBID: 4GD3, 6G94) (274, 275). These cytochrome-bound structures (Fig.5.3.B) are both very similar except for the FeS cluster proximal to the [Ni–Fe] active site, which causes the most recent structure (2018) to be oxygen-sensitive (275). In both these structures, only one cytochrome subunit is present for each (AB)<sub>2</sub> dimer, and the cytochrome contains just one b heme instead of the predicted two. In addition, a significant portion of the cytochrome and the C-terminal domain of the HyaA is missing.



**Figure 5.3.** **A.** Hya(AB)<sub>2</sub> X-ray structure, showing the arrangement of the HyaA and HyaB subunits and their FeS clusters. PDBID: 6FPO **B.** Hyd-1 X-ray structure with cytochrome HyaC attached showing a Hya(AB)<sub>2</sub>C arrangement. To be noted is the presence of only one of the two predicted hemes. PDBID: 4GD3.

#### 5.1.4 The function of the cytochrome subunit

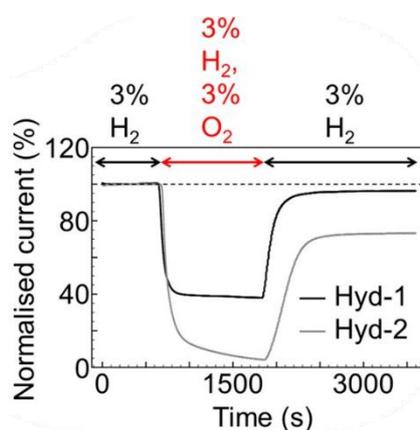
The physiological role of Hyd-1 is to couple H<sub>2</sub> oxidation (primarily produced by the membrane-bound formate hydrogenlyase, FHL (306)) to the reduction of the quinone pool in the inner membrane (Fig.5.4) (307, 308). The HyaC subunit is the point of contact for quinones as can be predicted by homology with the *E. coli* formate dehydrogenase-N and nitrate reductase-A (4, 309). Indeed, it was shown that the purified heterodimeric form of Hyd-1, Hya(AB)<sub>2</sub>, lacking the cytochrome was active with benzyl viologen as an artificial electron acceptor but unable to reduce quinones (310).



**Figure 5.4.** Schematic diagram showing how Hyd-1 and other Group1d hydrogenases couple the oxidation of  $H_2$  to quinone reduction via the HyaC transmembrane cytochrome *b*. Image adapted from (92).

### 5.1.5 Mechanism of oxygen tolerance in Hyd-1

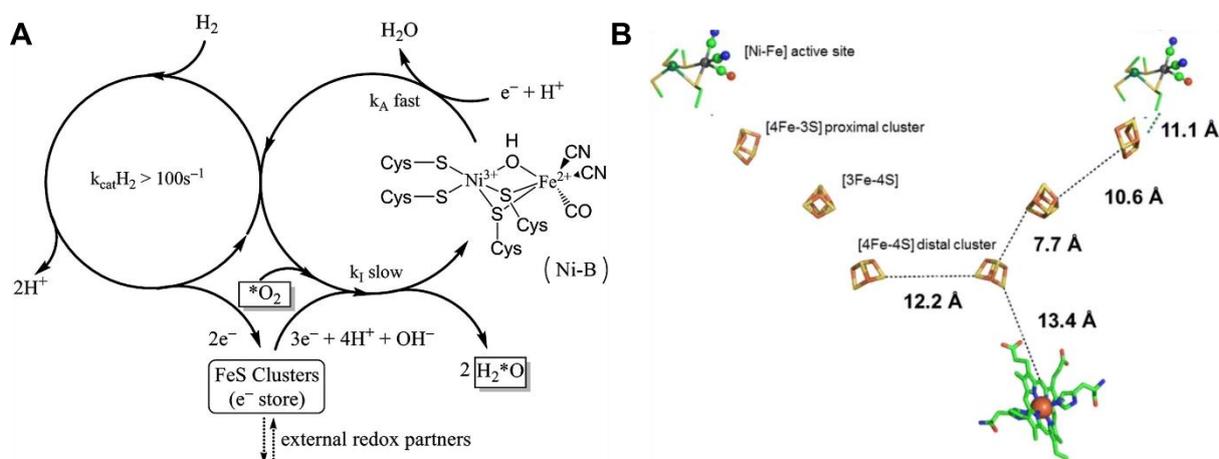
Oxygen-tolerant [NiFe]-hydrogenases can catalyze  $H_2$  oxidation under aerobic conditions without damaging the active site although their activity may be reduced (311). The  $H_2$  oxidation activity of the oxygen-tolerant Hyd-1 compared to the oxygen-intolerant Hyd-2 is shown in Fig.5.5 which demonstrates almost complete recovery of initial activity in Hyd-1 after a period of oxygen exposure (312).



**Figure 5.5** Constant potential experiments measuring the percentage of initial  $H_2$  oxidation activity which is sustained when the enzyme is exposed to  $O_2$ , and also quantifying the reversibility of the inhibition. For Hyd-1 a voltage of +0.06 V versus SHE is used, and for Hyd-2 -0.16 V versus SHE. Image from (312).

The key mechanism accounting for this special property is the rapid conversion of  $O_2$  molecules attacking the active site to water ( $2H_2 + O_2 \rightarrow 2H_2O$ ) (313).

This reaction relies on the capability to rapidly provide the electrons (and necessary protons) required to ensure the complete reduction of an  $O_2$  molecule each time it attacks (Fig.5.6.A). In Hyd-1 this is based on the unique properties of the [4Fe-3S] and [3Fe-4S] clusters and protein environment around the active site (314). It was also proposed that  $O_2$  tolerance depends on the presence of a dimeric quaternary structure in which the electron-transfer relay chain of one partner is electronically coupled to that of the other (Fig.5.6.B) (313).



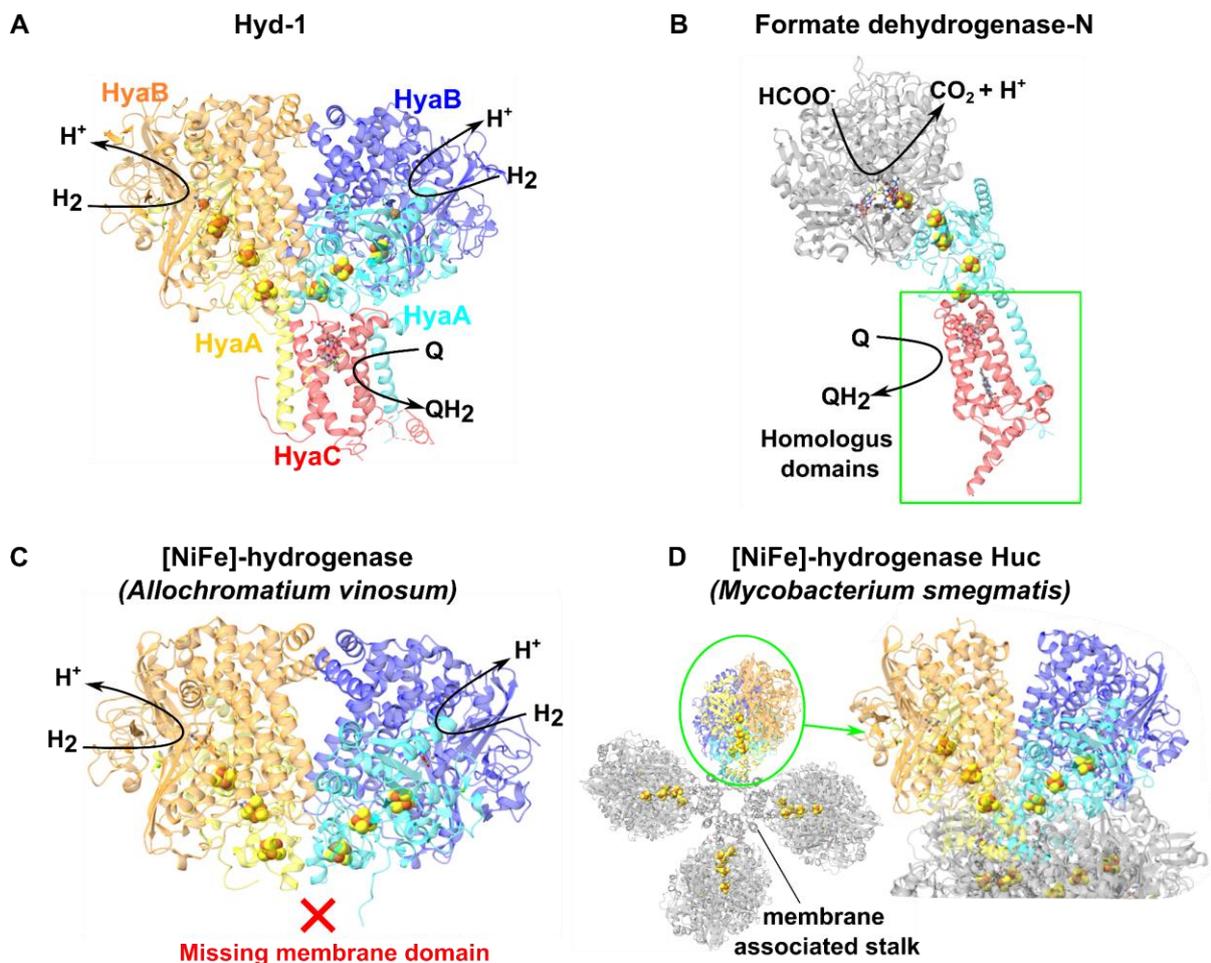
**Figure 5.6. A.** Simplified catalytic cycles of Hyd-1. Hydrogen oxidation (Left) proceeds at rates exceeding  $100\text{ s}^{-1}$  and feeds electrons into the relay system of FeS clusters. Oxygen reduction to water (Right), which begins with the  $O_2$  attack on the active site, with the utilization of the electrons stored in the FeS relay. Image taken from (314). **B.** The redox centres in Hyd-1: [NiFe] active site and iron-sulfur clusters are labelled with edge-to-edge electron transfer distances. The position of the b-type heme in the membrane anchor is also shown. Image taken from (313).

### 5.1.6 Homologous hydrogenases to Hyd-1

Hyd-1 is homologous to the oxygen-sensitive Hydrogenase-2 (Hyd-2) from *E. coli* (PDBID: 6EHQ (315)), the  $O_2$ -tolerant membrane-bound hydrogenase-5 from *Salmonella enterica* (PDBID:4C3O (316)), the membrane-bound [NiFe]-hydrogenases from *Ralstonia eutropha* (PDBID: 3RGW (317)), *Hydrogenovibrio marinus* (PDBID: 3AYZ (318)), and *Allochromatium vinosum* (PDBID: 3MYR (319)), the [NiFeSe]-hydrogenase from *Desulfovibrio vulgaris* (PDBID: 5JSH (320)), the high-affinity [NiFe]-hydrogenase Huc from *mycobacterium*

*smegmatis* (PDBID: 7UTD (21)), and the *E. coli* membrane-bound formate dehydrogenase-N (PDBID: 1KQF (4)).

Of these homologous structures only the formate dehydrogenase-N one is resolved with the homologous membrane domain (Fig.5.7.B). In this X-ray structure, obtained from protein solubilized in DDM and purified by anion exchange chromatography (4), the cytochrome subunit is intact and contains both hemes. As in Hyd-1, the cytochrome subunit in formate dehydrogenase-N participates in the quinone pool (4).



**Figure 5.7.** **A.** Crystal structure of Hyd-1 from *E. coli* showing the cytochrome subunit (partially resolved). PDBID: 4GD3 (274). **B.** Crystal structure of Formate dehydrogenase-N from *E. coli* showing the membrane domain homologous to that of Hyd-1 (green square). PDBID:1KQF (4). **C.** Crystal structure of the membrane-bound [Ni-Fe]-hydrogenase from *A. vinosum*. The membrane domain is not resolved. PDBID: 3MYR (319). **D.** Cryo-EM structure of the high-affinity Huc hydrogenase complex from *M. smegmatis*. Four domains that are homologous to the soluble domain of Hyd-1 (green circle) are anchored to the membrane through a membrane-associated stalk. PDBID: 7UTD (21).

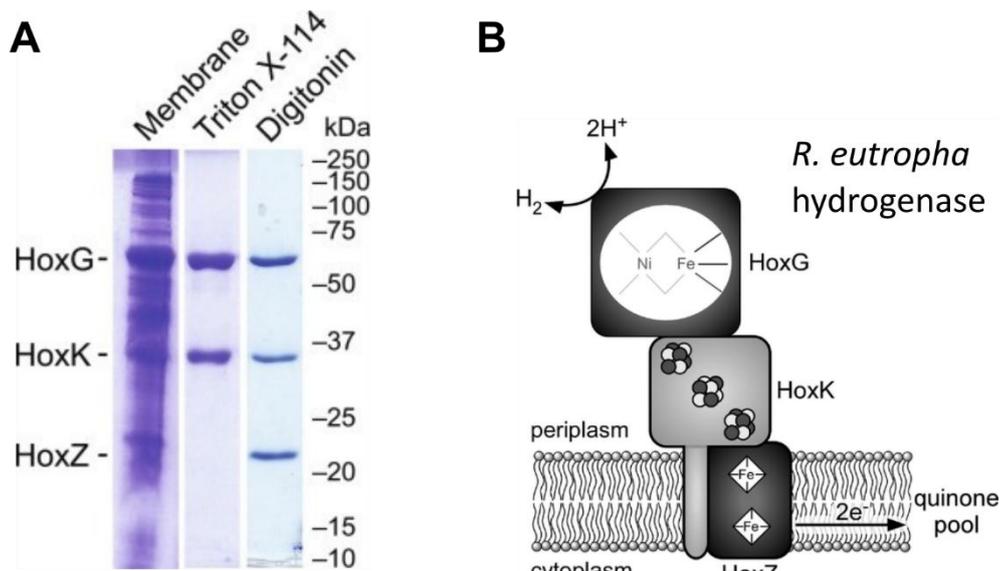
### 5.1.7 Purification Strategies and limitations

At the moment there is no available structure of the Hyd-1 with a fully intact membrane domain. Similarly, all the available structures of homologous cytochrome-containing membrane-bound [Ni-Fe] hydrogenases (e.g. from *R. eutropha*, *H. marinus*, *A. vinosum*, *S. enterica*) do not resolve the membrane domain (Table 5.3). From reported data, it appears that the cytochrome domain was particularly unstable being lost at some point between the solubilization with the detergent Triton-X and the affinity or ion exchange purification.

**Table 5.3** Purification strategies used for structural determination of Hyd-1 and homologous cytochrome-containing membrane-bound [Ni-Fe] hydrogenases.

Organism	Is the membrane domain resolved?	Method	Detergent used for solubilization	Purification strategy	Reference
<i>E. coli</i> Hyd-1	Partially	X-ray crystallography	Triton X-100	His-Tag at the C-terminus of HyaA	(274)
<i>R. eutropha</i>	No, the cytochrome subunit (HoxZ) is lost during purification.	X-ray crystallography	Triton X-114	Strep-Tag at the C-Terminus of HoxK	(317, 321)
<i>H. marinus</i>	No, the cytochrome subunit is lost during purification.	X-ray crystallography	Triton X-100	Ion exchange chromatography	(318)
<i>A. vinosum</i>	No, the cytochrome subunit is lost during purification.	X-ray crystallography	Triton X-100	Ion exchange chromatography	(319)
<i>S. enterica</i>	No, the cytochrome subunit is lost during purification.	X-ray crystallography	No detergent	His-Tag at the C-terminus of HydA replacing a portion of the transmembrane domain	(316)

Currently, only the *R. eutropha* membrane-bound [Ni-Fe]-hydrogenase was reported to be successfully purified in its fully intact oligomeric state by solubilization in digitonin (Fig.5.8.A), however, no structural studies have been performed on this improved preparation (322, 323).



**Figure 5.8.** **A.** SDS-PAGE gel showing the band pattern of purified heterodimeric and heterotrimeric *R. eutropha* membrane-bound [Ni-Fe]-hydrogenase. Proteins were natively expressed from a strain containing a Strep-Tag at the C-terminus of the HoxK subunit. The only difference is the use of Triton X-114 vs. digitonin for solubilization of the membrane fractions. Taken from (322) **B.** Scheme subunit composition of the *R. eutropha* membrane-bound [Ni-Fe]-hydrogenase. Image taken from (323).

### 5.1.8 Cryo-EM to study membrane proteins

For this project, we chose to use cryogenic electron microscopy (cryo-EM) to try to determine the structure of the cytochrome-bound Hyd-1. Cryo-EM is a powerful tool to investigate the structures of membrane proteins (MP) as they are generally hard to crystalize and the purified yield is often low (324). However, membrane protein aggregation and adoption of a preferred orientation remain significant challenges that are dependent on many parameters such as type and concentration of detergents, salt concentration, protein concentration, grid type, and blotting force and time (324). These parameters need to be optimized for every protein analyzed as every MP behaves differently and currently no tool is available for the prediction of optimized parameters.

The structural determination of membrane proteins (MP) is increasingly important given their fundamental role in physiological processes with more than half of the FDA-approved drugs targeting MP (325), therefore, improving protocols and methods in cryo-EM instrumentation and sample preparation will be of fundamental importance.

## 5.2 Aims of the chapter

This project started as a way to investigate further the role and structure of the cytochrome HyaC in Hyd-1 from *E. coli*. All the studies so far have been focused on the HyaA and HyaB subunits mainly regarding the oxygen tolerance mechanism and active site. Further information on the cytochrome structure, arrangement, and quinone binding site may be useful in understanding how this enzyme operates physiologically *in vivo*.

In this result chapter, the purification and cryo-EM results for Hyd-1 are presented. Hyd-1 has been purified both from a His-Tagged and Strep-Tagged strain (prepared as described in Chapter 4) and cryo-EM data has been collected using different detergents and parameters.

The chapter aims are:

- 1) To establish a better purification protocol for the purification of the intact Hyd-1.
- 2) To determine the Cryo-EM structure of Hyd-1.

## 5.3 Results: His-Tagged preparation

### 5.3.1 Overview of the purification strategy

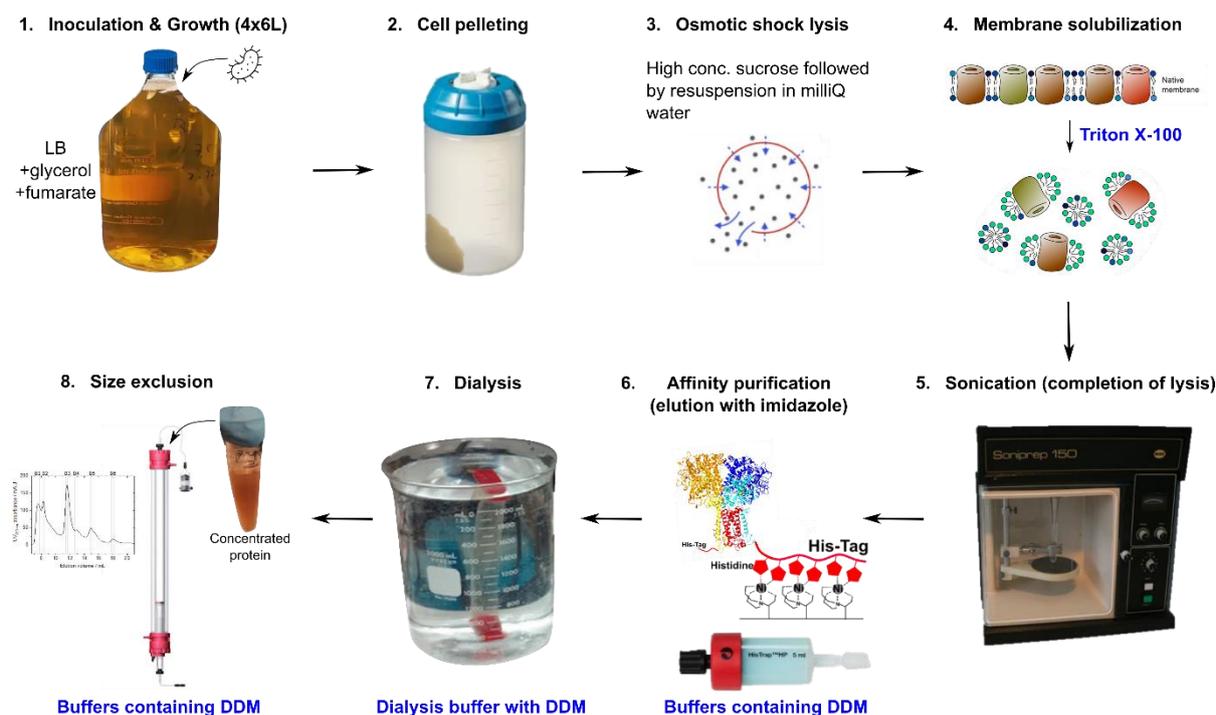
*E. coli* cells (K12 strain, LAF-003 mutant (165)), featuring a hexa-histidine tag at the 3'-terminus of the genomic gene encoding the Hyd-1 small subunit (HyaA) were used (Chapter 4.3.2, Fig.4.11.B). The cells were grown under anaerobic conditions at 37 °C in LB medium with the addition of 4 g/L of fumarate, as a terminal electron acceptor, and 5 mL/L of glycerol as a carbon source. These growing conditions were used to increase the native Hyd-1 production, however, their optimization was not attempted at this stage.

After growth, the cells were pelleted and lysed using osmotic shock. Then Triton X-100 was added to solubilize the membranes (overnight). Following solubilization, the suspension was further lysed using sonication. After centrifugation, the supernatant, containing cytoplasmic and solubilized membrane proteins, was loaded onto a HisTrap column (affinity purification) and the fractions containing the Hyd-1 protein were dialyzed in a buffer containing DDM.

Finally, the dialyzed solution was concentrated and loaded on a size-exclusion column to separate the oligomeric states.

In one variation of the protocol (detergent-free) no detergent was used from and including the affinity purification stage.

The entire purification protocol is schematized in Fig.5.9.

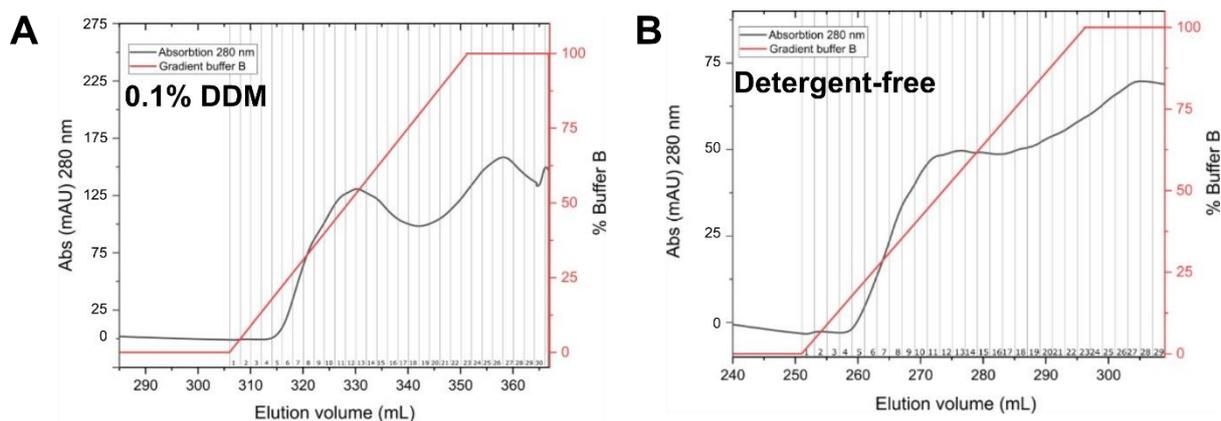


**Figure 5.9.** Schematic of the purification protocol used to purify Hyd-1 using a His-tagged strain and native expression. In a detergent-free variation of the protocol, DDM is not used at any stage. Membrane solubilization and dialysis images adapted from (326, 327).

### 5.3.2 Affinity purification

The enzyme Hyd-1 was purified using a HisTrap column. The small subunit, HyaA, has a polyhistidine tag (six histidine residues) at its C-terminus, increasing the affinity for the  $\text{Ni}^{2+}$  ions in the HisTrap column. The proteins that don't bind specifically to the column are washed and collected as flowthrough using buffer A, which contains a small concentration of imidazole (30 mM) to remove the weakly bound proteins. The strongly bound enzyme was then eluted using a linear gradient of 30 mM to 500 mM imidazole (increasing the % of buffer B) and UV absorbance at 280 nm was measured to monitor the detachment of the enzyme from the column.

In one preparation, both buffers A and B contained 0.1% w/v DDM (Fig.5.10.A). In another preparation (detergent-free) no DDM was used (Fig.5.10.B). In both cases, the protein (monitored at 280 nm) started to elute at ~10% Buffer B (corresponding to ~90 mM imidazole concentration) forming a broad elution peak. 30 fractions were collected for each preparation.

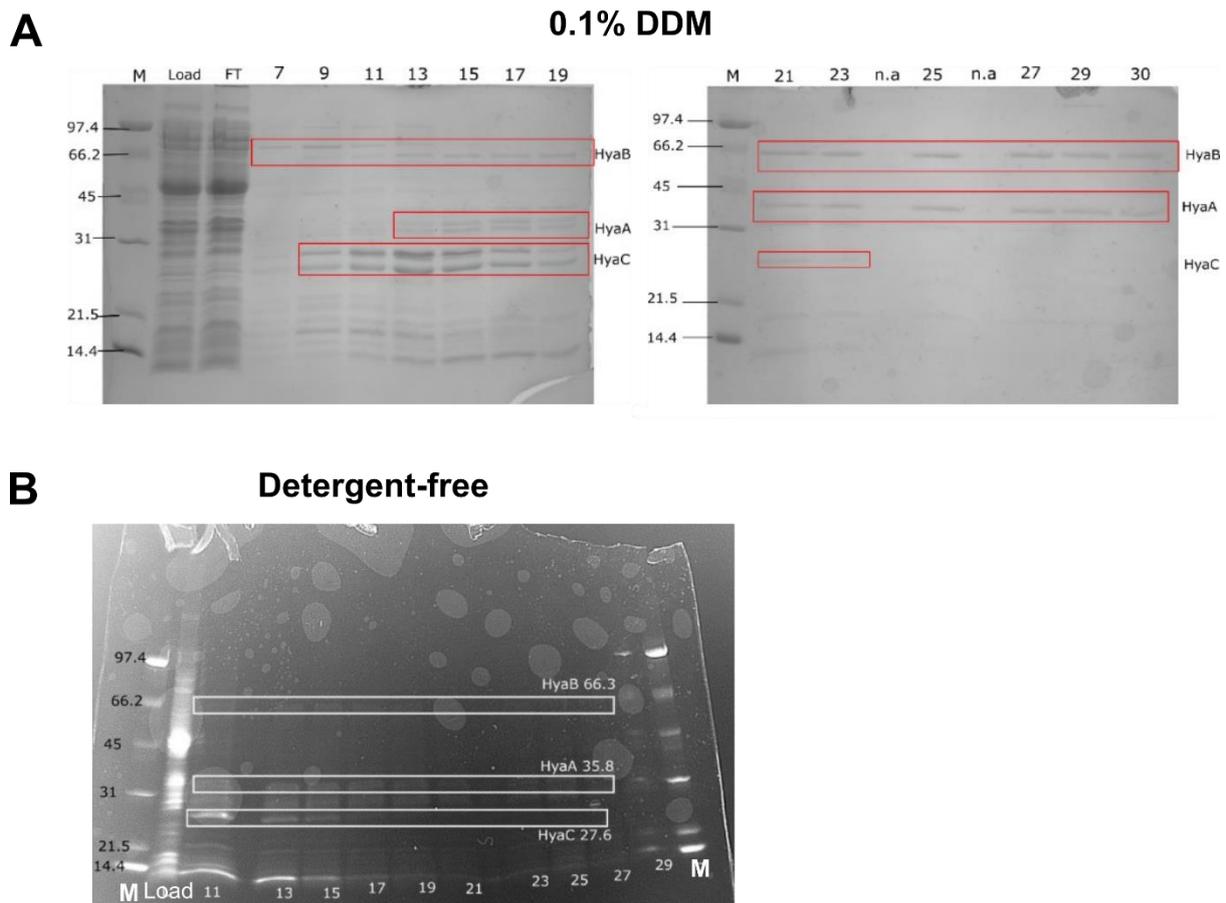


**Figure.5.10.A** Purification of Hyd-1 using Ni<sup>2+</sup> affinity column. **A.** Chromatogram showing elution volume against UV absorbance at 280 nm (black) and gradient of elution buffer (Red). The binding (buffer A) and elution buffer (buffer B) contained 0.1% w/v DDM. 30 fractions of 2 ml each (grey vertical lines) were collected. **B.** Analogous chromatogram, however, no DDM (or any detergent) was used in both buffer A and B. In each preparation, the initial mass of the cell pellet was ~ 30-35 g.

SDS-Page gels were used to analyze fractions from both preparations (DDM and detergent-free). In the preparation with DDM, it was observed that fractions 7-19 appeared to have different subunits/contaminants than fractions 21-30 (Fig.5.11.A), and therefore they were collected separately in two pools (Pool1 and Pool2). Fractions 7-19 also appear to have double bands possibly due to aggregation of the subunits with other peptides. Due to the high degree of contamination and double bands, it was difficult to verify the presence of all three subunits.

In the detergent-free preparation, the relatively high amount of contaminants made difficult the band assignment, however, it seems that all three subunits are present across all fractions, although with significantly different concentrations and ratios.

Variability in cytochrome presence across the fractions and relatively high presence of contaminants were found also in similar preparations from other members of the lab.



**Figure 5.11.** SDS-Page gels of fractions after affinity purification. **A.** On the left SDS-page gel with the fractions from 7 to 19 (Pool 1). On the right SDS-Page gel with fractions 21 to 30 (Pool 2). These fractions are obtained from affinity purification in the presence of DDM **B.** SDS-page gel with fractions from 11 to 29 obtained from affinity purification in the absence of detergents. Lane M: marker ladder. Lane 'Load': from loaded suspension (before affinity purification). Bands have been tentatively assigned (low confidence).

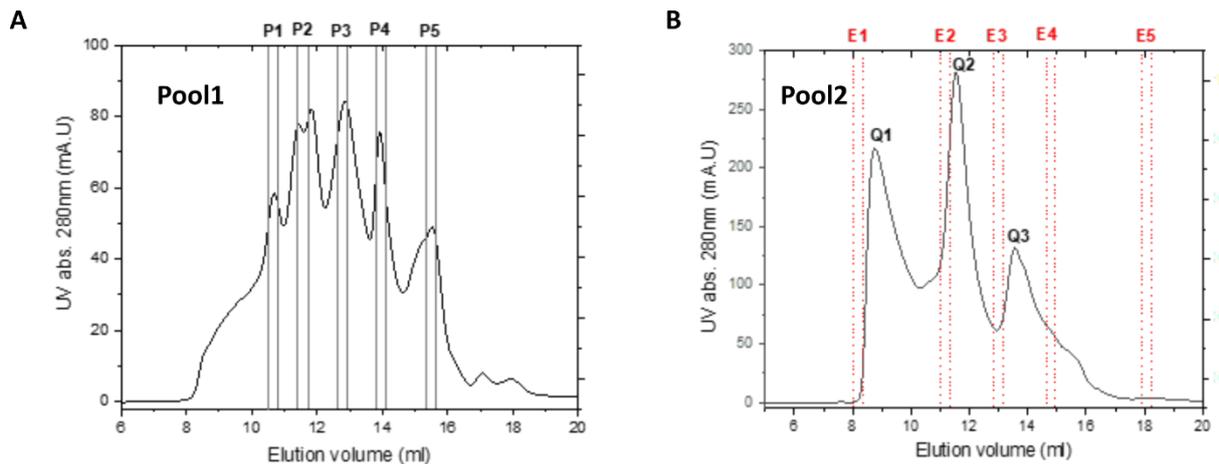
### 5.3.3 Size exclusion chromatography (separation of oligomeric states)

#### *In the presence of DDM*

Pool1 and Pool2 from the affinity purification in the presence of DDM were dialyzed separately and concentrated to 9.3 and 3.7 mg/mL, respectively. As this concentration was measured using the Bradford assay in the presence of detergent, the values should be taken as approximate. These protein samples were then loaded onto a size-exclusion chromatographic (SEC) column to separate proteins based on size (larger molecules elute quicker through the column). The SEC buffer contained 0.03% DDM to maintain the membrane domain stable.

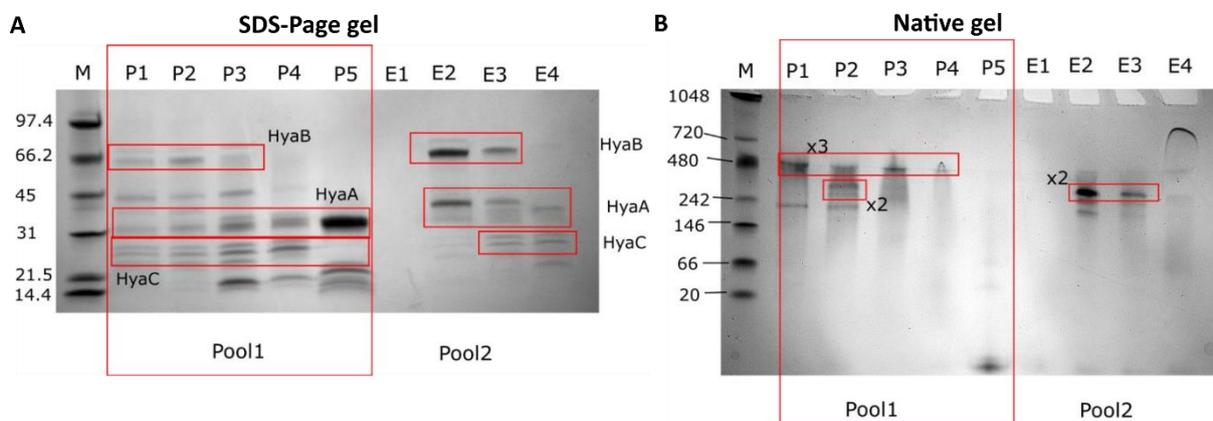
Pool1 and Pool2 resulted in very different SEC traces (Fig.5.12, A and B). Pool2 trace (Fig.5.12.B) is characteristic of Hyd-1 from previous work in the group, while the Pool1 trace (Fig.5.12.A) is instead very different from previous lab members' results with peaks that are difficult to resolve. This discrepancy may be because of the use of DDM or because previously all the fractions containing protein after the affinity purification were collected together.

### 0.03% DDM



**Figure 5.12.** Size-exclusion chromatography (SEC) traces showing elution of post HisTrap column of (A) Pool1 and (B) Pool2. Both samples were run using a buffer C containing 0.03% DDM. Traces show the volume eluted from a Superdex® 200 gel filtration column against UV absorption recorded at 280 nm. Fractions corresponding to significant peaks are labelled above the plots.

SDS-page (Fig.5.13.A) and native gels (Fig.13.B) were obtained for the SEC peaks in both Pool1 and Pool2, however, due to high levels of contamination and double bands the confidence in band assignment remains low. Nevertheless, it was clear that the protein eluted in multiple oligomeric states with a variety of compositions.

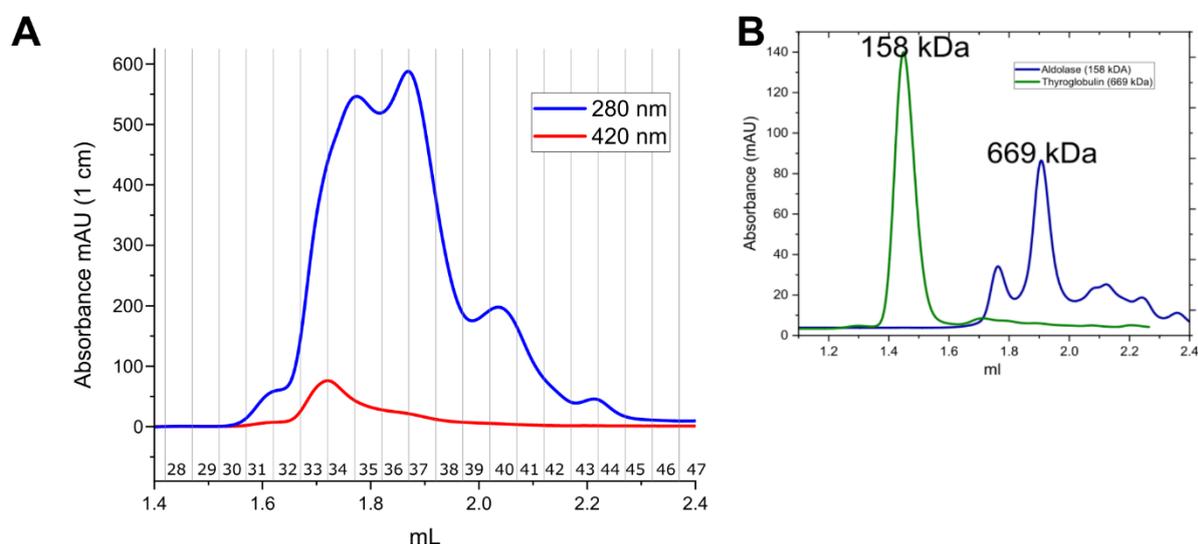


**Figure 5.13** SDS-page gel of (A) Pool1 and Pool2 of the fractions corresponding to the peaks in the SEC trace. (B) The native gel of SEC peaks in Pool1 and Pool2.

### *Detergent-free*

Fractions 7-30 from the affinity purification in the absence of detergents were collected and dialyzed together. Then these were loaded onto a size-exclusion chromatographic (SEC) using the Micro kit for ÄKTA without using any detergents in the buffers. The advantage of using the Micro kit is that less sample needs to be loaded (20-50  $\mu\text{L}$ ) and the fractions are considerably less diluted allowing for direct use in the cryo-EM grid preparation.

The SEC trace obtained using this method (Fig.5.14.A) shows that fractions 33-34 have the cytochrome present as monitored by the 420 nm absorbance (corresponds to the Soret band of hemes). Fractions 33-34 and 36-37 were directly used for Cryo-EM grid preparation. BCA assay revealed that these fractions had a protein concentration of  $\sim 0.10$  mg/mL.



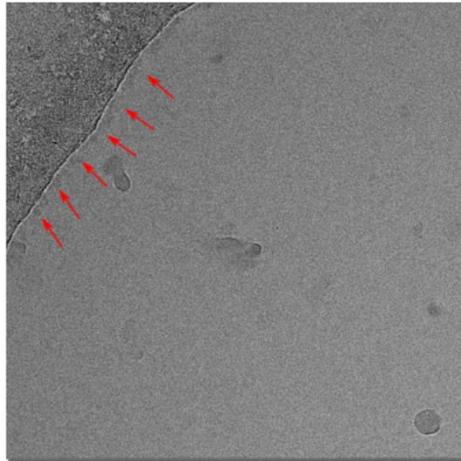
**Figure.5.14. A.** Size-exclusion chromatography (SEC) traces showing elution of post affinity column. All the traces plot volume eluted from a Superdex 200 Increase 3.2/300 filtration column against UV absorption recorded at 280 nm and 420 nm. Fractions are numbered below the plot. **B.** Calibration curves using aldolase (158 kDa, peak at 1.90 mL) and thyroglobulin (669 kDa, peak at 1.45 mL) standards.

### **5.3.4 Cryo-EM analysis**

#### *After SEC in the presence of DDM*

Peak Q2 from the SEC trace of Pool2 (Fig.5.12.B) was used to prepare cryo-EM grids. This peak was likely corresponding to the Hya(AB)<sub>2</sub> dimer and was consistent in all preparations from previous lab members.

The measured concentration before grid preparation was 2.0 mg/ml (Bradford assay), however, once the sample was imaged in the microscope (Fig.5.15) it appeared clear that the enzyme was aggregating at the border of the grid hole and the concentration appeared to be significantly lower than the measured one (the Bradford assay is known to be less reliable in the presence of detergents (328)). Data processing did not result in a sufficient number of particles to progress with the analysis.

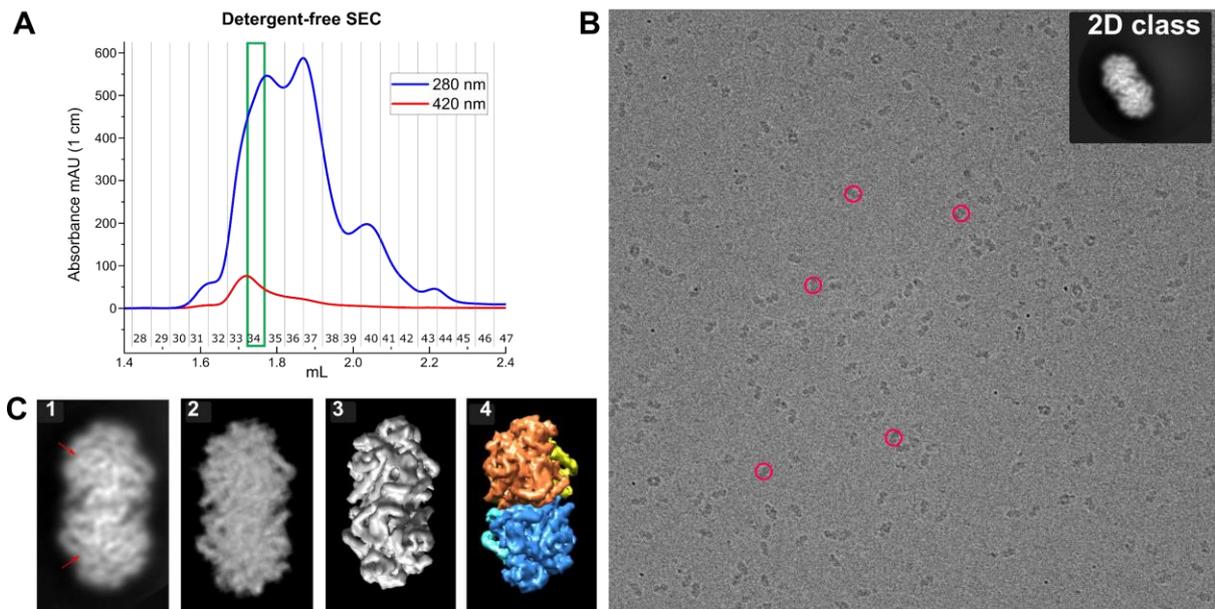


**Figure 5.15.** Micrograph of the highest concentrated sample (2.0 mg/ml) in a 1.2/1.3 Quantifoil grid. The magnification used was x150,000.

#### *After SEC in the absence of detergents*

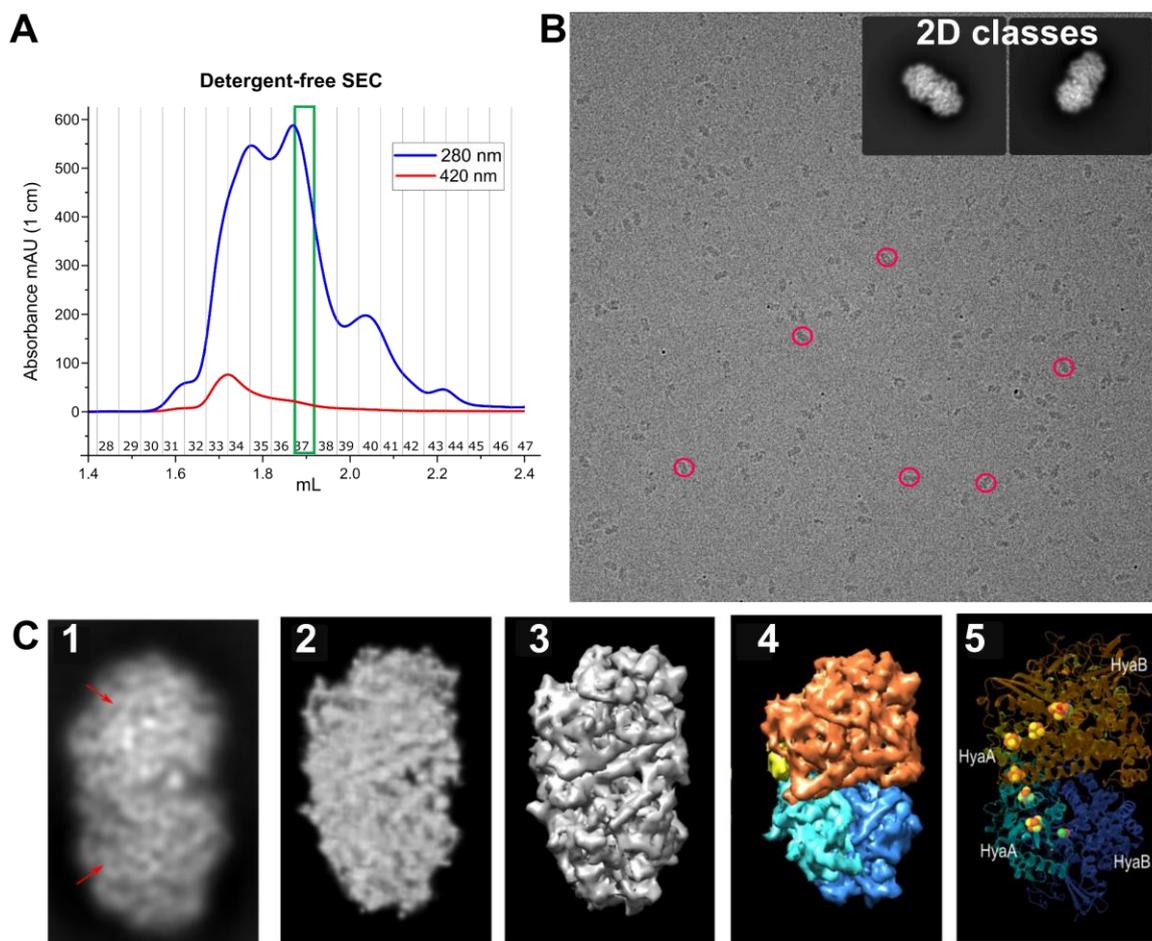
Fraction 34 (with cytochrome, see the 420 nm signal in Fig.5.14.A) from the detergent-free SEC trace was selected to prepare cryo-EM grids using 1.2/1.3 UltrAuFoil grids. 9 grid squares were used for collection and 904 EER movies were collected. About 203,000 particles were picked and used for 2D classification. Only a single 2D class was found (from 50,000 particles) (Fig.5.16.C, panel 1) due to the strong preferred orientation of the sample in the ice layer. As only a single 2D class was found, a 3D reconstruction could not be obtained.

The single 2D class was consistent with a dimer Hya(AB)<sub>2</sub> as shown from the comparison with the X-ray crystallographic data (Fig.5.16.C), however, the cytochrome could not be visualized due to the particular position assumed by the particle (it is expected to be below it).



**Figure 5.16.** **A.** SEC trace from the detergent-free preparation showing the fraction used for cryo-EM grid preparation (green square). **B.** Micrograph of the chosen fraction collected at magnification  $\times 120,000$ . Some representative Hyd-1 particles are shown inside red circles. Particles are evenly distributed across the grid. On the top right corner, there is an image showing the only 2D class obtained **C.** (1) Hyd-1 2D class obtained from the cryo-EM data using RELION 3.1, (2) solid map created using existing crystallographic structures showing the similarity to the 2D experimental class, (3) 3D display of the solid map (4) 3D display of the solid map but with coloured subunits (HyaA: yellow and cyan, HyaB: orange and blue).

Subsequently, Fraction 37 (without cytochrome) from the detergent-free SEC trace (Fig.5.17.A) was selected to prepare cryo-EM grids using 1.2/1.3 UltrAuFoil grids. 9 grid squares were used for collection and 918 movies were obtained. About 213,000 particles were found and used for 2D classification. Only two 2D classes were found, of which one was identical to the class obtained from fraction 34 (Fig.5.16.C, panel 1) while the other was a new addition (Fig.5.17.C, panel 1). It was clear that there was a persistent preferred orientation tendency, however, the situation was slightly better than with the previous fraction (two vs. one 2D class). As in the previous case, a 3D reconstruction could not be obtained due to the limited 2D classes available.

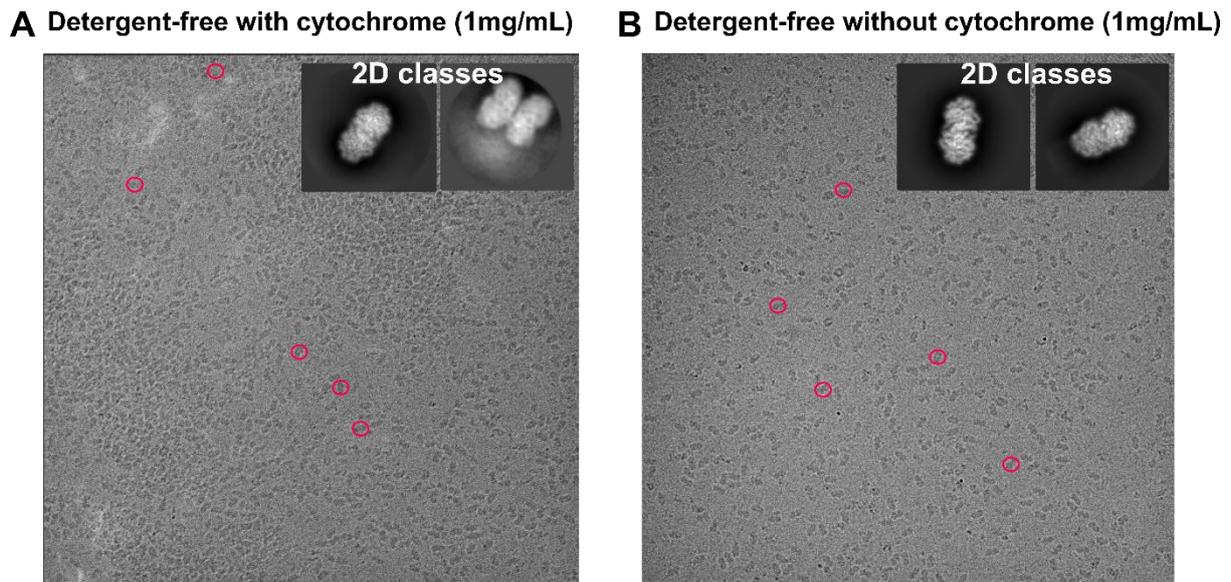


**Figure 5.17.** **A.** SEC trace from the detergent-free preparation showing the fraction used for cryo-EM grid preparation (green square). **B.** Micrograph of the chosen fraction collected at magnification  $\times 120,000$ . Some representative Hyd-1 particles are shown inside red circles. Particles are evenly distributed across the grid. On the top right corner, there is an image showing the only two 2D classes obtained **C.** (1) one of the two Hyd-1 2D classes obtained from the cryo-EM data using RELION 3.1, (2) solid map created using existing crystallographic structures showing the similarity to the 2D experimental class, (3) 3D display of the solid map, (4) 3D display of the solid map but with coloured subunits (HyaA: yellow and cyan, HyaB: orange and blue), (5) ribbon structure of Hyd-1 used to create the maps.

At a later stage, the detergent-free purification procedure was repeated, however, with samples exchanged into an imidazole-free buffer on the same day as running the HisTrap column (rather than using overnight dialysis) (170). This modified purification was performed by another lab member and two fractions (with and without the cytochrome band) were concentrated to 1 mg/mL and then analyzed using the same cryo-EM settings as the ones described above. The results were very similar except for the expected higher concentration of particles on the grid (Fig.5.18) which resulted in some protein aggregation observed in the cytochrome-containing fraction (Fig.5.18.A). The 2D classes results were identical to the respective ones obtained in

the previous detergent-free datasets (Fig.5.16 and 5.17) with the only exception that a few particles (~5,000) from the partially aggregated cytochrome fraction displayed a higher order tetrameric structure (Fig.5.18.A, top right corner). This tetrameric state, with two weakly bound dimers, was also observed in the crystallographic cytochrome-bound structures (PDBID: 6G94 (275)) and it is likely the result of the high concentration that forces proteins to be in close proximity.

It was concluded that also this sample behaved similarly as in the above cases.

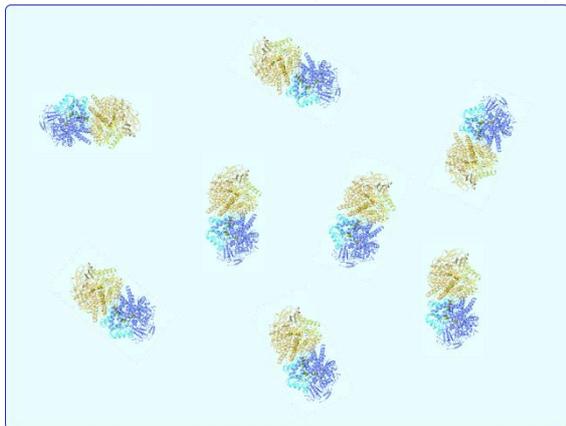


**Figure 5.18.** **A.** Micrograph of the fraction containing the cytochrome band collected at magnification x120,000. Some representative Hyd-1 particles are shown inside red circles. Some protein aggregation is visible. On the top right corner, there is an image showing the only two 2D classes obtained. **B.** Micrograph of the fraction without the cytochrome band collected at magnification x120,000. Some representative Hyd-1 particles are shown inside red circles. Particles are evenly distributed across the grid. On the top right corner, there is an image showing the only two 2D classes obtained.

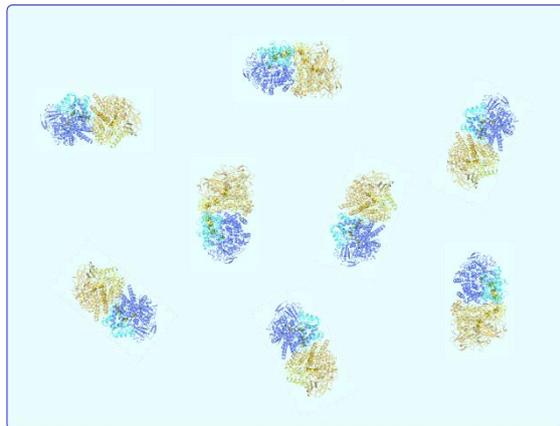
### 5.3.5 Preferred orientation: use of the CHAPSO detergent

The cryo-EM analysis of detergent-free samples indicated a consistently strong preferred orientation assumed by the Hyd-1 particles in the vitreous ice layer (Fig.5.19). The air-water interface tends to adsorb proteins and frequently causes preferred orientation problems in cryo-electron microscopy (329) which can be overcome by stabilizing the protein complex and by using detergents to modify the properties of the air-water interface.

**A. With cytochrome (one orientation)**



**B. Without cytochrome (two orientations)**

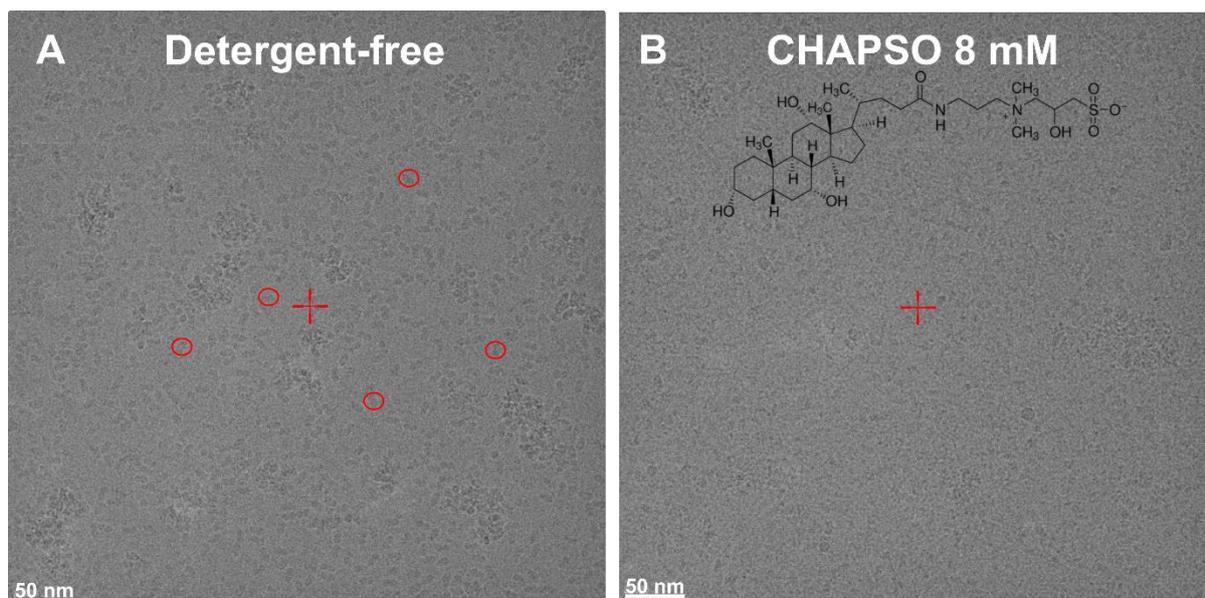


**Figure 5.19.** Schematic showing Hyd-1 particles in limited orientations in the vitreous ice layer (view from the top). **A.** Only one orientation was visible using the fraction containing the cytochrome (from SDS-page analysis and 420 nm signal). **B.** Only two orientations were visible using the sample without the cytochrome.

One reported detergent used to overcome preferred orientation issues is CHAPSO, which is a non-denaturing zwitterionic detergent. The addition of CHAPSO at the critical micelle concentration (CMC) of 8 mM was shown to completely prevent the adsorption of several bacterial RNA polymerase complexes to the air/water interface dramatically broadening the distribution of particle orientations (330).

Here, CHAPSO detergent, at various concentrations from 4 mM to 12 mM, was added to the purified samples just before the deposition onto the cryo-EM grids. CHAPSO was added to fractions with and without the cytochrome (1 mg/mL), however, in all grids analyzed a consistent denaturation effect was observed (Fig.5.20) indicating that the CHAPSO detergent negatively affected the stability of the Hyd-1 protein. Grids prepared using aliquots of the same protein samples but without the addition of CHAPSO did not display any denaturing effect. The use of CHAPS (a similar detergent) and CHAPSO on strep-tagged Hyd-1 samples resulted in the same denaturing effect.

The conditions experienced by the protein during grid preparation are far from those found in bulk solutions and therefore unexpected interactions with detergents or other compounds may emerge (331).



**Figure 5.20.** **A.** Micrograph obtained from a detergent-free Hyd-1 sample. Clear particles are visible (red circles). **B.** Micrograph obtained from the same Hyd-1 sample in A, however, with the addition of CHAPSO 8mM just before grid preparation. No clear particles are visible, and the appearance is consistent with the presence of ‘fallen apart’ proteins. The structural formula of CHAPSO is shown in black.

### 5.3.6 Discussion and summary: His-Tagged preparation

The His-tagged Hyd-1 protein was purified following established protocols in the lab which resulted in the expected broad range of oligomeric states (see SEC traces) and loss of the cytochrome subunit in a significant portion of the protein sample. The cytochrome subunit was lost in most of the sample even when DDM (a detergent) was used in both affinity purification and SEC steps.

Cryo-EM analysis of detergent-free samples showed the presence of a strong preferred orientation which may have been caused by the intrinsic properties of the Hyd-1 protein and/or by a destabilized protein sample. Indeed, it is expected that the membrane domain in the absence of detergent is unstable. In addition, based on published data (274, 275), it seems that the current preparation with the His-tag positioned at the C-terminus of HyaA combined with the use of imidazole (used for HisTrap elution) results in a partially denatured membrane domain even in the presence of detergents. In fact, similar protein preparations resulted in crystallographic structures (PDBID: 4GD3, 6G94) missing one heme and with a partially unresolved membrane domain.

It was decided that it was best to rethink the protein preparation in order to maximize the stability of the membrane domain. The first step involved redesigning a genetically edited strain to produce a strep-tagged Hyd-1 (Chapter 4). This produced only a single cytochrome-bound oligomeric state with a stronger heme signal (indicating both hemes are present).

## 5.4 Results: Strep-tagged preparation

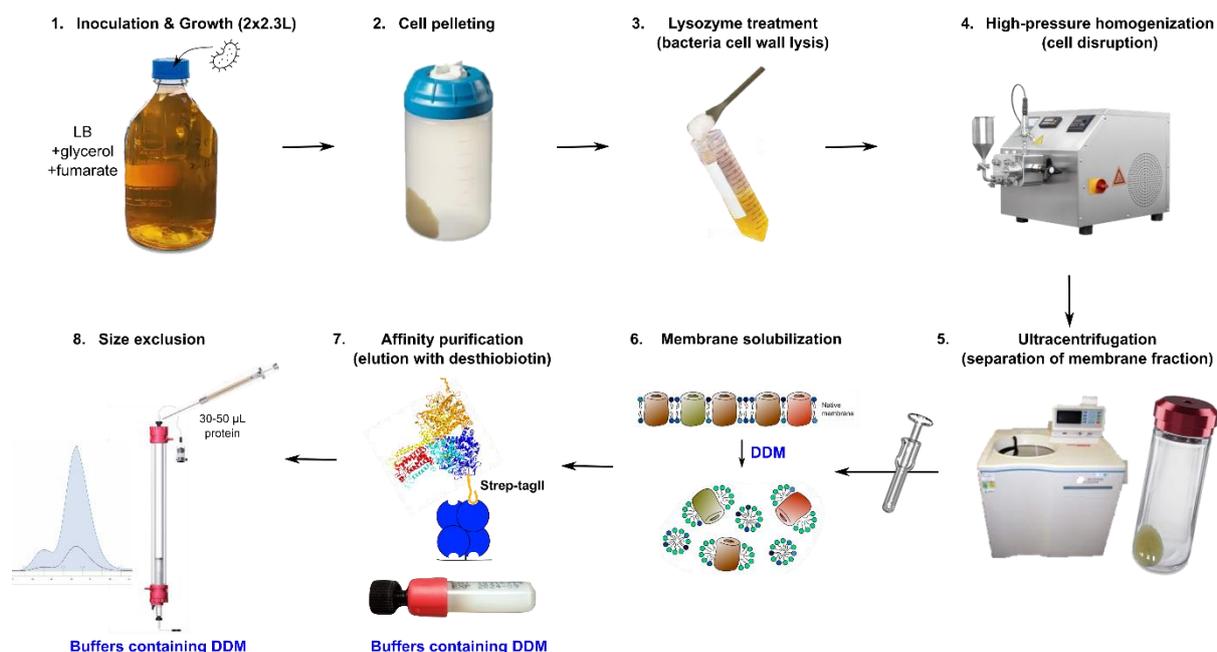
### 5.4.1 Overview of the purification strategy

*E. coli* cells (K12 strain, CF-001 mutant, see Chapter 4), featuring a Strep-tag®II at the 5'-terminus of the genomic gene encoding the Hyd-1 large subunit (HyaB) were used. The cells were grown under anaerobic conditions at 37 °C in LB medium with the addition of glycerol (5 ml/L), and fumarate (5 g/L). In some preparations the different growing conditions were tested; in particular, the effect of concentration and presence of fumarate and formate was tested while the effect of pH and supplementation of Fe and Ni ions (272) was not.

With Covid restrictions lifted, it was easier to access a greater variety of instrumentation than in the preceding sections of this chapter, including an ultracentrifuge and a cell disruptor. Access to these instruments allowed a different approach to purification, whereby cell lysis was achieved using lysozyme and cell disruption, which is quicker and easier and often gives better yields. Additionally, the ultracentrifuge allowed a fraction greatly enriched in membranes to be pelleted away from soluble cellular components and cell wall debris. This membrane fraction was then resuspended using a Dounce homogenizer. The resuspended membrane fraction was then diluted to a constant protein concentration of 5 mg/mL. The detergent Triton X-100 was avoided for solubilisation, instead, DDM was added to reach a 1% concentration to solubilize the membrane proteins. Diluting is important to avoid using a too high concentration of DDM which can be denaturing.

The solubilized membrane proteins were then loaded onto the StrepTrap affinity column using DDM-containing buffers (or GDN in a variation), and fractions were collected. The fractions containing the proteins were directly used for cryo-EM grid preparation or concentrated and loaded onto a size-exclusion column using a DDM-containing buffer (or LMNG in a variation).

The entire purification protocol is schematized in Fig.5.21.



**Figure 5.21.** Schematic of the purification protocol used to purify Hyd-1 using a Strep-tagged strain and native expression. In variations of the protocol, GDN and LMNG are used instead of DDM at the affinity purification and size exclusion steps, respectively. Membrane solubilization image adapted from (326).

## 5.4.2 Affinity purification

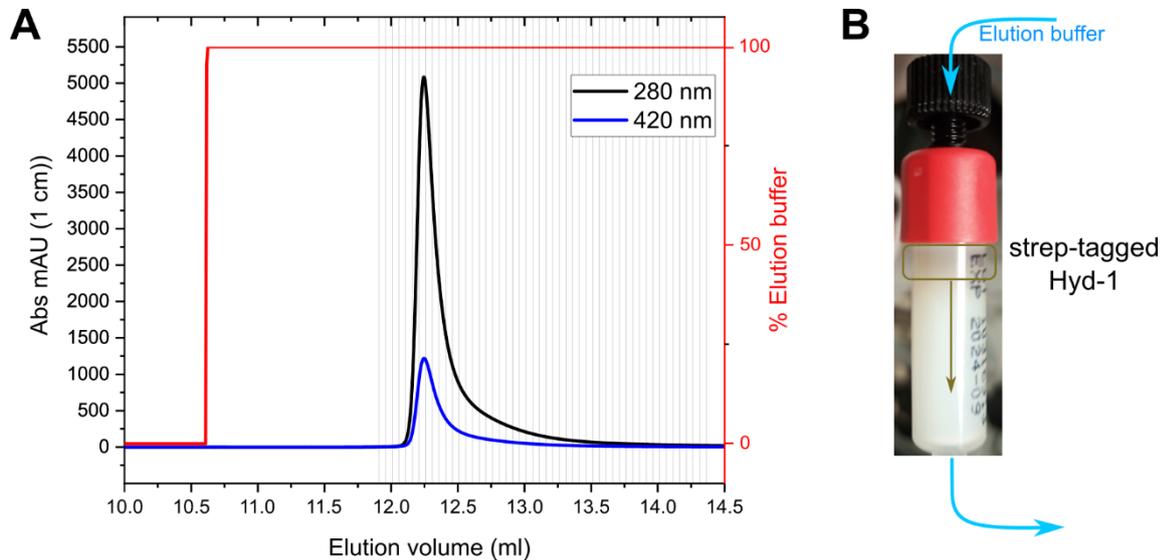
The enzyme Hyd-1 was purified using a StrepTrap affinity column (Fig.5.22.B). The large subunit, HyaB, has a Strep-tag®II at its N-terminus, which binds to the streptavidin in the affinity column. The proteins that do not bind specifically to the column are washed and collected as flowthrough using the binding buffer. The strongly bound enzyme was then eluted using the elution buffer containing desthiobiotin (no gradient works best as it produces higher concentration fractions). Absorbance at 280 nm and 420 nm was measured to monitor the detachment of the enzyme from the column and the presence of the cytochrome subunit, respectively.

Both binding and elution buffers contained 0.04 % w/v DDM. In another preparation, 0.01% GDN (a synthetic substitute for Digitonin) was used instead.

The affinity purification traces were very similar in all preparations (Appendix Fig.A.3), except for different 280 nm absorption levels depending on medium composition which reflect the different amounts of Hyd-1 produced. The common characteristic was a single sharp peak after the void with strong cytochrome absorption (the 280 nm to 420 nm absorption ratio is 4.2) as shown in a representative trace in Fig.5.22.A.

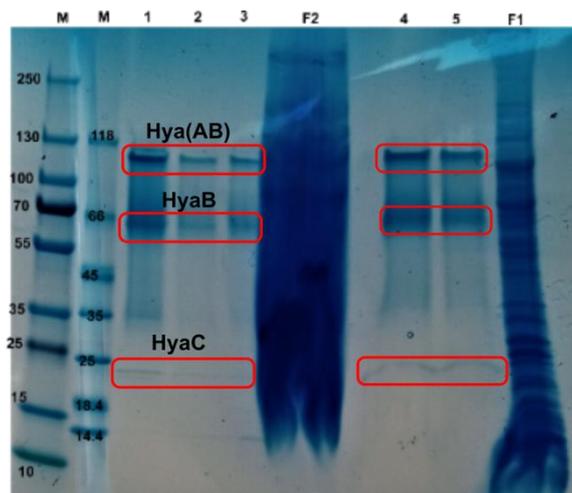
The advantages of the StrepTag over the HisTag affinity purification were:

1. A highly concentrated pure sample (visibly brown) directly eluted from the StrepTrap column (as expected by the sharp absorbance peak).
2. No use of imidazole and easy column regeneration.



**Figure.5.22.** A Purification of Hyd-1 using a StrepTrap affinity column. Chromatograms showing elution volume against UV absorbance at 280 nm (black) and 420 nm (blue) and gradient of elution buffer (red). 50 µL fractions were collected (grey vertical lines). **A.** From 4.6 L growth in LB medium supplemented with 5 mL/L glycerol, 2.5 g/L of sodium fumarate, and 1 g/L sodium formate. Buffers containing 0.04% DDM. **B.** StrepTrap column showing the visibly brown bound Hyd-1. The brown band compactly moves along the column during elution.

SDS-Page gels (Fig.5.23) indicated high sample purity, however, incomplete denaturation was observed (possibly due to skipping the heating step when SDS was placed on the protein samples). Strong bands, likely, from a Hya(AB) dimer (MM = 103 kDa) were observed.



**Figure.5.23** SDS-page of hyd-1 fractions (Lane 1, 2, 3, 4, 5) from various purifications and different dilutions. F1 and F2 lanes: flow-throughs. M lane: markers. Incomplete denaturation was observed. The expected molecular mass for the Hya(AB) complex is 103 kDa, HyaB is 66.3 kDa, and HyaC is 28.8 kDa.

### 5.4.3 Protein yields and the effect of medium composition

The initial pH of the unbuffered medium was 7 which decreases to ~6.5 at the end of all growth-containing glycerol regardless of the presence/absence of formate and/or fumarate. However, when glucose (2.5g/L) was supplemented the final pH decreased to 5.5. The higher acidity of the glucose-containing medium is likely caused by acetic acid which is a main fermentation product of glucose.

It was also observed that the addition of sodium formate (1 g/L) resulted in significantly more bubbles compared to the respective growths not containing it. This is expected, as formate is used by the FHL complex to produce CO<sub>2</sub> and H<sub>2</sub>.

Overall, the highest protein yield among all the conditions tested was obtained for the medium supplemented with 5 mL/L glycerol and 5 g/L Na fumarate (Table 5.4). This is in contrast to what is expected from published literature which predicts a higher *Hya* operon expression in the presence of both formate and fumarate (294). However, the conditions in this comparison experiments were not carefully monitored (i.e. by making sure the same batch of LB is used and by using the same precise incubation time) and there is not sufficient data for statistical relevance. Nevertheless, this data indicates that adding formate does not give a significant boost in protein production sufficient to justify the extra step of preparation and addition of the sterilized sodium formate.

**Table 5.4** Protein yields expressed as mg of purified Hyd-1 obtained per litre of culture. Values were calculated from the area below the affinity purification trace using the 1 cm path length reported by the instrument. The absorption coefficient, determined for the Hya(AB)<sub>2</sub>C complex is based on the protein sequence ( $410,000 \text{ M}^{-1} \text{ cm}^{-1} = 1'680 \text{ mg}^{-1} \text{ mL cm}^{-1}$ ).

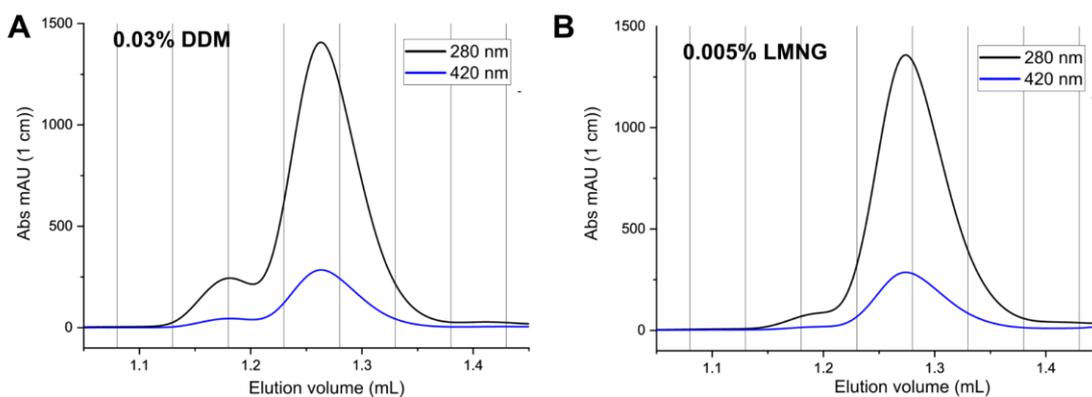
Growing conditions	Volume growth	Area (mL·AU)	Protein yield	280/420nm ratio
LB + 2.5mL/L glycerol	2.3L	921	0.24 mg/L	4.27
LB + 5 mL/L glycerol + 2.5 g/L Na fumarate	2.3L	790	0.21 mg/L	n.a.
LB + 5mL/L glycerol + 5g/L Na fumarate	4.6L	2100	0.27 mg/L	4.18
LB + 5mL/L glycerol + 5g/L Na fumarate + 1g/L formate	4.6L	1223	0.16 mg/L	4.17
LB + 5 mL/L glycerol + 2.5 g/L Na fumarate + 2 g/L formate + 2.5 g/L glucose	4.6L	540	0.070mg/L	5.0

#### 5.4.4 Size exclusion chromatography

30  $\mu\text{L}$  of concentrated fractions from the affinity purification peak was loaded on a Superdex 200 Increase 3.2/300 column coupled with the micro Akta kit. The size exclusion column was run with DDM-containing buffer (0.03% w/v) to keep the membrane domain stable (Fig.5.24.A). In a variation, the size exclusion column was run with LMNG-containing buffer (0.005% w/v) instead of DDM (Fig.5.24.B).

The 280/420 nm absorbance ratio measured at the peak was 4.9 and 4.75 for the DDM- and LMNG-containing runs respectively. This is significantly lower than the ratio obtained using the His-tagged protein (Fig.5.14.A) indicating a higher heme retention.

In addition, the size exclusion profiles of the strep-tagged Hyd-1 (Fig.5.24) only show one peak (excluding the void) indicating that a single cytochrome-containing oligomeric state is present. This is in contrast to the numerous peaks found in the His-tagged preparation (Fig.5.12 and Fig.5.14.A).



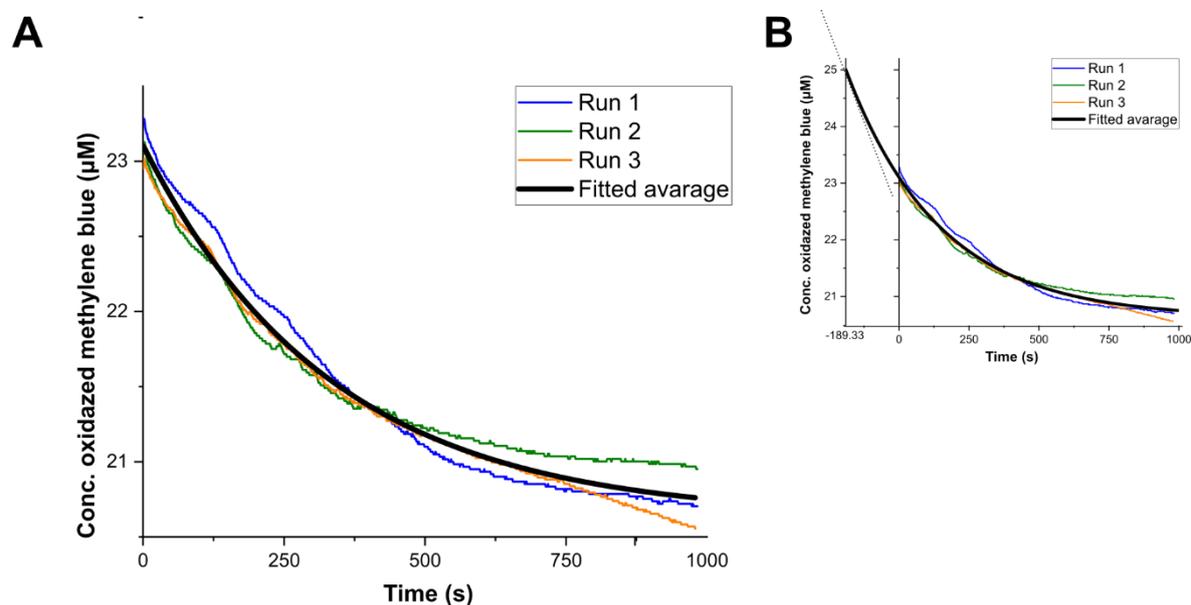
**Figure.5.24.** Size-exclusion chromatography (SEC) traces showing elution of post StrepTrap column using Buffer C containing (A) 0.03% DDM and (B) 0.005% LMNG w/v. Traces show the volume eluted from a Superdex 200 Increase 3.2/300 gel filtration column against UV absorption recorded at 280 nm and 420 nm. Fractions are between vertical light grey lines.

#### 5.4.5 Activity assay

The maximum H<sub>2</sub>-oxidation activity of as-isolated Strep-tagged Hyd-1 samples after affinity purification was measured using a methylene blue assay (Fig.5.25.A). In this assay, the hydrogenase catalyzed the oxidation of H<sub>2</sub>, and the released electrons were used to reduce methylene blue leading to the decolourization of the solution (monitored spectrophotometrically).

Using the total protein concentration determined via Bradford assay as a proxy for Hyd-1 concentration (factoring dilutions: ~ 4 nM), a turnover rate ( $k_{cat}$ ) of 3.3 s<sup>-1</sup> was determined (Fig.5.25.B). The turnover rate is calculated by dividing the maximum oxidation rate (the derivative of the fitted function at the initial concentration of oxidized methylene blue) by the concentration of the enzyme as per the definition of  $k_{cat}$  in the Michaelis–Menten model. It was necessary to extrapolate back to obtain an initial rate due to the delay in preparing the reaction and starting measurements. Whilst this approach would be unacceptable for detailed kinetic measurements, for indicative measures of activity it is sufficient. This as-isolated turnover rate is comparable to previous measurements reported by previous lab members on the His-tagged enzyme (332). As the activity is not influenced by the presence of the cytochrome a comparable activity was expected and indicates that the Strep-tagged Hyd-1 was produced in a correctly folded state. In the future it would be valuable to explore the cytochrome-dependant activity i.e. quinone reduction, using activity assays based around

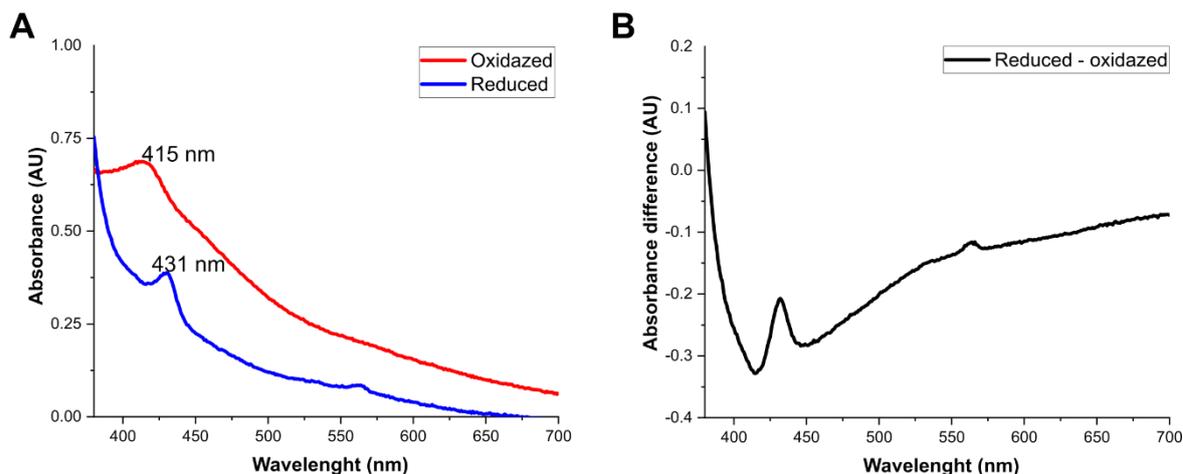
providing the enzyme with H<sub>2</sub> and menaquinone/ubiquinone either using water-soluble forms such as menadione (333) or vesicle-based systems with an enzyme to re-oxidise the quinol (334).



**Figure 5.25** **A.** Solution activity assay data showing the H<sub>2</sub>-oxidation activity of Hyd-1 determined via the reduction of methylene blue monitored spectrophotometrically. Data are shown for as-isolated Hyd-1 enzymes after StrepTrap affinity purification (peak fraction). Three repeated measurements (blue, green, orange) are averaged together. The averaged plot is fitted with an exponential decay function (thick black curve,  $y = 20.63 + e^{-t/333.43}$ ). **B.** Extension of the fitted curve to reach the initial concentration of oxidized methylene blue (25 µM) added and reduction rate at that point (slope of the tangent line  $y = 0.013107 \cdot t$ ).

#### 5.4.6 Heme spectroscopy

Spectra of as-isolated Hyd-1 proteins after affinity purification with DDM were taken in the absence and presence of sodium dithionite (reducing agent). The spectra (Fig.5.26) show the characteristic Q-bands shift upon reduction with sodium dithionite, however, significant background scatter was observed possibly due to the interference with the detergent.

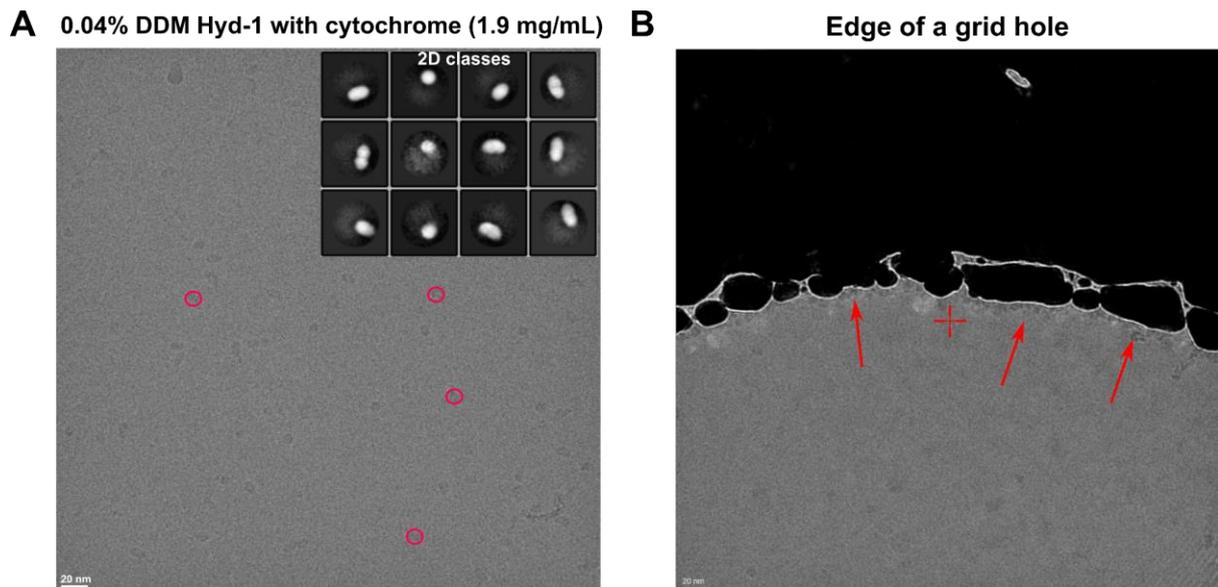


**Figure 5.25** **A.** Visible spectra showing the shift in the heme Q-bands upon reduction with sodium dithionate. **B.** Difference spectra representation (reduced – oxidized) of panel A.

#### 5.4.7 Cryo-EM: Hyd-1 after affinity purification with DDM

1.9 mg mL<sup>-1</sup> Hyd-1 aliquots obtained directly (without concentrating) from the affinity purification peak in the presence of 0.04% w/v DDM were defrosted and 2.5 μL placed onto a 1.2/1.3 UltrAuFoil grid selecting a blot time of 3 s with blot force at -5. The Hyd-1 concentration on the grid appeared to be very low (Fig.5.26.A) with only 8,000 particles found in total. It is known that the concentration of proteins on the grid does not always reflect the bulk concentration due to their interaction with the air-water interface (331). In this case, it appeared that Hyd-1 in the presence of DDM binds strongly to some other part of the grid or the filter paper used in blotting (Fig.5.26.B). A similar observation was made in the His-tagged preparation where aggregation at the border of the grid hole was observed in the presence of detergent (Fig.5.15).

The 2D classes obtained (Fig.5.26.A, top right panel) appeared to be poorly resolved potentially due to the very low particle number available. Next, the preparation of PEGylated Cryo-EM grids was attempted to increase the particle density (120, 335).

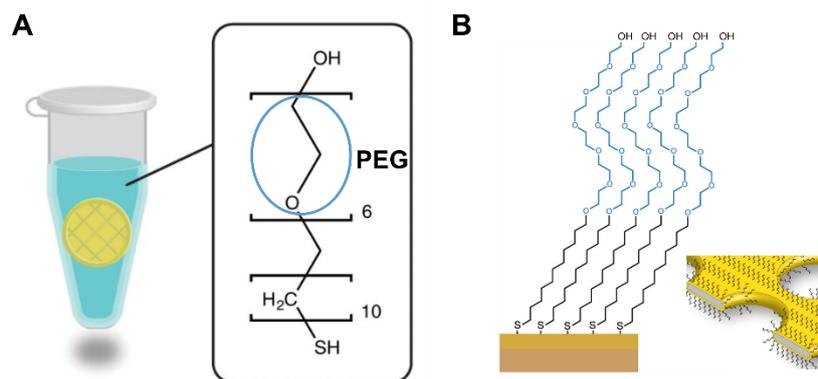


**Figure 5.26** **A.** Micrograph of the highest concentrated fraction (1.9 mg/ml) containing 0.04% w/v DDM in a 1.2/1.3 UltrAuFoil gold grid. The magnification used was x150,000. Some particles are shown inside red circles. 2D classes are shown on the top right panel. **B.** Micrograph showing protein aggregation at the edge of the grid hole (red arrows).

#### 5.4.8 Cryo-EM: PEGylated gold grids

To increase particle density, the surface of cryo-EM gold grids was coated with PEG groups. The treatment results in increased hydrophilicity compared to both plasma-cleaned amorphous carbon or gold film cryo-EM grids (120).

To perform the surface coating, gold grids were plasma cleaned and then immersed in a solution containing (11-Mercaptoundecyl)hexa(ethylene glycol) inside an anaerobic glovebox. The thiol groups bind to the gold surface and the molecules self-assembled forming a hydrophilic monolayer (Fig.5.27).

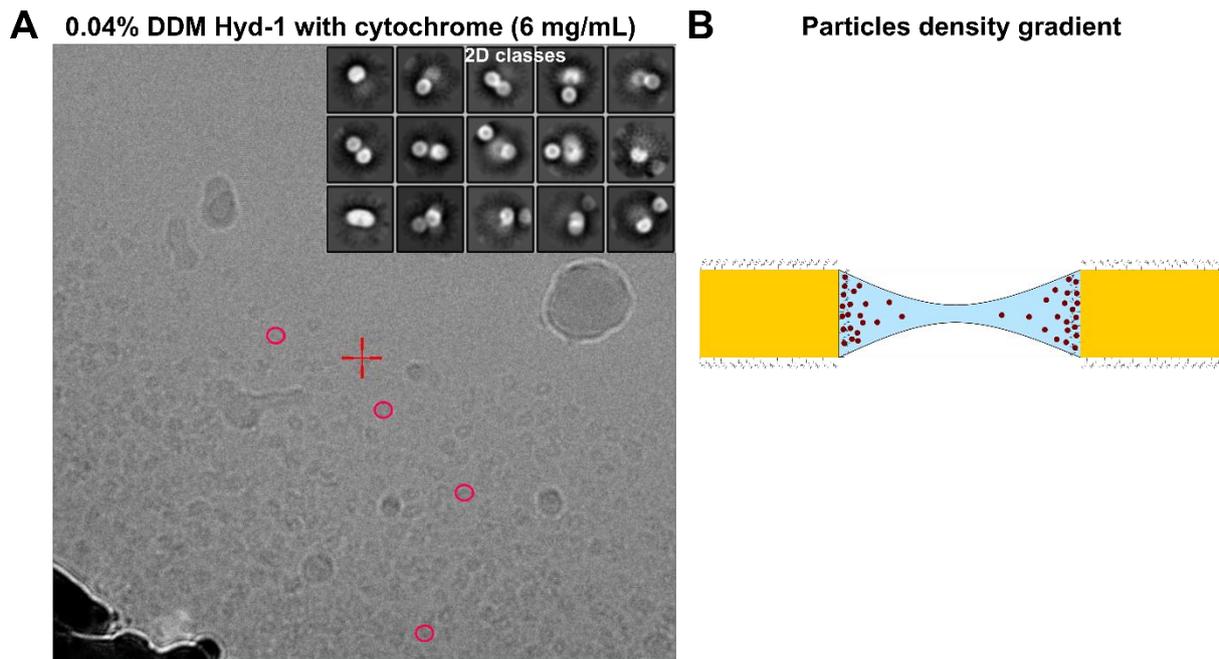


**Figure 5.27.** **A.** Schematic of a gold cryo-EM grid immersed in a solution containing (11-Mercaptoundecyl)hexa(ethylene glycol). **B.** The (11-Mercaptoundecyl)hexa(ethylene glycol) self-assembles forming a monolayer on top of the gold surface. Images adapted from (120, 336).

After about 48h the PEGylated gold grids were extracted and used for cryo-EM sample preparation (aerobic); concentrated strep-tagged Hyd-1 aliquotes (6 mg/mL) obtained after affinity purification in the presence of 0.04% DDM were defrosted and 2.5  $\mu$ L placed onto a 1.2/1.3 PEGylated UltrAuFoil grid selecting a blot time of 10 s with blot force at -5.

The Hyd-1 concentration on the grid greatly improved, however, a particle density gradient was observed with particle density decreasing towards the centre of the grid hole (Fig.5.28).

The 2D classes obtained from 50k particles (Fig.5.28.A, top right panel) appeared to be poorly resolved with an odd duplication effect potentially due to particle vertical overlap or other detergent effects. Therefore, it was decided to attempt increasing particle concentration simply by applying a high-concentration Hyd-1 sample on Quantifoil grids instead.

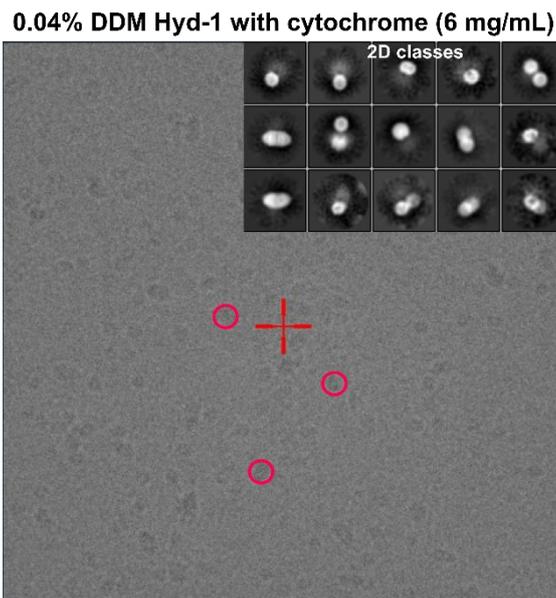


**Figure 5.28** **A.** Micrograph of a concentrated fraction (6 mg/mL) containing 0.04% w/v DDM in a 1.2/1.3 UltrAuFoil PEGylated grid. The magnification used was x240,000. Some particles are shown inside red circles. 2D classes are shown on the top right panel. **B.** Schematic showing how PEGylation treatment potentially affected the observed particle density gradient.

#### 5.4.9 Cryo-EM: concentrated Hyd-1 after affinity purification with DDM

6 mg mL<sup>-1</sup> concentrated Hyd-1 aliquots after affinity purification in the presence of 0.04% w/v DDM were defrosted and 2.5 µL placed onto a 1.2/1.3 Quantifoil grid selecting a blot time of 3 s with blot force at -5. The Hyd-1 concentration on the grid significantly improved (Fig.5.29) compared to using unconcentrated fractions (Chapter 5.4.7) and 77k particles were found in total.

The 2D classes obtained (Fig.5.29 top right panel) were very similar to the ones obtained from the PEGylated grids dataset with the characteristic strange image distortions and blurriness. It was concluded that this effect may have been caused by the increased detergent concentration resulting from using regular spin columns to concentrate the sample. In fact, it is common lab knowledge that detergents tend to be co-concentrated when regular spin columns are used. A subsequent attempt involved using a detergent removal spin column to concentrate the sample followed by the addition of LMNG.



**Figure 5.29** Micrograph of a concentrated fraction (6 mg/mL) containing 0.04% w/v DDM in a 1.2/1.3 Quantifoil grid. The magnification used was x240,000. Some particles are shown inside red circles. 2D classes are shown on the top right panel.

#### **5.4.10 Cryo-EM: Hyd-1 with detergent removal and addition of LMNG**

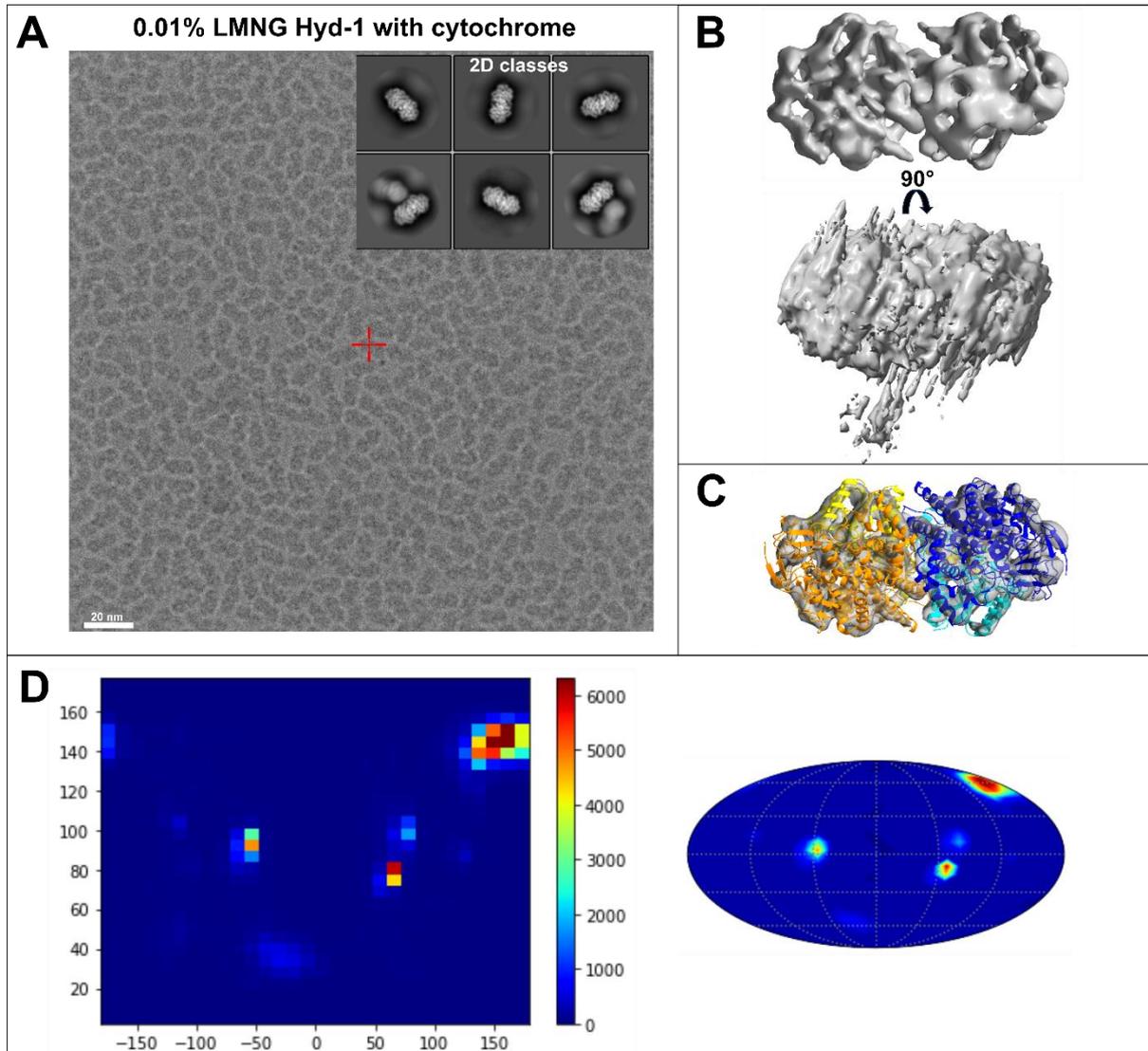
In another attempt Hyd-1 fractions after affinity purification were concentrated using a detergent removal spin column (Chapter 2.4.2) and then LMNG (0.01% w/v) was added.

The concentrated Hyd-1 sample in the presence of LMNG was used to prepare grids; 2.5  $\mu$ L were placed onto a 1.2/1.3 Quantifoil grid selecting a blot time of 2.5 s with blot force at -5. The hyd-1 concentration on the grid was excessively high (Fig.5.30), however, a sufficient number of isolated particles were found and used for further data processing.

The 2D classes obtained (Fig.5.30.A, top right panel) were similar to the ones obtained from the detergent-free datasets (Chapter 5.3.4) showing two main orientations by visual analysis. 120k particles after 2D classification were used to produce a 3D reconstruction (Fig.5.30.B) which displayed the ‘elongated’ artefacts characteristic of datasets with strong preferred orientation and extreme anisotropy (Fig.5.30.B, bottom image). Nevertheless, in the 3D reconstruction, a hint of the cytochrome in the expected position was observed (Fig.5.30.B, bottom image). Due to this effect, it was not possible to estimate the resolution of the map.

Overall this dataset could not be used for atomic model building of the Hyd-1 enzyme due to the strong preferred orientation (Fig.5.30.D), however, it was considered an improvement compared to previous results.

As a final attempt, anaerobic grid preparation was attempted using GDN (synthetic digitonin) and LMNG after initial solubilisation in DDM; no improvement was seen.



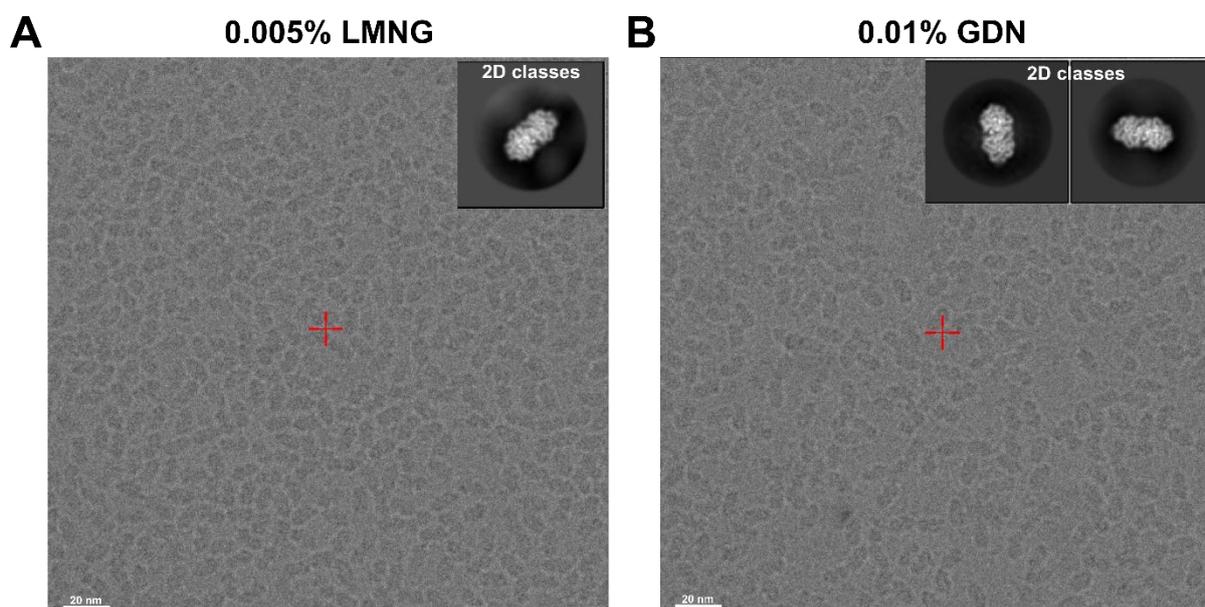
**Figure 5.30** **A.** Micrograph of a concentrated fraction containing 0.01% w/v LMNG in a 1.2/1.3 Quantifoil grid. The magnification used was x240,000. On the top right corner, the 2D classes are shown. **B.** Top view (from the opposite side of the cytochrome) and side view of the 3D reconstruction obtained from this experimental dataset. The side view (bottom image) shows a density potentially corresponding to the cytochrome subunit. The elongation of the map is an artefact due to preferred orientations. **C.** Top view of the reconstructed map fitted with the published PDB structure of Hyd-1. **D.** Angular distribution plot of all Hyd-1 particles that contributed to the reconstructed map.

### 5.4.11 Cryo-EM: Anaerobic Hyd-1 grids with LMNG and GDN

In another ‘final’ attempt, Hyd-1 fractions from size exclusion with 0.005% w/v LMNG (Fig.5.24.B) were concentrated and used to prepare grids inside an anaerobic glovebox. This was attempted as there was concern that the H<sub>2</sub>O<sub>2</sub> produced in the air-water interface during grid preparation could be damaging the complex, particularly at bioinorganic centres (337). Another set of grids was prepared anaerobically from fractions after affinity purification with buffers containing 0.01% w/v GDN.

In both cases, 2.5 µL of the samples were placed onto a 1.2/1.3 Quantifoil grid selecting a blot time of 2.5 s with blot force at -5. The Vitrobot was operated inside an anaerobic glovebox.

The 2D classes obtained (Fig.5.31.A, B top right panels) were identical to the ones obtained from the detergent-free datasets (Chapter 5.3.4) showing only one or two orientations. It was concluded that anaerobic preparation does not result in any improvement in particle orientations and neither does the use of GDN as a detergent.



**Figure 5.31** **A.** Micrograph of a Hyd-1 fraction containing 0.005% w/v LMNG in a 1.2/1.3 Quantifoil grid. **B.** Micrograph of a Hyd-1 fraction containing 0.01% w/v GDN in a 1.2/1.3 Quantifoil grid. The magnification used was x240,000. On the top right corners, the 2D classes are shown.

## 5.5 Conclusions

Both His- and Strep-tagged Hyd-1 proteins were used to make cryo-EM grids using different detergents (Triton X-100, DDM, LMNG, GDN), however, all strategies tested resulted in a strong preferred orientation of the particles. Hyd-1 seems to be particularly susceptible to air-water interface absorption even in the presence of the tested detergents. The strong preferred orientation tendency may also be a factor at play in the formation of stable ‘films’ on graphite electrodes when dimeric membrane-bound hydrogenases are used (as per Alison Parkin’s lab experience).

Overall, the best results were found using a strep-tagged complex, with the mild detergent LMNG, where a hint of the membrane-bound cytochrome was observed.

A possible strategy to overcome the preferred orientation challenge would be using stage tilt (338), however, this would result in a lower resolution structure (Chapter 1.8.8). As an alternative, other detergents or detergent analogues could be tested. In particular, detergent analogues such as the amphiphilic protein saposin (339), peptidics (340), or SMALPS (341) have been shown to result in remarkably stable isolated membranes.

## 6. Discussion and future perspectives

The work in this thesis has concerned the cryo-EM analysis of *Thermotoga maritima* hydrogenase (*TmHydABC*) and *E. coli* hydrogenase-1 (*Hyd-1*).

In the case of *TmHydABC*, a cryo-EM structure was determined showing a tetramer of trimer arrangement and CT flexible domains in each subunit (*HydA*, *HydB*, *HydC*) with at least one iron-sulfur cluster (171). Most likely the mobile domain in *HydA* ‘electrically connects’ the two *HydABC* protomers in a *Hyd(ABC)<sub>2</sub>* unit while the mobile domain in *HydB* is the binding site for *Fd* (as per a here presented *Fd*-bound structure). As only the mobile *HydB* subunit is conserved in homologous enzymes, it was proposed that only *HydB* is necessary for coupling. In addition, a Zn site in *HydB* was found which seems to act as a rigid hinge.

However, structure and cofactor arrangement of *TmHydABC* were shown to be incompatible with a thermodynamically bifurcating mechanism. In addition, upon further investigation, no experimental activity assay evidence was found in its support. Instead of forcing and claiming a thermodynamic bifurcation mechanism as previously done (171, 207), another approach has been taken and it was theorized that *TmHydABC* and similar trimeric hydrogenases could operate under a purely kinetic coupling mechanism. If confirmed, this would be the first example of this type of coupling, however, activity assays are required for definitive proof. In particular, pH and/or  $H_2$  concentration needs to be carefully chosen to ensure that the least exergonic reaction does not go to or near completion purely based on thermodynamical considerations.

Regarding *Hyd-1*, this thesis demonstrated the possibility of purifying *Hyd-1* in a single oligomeric state using a strep-tag approach. A Strep-tag ®II was introduced using CRISPR-Cas assisted genome editing at the N-terminus of *HyaB* (far from the delicate cytochrome subunit) conserving translational coupling. However, despite the efforts and multiple attempts, a cryo-EM structure of *Hyd-1* could not be obtained due to the strong preferred orientation observed. In future work, stage tilt can be attempted or alternative purification strategies may be tested using other combinations of detergents. In particular, detergent analogues such as the

amphiphilic protein saposin (339), peptidics (340), or SMALPS (341) have been shown to result in remarkably stable isolated membranes which may reduce preferred orientation issues. However, finding the right combination of detergents remains a ‘test and see’ strategy as no prediction tools exist at the moment to make more directed choices.

This strep-tagged Hyd-1 preparation could also be used for X-ray structural studies and, to increase protein yield, the strep-tagged strain (CF-001) is currently being edited so that Hyd-1 is expressed under the control of the promoter of the Lpp lipoprotein, which is one of the most abundant cellular protein (342). I am hoping that this approach maximises protein yield whilst minimizing loss of native conformation by controlling the level of expression; in particular, this promoter sequence can be modified to tune up and down expression as previously demonstrated by S. Inouye and M. Inouye’s work (342).

The improved Hyd-1 preparation developed here also sets the scene for further functional studies of the cytochrome domain to understand how this enzyme sustains its host. In particular, it will be possible to study how the binding of quinones (likely the rate-limiting step) affects enzyme activity.

In addition, the use of CRISPR-Cas editing has been here tested on Hyd-1 and was shown to be particularly efficient and less time-consuming compared to traditional genome editing techniques and can be therefore used for future work; for example for the tagging of Hyd-2 or Hyd-3, or to generate aminoacids mutations in Hyd-1 for functional studies. On a more general practical level, the CRISPR-Cas method can be used to more conveniently edit the bacterial genome in such was that proteins are overexpressed from the genome instead of from a plasmid. Industrially, this is of particular interest as the use of antibiotics for plasmid maintenance can be avoided. The use of antibiotics on an industrial scale is problematic both in terms of additional costs and environmental issues.

An overarching theme during my PhD has been the observation that while cryo-EM offers invaluable insights into how enzymes function, without complementary functional assays these are of limited use. Moving forward, it will be critical to develop better and more detailed functional descriptions of enzymes avoiding placing too much weight on structural studies.

During my PhD I also placed significant consideration on weighting the practical implications of research, in particular, here I analyzed the hydrogenase field as a whole and found to be of

limited commercial potential (Chapters 1.4 to 1.7) especially when factors such as costs and stability of hydrogenases are taken into considerations. I believe claims of real-world relevance should be more carefully evaluated and demonstrated when written or presented in academic contexts.

# 7. Appendix

**Table A.1** Groups and subgroups of the hydrogenases and examples of organisms where they are found. Adapted from (26).

<b>Group</b>	<b>Proposed function</b>	<b>Organisms</b>
<b>[NiFe]-hydrogenases</b>		
<b>Group 1: membrane-bound H<sub>2</sub>-uptake [NiFe]-hydrogenases</b>		
Group 1a: ancestral	Liberates electrons for sulphate, metal, organohalide and methanogenic heterodisulphide respiration. Includes [NiFeSe] variants.	<i>Desulfovibrio vulgaris</i> <i>Desulfomicrobium baculatum</i> <i>Methanosarcina mazei</i>
Group 1b: prototypical	Liberates electrons for sulphate, fumarate and nitrate respiration.	<i>Desulfovibrio gigas</i> <i>Wolinella succinogenes</i> <i>Helicobacter pylori</i>
Group 1c: Hyb type	Liberates electrons primarily for fumarate respiration. Possibly bidirectional.	<i>Escherichia coli</i> <i>Salmonella enterica</i>
Group 1d: oxygen tolerant	Electron input for aerobic respiration and oxygen-tolerant anaerobic respiration.	<i>Ralstonia eutropha</i> <i>Escherichia coli</i> <i>Aquifex aeolicus</i>
Group 1e: Isp type	Liberates electrons primarily for sulfur respiration. Possibly bidirectional.	<i>Allochromatium vinosum</i> <i>Aquifex aeolicus</i> <i>Thiocapsa roseopersicina</i>
Group 1f: oxygen protecting	Unresolved. May liberate electrons to reduce reactive oxygen species.	<i>Geobacter sulfurreducens</i> <i>Frankia</i> sp.
Group 1g: Crenarchaeota type	Unresolved. May liberate electrons primarily for sulfur respiration.	<i>Pyrodictium brockii</i> <i>Acidianus ambivalens</i>
Group 1h/5: Actinobacteria type	Scavenges electrons from tropospheric H <sub>2</sub> to sustain aerobic respiration during starvation.	<i>Ralstonia eutropha</i> <i>Mycobacterium smegmatis</i> <i>Streptomyces avermitilis</i>
<b>Group 2: Cytosolic H<sub>2</sub>-uptake [NiFe]-hydrogenases</b>		
Group 2a: Cyanobacteria type	Electron input for aerobic respiration and recycling H <sub>2</sub> produced by cellular processes (for example, nitrogenase, fermentation).	<i>Anabaena</i> sp. <i>Mycobacterium smegmatis</i> <i>Nitrospira moscoviensis</i>
Group 2b: HK linked	Senses H <sub>2</sub> and activates two-component cascade controlling hydrogenase expression.	<i>Ralstonia eutropha</i> <i>Rhodobacter capsulatus</i>
Group 2c: DGC linked (putative)	Unknown. Predicted to sense H <sub>2</sub> and induce cyclic di-GMP production.	Uncharacterised.
Group 2d: Aquificae type	Unknown. May generate reductant for carbon fixation or have a regulatory role.	<i>Aquifex aeolicus</i>
<b>Group 3: Cytosolic bidirectional [NiFe]-hydrogenases</b>		
Group 3a: F <sub>420</sub> coupled	Directly couples oxidation of H <sub>2</sub> to reduction of F <sub>420</sub> during methanogenesis. Reverse reaction may also occur. Includes [NiFeSe] variants.	<i>Methanothermobacter marburgensis</i> <i>Methanosarcina barkeri</i>
Group 3b: NADP coupled	Directly couples oxidation of NADPH to evolution of H <sub>2</sub> . May be reversible. Some complexes are proposed to have sulfhydrogenase activity.	<i>Pyrococcus furiosus</i> <i>Thermococcus kodakarensis</i> <i>Mycobacterium smegmatis</i>
Group 3c: HDR linked	Bifurcates electrons from H <sub>2</sub> to heterodisulphide and ferredoxin in methanogens without cytochromes.	<i>Methanothermobacter marburgensis</i>
Group 3d: NAD coupled	Directly interconverts electrons between H <sub>2</sub> and NAD depending on redox state.	<i>Anabaena</i> sp. <i>Ralstonia eutropha</i> <i>Thiocapsa roseopersicina</i>

<b>Group 4: Membrane-bound H<sub>2</sub>-evolving [NiFe]-hydrogenases</b>		
Group 4a: formate hydrogenlyases	Couples oxidation of formate to fermentative evolution of H <sub>2</sub> . Hyf-type complexes may translocate protons via antiporter modules.	<i>Escherichia coli</i> <i>Salmonella enterica</i>
Group 4b: Mrp linked	Couples oxidation of formate or carbon monoxide to proton reduction. Generates sodium-motive force via Mrp antiporter modules.	<i>Pyrococcus furiosus</i> <i>Thermococcus onnurineus</i> <i>Thermococcus kodakarensis</i>
Group 4c: CODH linked	Forms complex with carbon monoxide dehydrogenase to anaerobically respire CO using protons as terminal electron acceptors.	<i>Carboxydotherrmus hydrogenoformans</i> <i>Rhodospirillum rubrum</i>
Group 4d: Eha/Ehb type	Multimeric complexes that couple H <sub>2</sub> oxidation to ferredoxin reduction for anaplerotic (Eha) and anabolic (Ehb) purposes. H <sup>+</sup> /Na <sup>+</sup> driven.	<i>Methanococcus maripaludis</i>
Group 4e: Ech type	Couples ferredoxin oxidation to H <sub>2</sub> evolution. This process is physiologically reversible via H <sup>+</sup> /Na <sup>+</sup> translocation.	<i>Methanasarcina barkeri</i> <i>Desulfovibrio gigas</i> <i>Thermoanaerobacter tengcongensis</i>
Group 4f: Ehf type (putative)	Unknown. May couple oxidation of a one-carbon compound to proton reduction concurrent with proton translocation. Related to Ehr complexes.	Uncharacterised.
<b>Group 5: High-affinity H<sub>2</sub>-uptake [NiFe]-hydrogenases</b>		
Group 5	Oxidases atmospheric H <sub>2</sub> to sustain electron input into the respiratory chain during energy starvation. Characterized by high affinity to H <sub>2</sub> , oxygen insensitivity, and thermostability.	<i>Streptomyces avermitilis</i> <i>Mycobacterium smegmatis</i> <i>Rhodococcus equi</i>
<b>[FeFe]-hydrogenases</b>		
Group A1: prototypical	Couples oxidation of ferredoxin to fermentative or photobiological evolution of H <sub>2</sub> .	<i>Clostridium pasteurianum</i> <i>Desulfovibrio desulfuricans</i> <i>Chlamydomonas reinhardtii</i>
Group A2: glutamate synthase linked (putative)	Unknown. Predicted to transfer electrons from H <sub>2</sub> to NAD, generating reducing equivalents for glutamate synthase.	Uncharacterised.
Group A3: bifurcating	Reversibly bifurcates electrons from H <sub>2</sub> to ferredoxin and NAD in anaerobic bacteria.	<i>Thermotoga maritima</i> <i>Acetobacterium woodii</i> <i>Moorella thermoacetica</i>
Group A4: formate dehydrogenase linked	Couples formate oxidation to evolution of H <sub>2</sub> . Some can also bifurcate electrons from H <sub>2</sub> to ferredoxin and NADP.	<i>Clostridium autoethanogenum</i>
Group B: ancestral (putative)	Unknown. May couple oxidation of ferredoxin to fermentative evolution of H <sub>2</sub> .	Uncharacterised.
Group C: sensory (putative)	Unknown. Predicted to sense hydrogen and induce cascades via co-transcribed regulatory elements, for example, kinases and phosphatases.	<i>Thermoanaerobacterium saccharolyticum</i> <i>Ruminococcus albus</i>
<b>[Fe]-hydrogenases</b>		
Hmd: methenyl-H4MPT dehydrogenase	Couples oxidation of H <sub>2</sub> to reduction of 5,10-methenyltetrahydromethanopterin in methanogens. Physiologically-reversible and important during nickel limitation.	<i>Methanocaldococcus jannaschii</i> <i>Methanothermobacter thermoautotrophicum</i>

## Supplementary text A.1

### *Commercial evaluation (cell-free enzymatic pathway for the production of H<sub>2</sub>)*

A hydrogen production cost below ~2 USD/kg H<sub>2</sub> is essential to ensure a competitive hydrogen selling price of 2–3 USD/kg H<sub>2</sub>. Here, I will show that even with substantial underestimations in the production cost, reaching this threshold is unfeasible.

As a first major underestimation, I neglect all the production costs except those of feedstock, enzymes, and recycling cofactors. I will demonstrate that we are so far from economic feasibility that even just considering these three variables is sufficient.

As a second underestimation, I imagine that the feedstock is pure ready-to-use glucose and its price is the same as cellulosic biomass (one of the cheapest feedstocks). Cellulosic biomass can be purchased at a representative price of USD ~0.075 per 1 kg of dry mass (343, 344). Assuming that H<sub>2</sub> is produced at stoichiometric potential, then it can be easily derived that the feedstock cost contributes ~0.56 USD for every kg of H<sub>2</sub> produced.

As a third underestimation, I imagine that all the enzymes in the pathway have an infinite lifespan and only the hydrogenase enzyme needs to be replaced. Hydrogenases are likely to have the biggest contribution to the total enzyme cost both for the stoichiometry of their reaction (hydrogenase: 1 catalytic cycle = 1 molecule of H<sub>2</sub> vs. rest of the enzymes: 1 catalytic cycle = 6-12 molecules of H<sub>2</sub>\*) and their significant lower stability and oxygen sensitivity. The catalytical potential of the hydrogenase is numerically expressed by its Total Turnover Number (**TTN<sub>hyd</sub>**). The TTN is defined as the ratio of moles of product generated divided by the moles of biocatalyst used in a reaction; an equivalent definition is the number of catalytic events performed by one active site of one molecule of the enzyme during its lifespan (345). To simplify this model I assume that the chosen hydrogenase has a single catalytic site:

$$TTN_{Hyd} = \frac{\text{moles H}_2 \text{ produced}}{\text{moles hydrogenase}} \quad (A.1)$$

Therefore  $\frac{1}{TTN_{Hyd}}$  moles of hydrogenase are needed to produce one mole of H<sub>2</sub>.

By combining the  $TTN_{Hyd}$  with the molecular mass in kDa ( $M_{hyd}$ ) and the unit cost ( $P_{hyd}$ , USD/kg) of the hydrogenase enzyme, it is possible to determine its cost contribution in the production of 1kg of  $H_2$  as:

$$\frac{1}{TTN_{Hyd}} M_{hyd} \cdot P_{hyd} \cdot 500 \text{ USD/kgH}_2 \quad (A.2)$$

Most hydrogenases have a molecular mass above 100 kDa, however, there are a few that have molecular masses around 50-60 kDa (181). As a fourth underestimation, I choose  $M_{hyd} = 50$  kDa.

A reasonable  $P_{hyd}$  for a hydrogenase enzyme would be in the range of 1,000-4,000 USD per kg enzyme protein (346). This is based on the fact that hydrogenases are among the most expensive enzymes to produce due to their complex cofactors, low stability, and oxygen sensitivity. Therefore we are looking at a price that is in the mid to above range of that of bulk commodities industrial enzymes (346).\*\* As a fifth underestimation I select the lowest value of this range which is 200 USD/kg protein.

By substituting the above values in Formula A.2, it can be obtained that the underestimated cost contribution of the hydrogenase in the production of 1kg of  $H_2$  is:

$$\frac{1}{TTN_{Hyd}} 5,000,000 \text{ USD/kgH}_2 \quad (A.3)$$

As a sixth underestimation, I assume that no phosphate groups are used and only one recycling cofactor is utilized. This cofactor acts as a substrate for the reduction of protons by donating two electrons.

I can then combine the TTN of the recycling cofactor ( $TTN_R$ , moles of product produced per mol of cofactor used) with its molecular mass in g/mol ( $M_R$ ) and its unit cost ( $P_R$ , USD/g) and determine its cost contribution in the production of 1kg of  $H_2$  as:

$$\frac{1}{TTN_R} M_R \cdot P_R \cdot 500 \text{ USD/kgH}_2 \quad (A.4)$$

Here I choose NADPH as a recycling cofactor (as per Zhang *et al.* work), therefore  $M_R$  is equal to 744 g/mol. Based on Zhang *et al.* economical model,  $P_R$  for NAPH and NADH is around 19 USD/kg and 1.9 USD/g, respectively (48). However, it was found that Chinese suppliers sell NADPH at a reference price of 1-2 USD/g (Alibaba.com). A minimum value of 1 USD/g is

here chosen. I can substitute the values of  $M_R$  and  $P_R$  in Formula A.4 and obtain the underestimated cost contribution of the recycling cofactor in the production of 1kg of  $H_2$ :

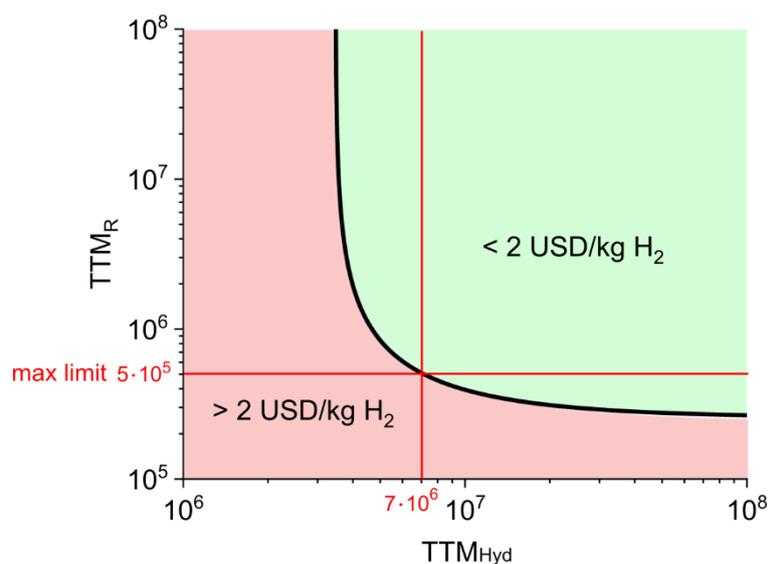
$$\frac{1}{TTN_R} 370,000 \text{ USD/kgH}_2 \quad (A.5)$$

By combining the cost of feedstock, hydrogenase enzyme (Formula A.3), and recycling cofactor (Formula A.5), it can be obtained that the by far underestimated production cost of 1 kg of  $H_2$  is:

$$0.56 + \frac{1}{TTN_{Hyd}} 5,000,000 + \frac{1}{TTN_R} 370,000 \text{ USD/kgH}_2 \quad (A.6)$$

We want this price (Formula A.6) to be below  $\sim 2$  USD/kg  $H_2$ . This can be achieved when the combination of values of  $TTN_{Hyd}$  and  $TTN_R$  falls within the green area in the plot of Figure A.1. We can immediately notice, that even for this by far underestimated model the TTM required are particularly high. First of all, we can see that  $TTN_{Hyd}$  must be in the order of millions or higher, more precisely above  $\sim 3,500,000$  if the cost of recycling cofactors is completely ignored. This would be particularly challenging for hydrogenase enzymes which tend to be among the least stable enzymes. To further convince ourselves, we can look at much more stable industrially used enzymes and see that only a few of the most stable and optimized enzymes reach TTMs of this order of magnitude. For example, in the production of high-fructose corn syrups, 1kg of the optimized industrial enzyme glucose isomerase (immobilized to increase stability) can convert  $\sim 4,500$  kg of glucose into fructose (347, 348) meaning its TTM is equal to  $\sim 1,250,000$ .

When we look at the  $TTN_R$  we can see that the situation is even less favourable. The TTN of NADH and NADPH can be considered to be  $\sim 1,000$  (349, 350). However, recent advances show systems in which the TTN of these cofactors can reach significantly higher values ranging from 50,000 to 500,000 (40, 350, 351). A  $TTN_R$  of 500,000 can be considered the upper limit of feasibility (as of yet this has only been reported in a glutathione-based recycling system (350)) resulting in a minimum  $TTN_{Hyd}$  of 7,000,000 (Fig. A.1) which, as discussed above, would be too high even for the most stable industrial enzymes. I have therefore demonstrated that cell-free pathways for the production of hydrogen are unlikely to be economically viable unless the protein production cost is substantially decreased and major breakthrough discoveries are made in enzyme stabilization.



**Figure A.1.** Plot showing the ranges of Total Turnover Numbers (TTN) of the enzyme hydrogenase (x-axis) and of the recycling cofactor NADPH (y-axis) that are compatible with a hydrogen production cost below  $\sim 2$  USD/kg  $H_2$  (green area) in an already substantially underestimated model. The plot is based on Formula A.6.

\* All the reactions in the enzymatic pathway are connected. We can therefore count how many  $H_2$  molecules end up being produced for each catalytical cycle in any enzyme of the pathway.

\*\* We should note that there are some exceptions to this range, for example, cellulase is far less expensive with production costs in the range between 3.80-8.80 USD/kg depending if it is an on-site or off-site production (352). However, cellulase is a very simple secretory (makes the purification less expensive) one-subunit enzyme that does not contain cofactors and, therefore cannot be used as a reference.

## Supplementary text A.2

### *Commercial evaluation (electro enzymatic hydrogen production)*

Any hydrogen produced using electrical energy (regardless of method) cannot currently compete with the hydrogen produced by steam methane reforming (sold at ~1.5 USD/kg H<sub>2</sub>) even if 100% of the electrical energy is converted into the chemical potential of H<sub>2</sub> molecules: Hydrogen gas has an energy density of approximately 120 MJ/kg, therefore 120 MJ of electricity is needed to produce 1kg of H<sub>2</sub> (100% energy conversion). The world average price of electricity is 0.16 USD per kWh, however, for this analysis, I choose a lower value of 0.08 USD/kWh (0.02 USD/MJ). This price is close to that found in China and Russia. With this value, it can be obtained that 2.7 USD of electricity are needed to produce 1 kg of H<sub>2</sub> in the ideal (but not realistic) case of 100% energy conversion.

Regardless of this result, I continue the analysis, as this will allow us to draw some comparisons with other hydrogen-producing methods that do not utilize fossil fuels and to understand the important effect of enzyme stability.

In particular, I want to analyse realistic future potentials and for this purpose, a comparison with standard electrolysis can be very valuable. Here I will only consider the production cost contributions of electricity and enzymes excluding important contributions such as costs of electron mediators, electrodes, surface treatments, maintenance of an anaerobic environment, etc. Furthermore, I assume a 100% energy conversion rate vs. ~70% of standard electrolysis. I will show that even with a super-efficient energy conversion of 100% the enzyme costs are sufficient to drive the production costs too far from those of standard electrolysis.

Similarly to the previous commercial analysis, I will utilize the Total Turnover Number to build the model. The catalytical potential of the hydrogen-producing enzyme (typically a hydrogenase) is numerically expressed by its Total Turnover Number (**TTN<sub>hyd</sub>**) which was defined in Formula A.1.

Therefore  $\frac{1}{\text{TTN}_{\text{Hyd}}}$  moles of hydrogenase are needed to produce one mole of H<sub>2</sub>.

By combining the **TTN<sub>Hyd</sub>** with the molecular mass in kDa (**M<sub>hyd</sub>**) and the unit cost (**P<sub>hyd</sub>**, USD/kg) of the hydrogenase enzyme, it is possible to determine its cost contribution in the production of 1kg of H<sub>2</sub> as:

$$\frac{1}{TTN_{Hyd}} M_{hyd} \cdot P_{hyd} \cdot 500 \text{ USD/kgH}_2 \quad (A.7)$$

Therefore, the cost contribution of electricity and enzymes in electro-enzymatic synthesis is:

$$2.7 + \frac{1}{TTN_{Hyd}} M_{hyd} \cdot P_{hyd} \cdot 500 \text{ USD/kgH}_2 \quad (A.8)$$

If these production cost contributions alone are above ~8 USD, around the top range of the current selling price of hydrogen produced with standard electrolysis (6-8 USD/kg H<sub>2</sub> (55)), then it is reasonable to assume that electro-enzymatic synthesis will be unlikely to outcompete electrolysis in the future energy competitor landscape. In particular, in standard electrolysis, electricity has the major contribution to the hydrogen production cost, followed by CapEX prices (i.e. physical assets, patents, etc.) which contribute to more than 50% of the remaining costs (33). This further confirms my choice of 8 USD as a very generous upper limit to assess future competitiveness.

For electricity and enzyme costs to be below ~8 USD it can be derived that:

$$TTN_{Hyd} > M_{hyd} \cdot P_{hyd} \cdot 94 \text{ USD/kgH}_2 \quad (A.9)$$

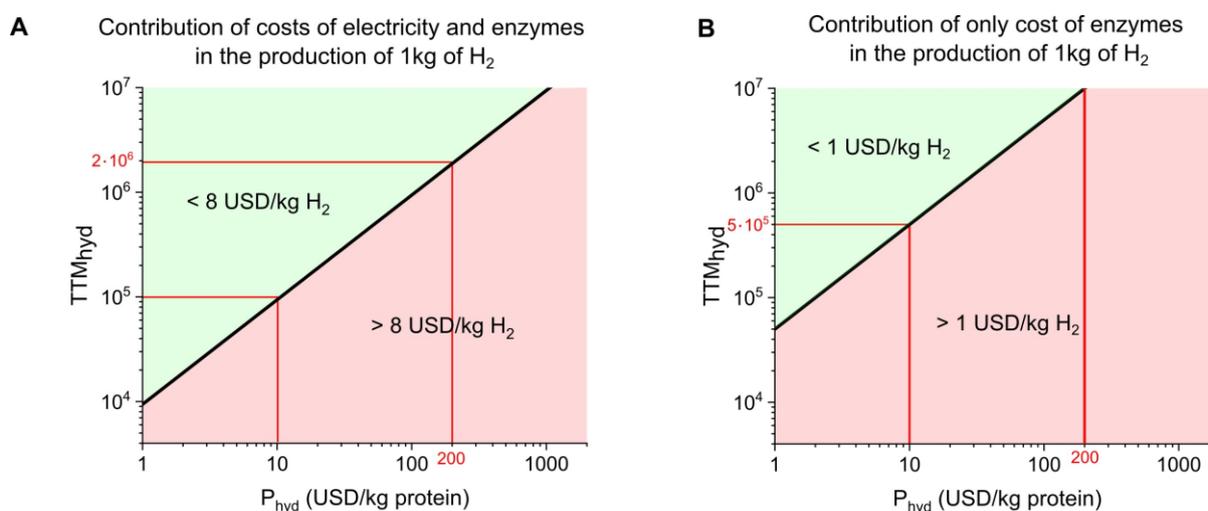
The molecular mass in kDa (**M<sub>hyd</sub>**) of the hydrogenase can be considered fixed as this does not vary a lot among hydrogenases. In particular here with **M<sub>hyd</sub>** we are referring to a protomer of the enzyme with one active site (as per the definition of TTM) and therefore even large hydrogenases that are complexes of multiple protomers with multiple active sites still have similar **M<sub>hyd</sub>**. Typically **M<sub>hyd</sub>** is in the range of 50-200 kDa (56), however, it is significantly less common for hydrogenases to have a **M<sub>hyd</sub>** below 100 kDa. I here consider **M<sub>hyd</sub>** to be 100 kDa which is lower or very close to the molecular mass of most hydrogenase protomers (e.g. CpI 134 kDa, 90 kDa protomer of Hyd-1 excluding cytochrome, etc.).

With **M<sub>hyd</sub>** = 100 kDa, I let the unit cost (**P<sub>hyd</sub>**, USD/kg protein) and the **TTN<sub>Hyd</sub>** of the hydrogenase vary to see whether it is feasible to achieve a cost contribution of electricity and enzymes below 8 USD (Formula A.9). These results are plotted in Fig. A.2.a. which shows that using a reasonable **P<sub>hyd</sub>** for a hydrogenase enzyme (1,000-4,000 USD per kg enzyme protein) requires **TTN<sub>Hyd</sub>** values above 10,000,000. This is extremely high and not compatible with even the TTM values of the most stable and optimized industrial enzymes (see Supplementary text A.2 above). Even using an underestimated price of 200 USD/kg protein, as in the previous

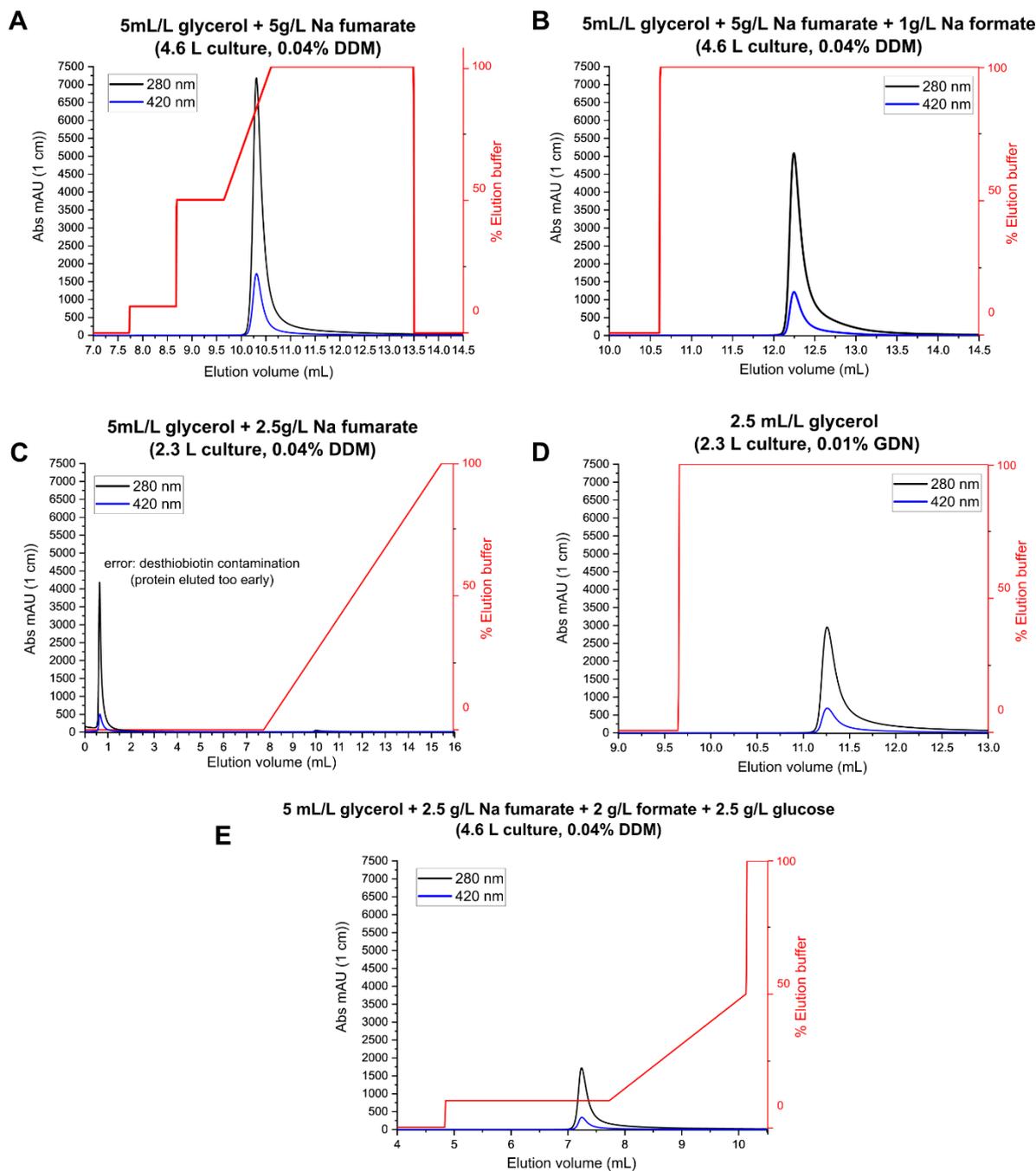
analysis, still requires too high  $TTN_{Hyd}$  value above 2,000,000. A more achievable  $TTN_{Hyd}$  of less than 100,000 requires that the  $P_{Hyd}$  is below 10 USD/kg protein. As discussed in the economic evaluation of cell-free enzymatic pathways, a threshold of 10 USD/kg is highly unrealistic, especially for less stable enzymes containing complex cofactors such as hydrogenases.

It, therefore, seems that for the foreseeable future, electro-enzymatic hydrogen production will face significant challenges requiring major breakthroughs in enzyme stabilization and production technologies. In particular, if we want to maintain the sole cost contribution of enzymes below 1 USD per kg of  $H_2$  produced (a reasonable threshold above which claiming an electro-enzymatic synthesis is investible would be of little sense) then we see that even with an extremely low protein production price of 10 USD/kg the TTN of the enzyme would need to be above 500,000 (Fig. A.2.b).

To conclude, I have demonstrated that the electro-enzymatic synthesis of hydrogen, like all electricity-based methods, is currently not commercially competitive to SMR. I have also shown that electro-enzymatic synthesis is by far less competitive than standard electrolysis.



**Figure A.2.** a. Plot showing the ranges of Total Turnover Numbers (TTN) of the hydrogenase enzyme (y-axis) and its unit cost in USD/kg protein (x-axis) that are compatible with a cost contribution of electricity and enzymes of less than 8 USD per kg of  $H_2$  produced (green area). The plot is based on Formula A.9. b. Plot showing the ranges of Total Turnover Numbers (TTN) of the hydrogenase enzyme (y-axis) and its unit cost in USD/kg protein (x-axis) that are compatible with a cost contribution of only the hydrogenase of less than 1 USD per kg of  $H_2$  produced (green area).



**Figure.A.3** Purification of Hyd-1 using a StrepTrap affinity column. Chromatograms showing elution volume against UV absorbance at 280 nm (black) and 420 nm (blue) and gradient of elution buffer (red). **A.** From 4.6 L growth in LB medium supplemented with 5 mL/L glycerol and 5 g/L of sodium fumarate. Buffers containing 0.04% DDM. **B.** From 4.6 L growth in LB medium supplemented with 5 mL/L glycerol, 5 g/L of sodium fumarate, and 1 g/L sodium formate. Buffers containing 0.04% DDM. **C.** From 2.3 L growth in LB medium supplemented with 5 mL/L glycerol and 2.5 g/L of sodium fumarate. Buffers containing 0.04% DDM. **D.** From 2.3 L growth in LB medium supplemented with 2.5 mL/L glycerol. Buffers containing 0.01% GDN. **E.** From 4.6 L growth in LB medium supplemented with 5 mL/L glycerol, 2.5 g/L sodium fumarate, 2 g/L sodium formate, and 2.5 g/L glucose. Buffers containing 0.04% DDM.

## Sequences (Hyd-1)

*Escherichia coli* K-12 MG1655 genome sequence (GenBank: U00096.3) was used to retrieve the full operon sequence.

### HyaA (signal peptide in blue)

MNNEETFYQAMRRQGVTRRSFLKYCSLAATSLGLGAGMAPKIAWA (cleave) LENKPRI PVVWIHGLECTCCTE  
SFIRSAHPLAKDVILSLISLDYDDTLMAAAGTQAEVFEDEITQYNGKYILAVEGNPPLGEQGMFCISSGRPFIE  
KLKRAAAGASAI IAWGTCASWGCVQAARPNPTQATPIDKVIITDKPI IKVPGCPPIPDVMSAIITYMVTFDRLPDV  
DRMGRPLMFYQRIHDKCYRAHFDA GEFVQSWDDDAARKGYCLYKMGCKGPTTYNACSSSTRWNDGVSFPIQSGH  
GCLGCAENGFWRGFSFYSRVVDIPQMGTHSTADTVGLTALGVVAAA VGVHAVASAVDQRRRHQQPTETEHQPGN  
EDKQA

### HyaB

MSTQYETQGYTINNAGRRLVVDPI TRIEGHMRCEVNINDQNVITNAVSCGTMFRGLEIILQGRDPRDAWAFVERI  
CGVCTGVHALASVYAIEDAIGIKVPDNANI IRNIMLATLWCHDHLVHFYQLAGMDWIDVLDALKADPRKTSELAQ  
SLSSWPKSSPGYFFFDVQNRLLKFFVEGGQLGIFRNGYWGHPQYKLPPEANLMGFAHYLEALDFQREIVKIHAVFGG  
KNPHPNWIVGGMPCAINIDESGAVGAVNMERLNLVQSI ITRTADF INNVMI PDALAIGQFNKPWSEIGTGLSDKC  
VLSYGAFPDIANDFGEKSLMPGGAVINGDFNNVLPVDLVDPPQVQEFVDHAWYRYPNDQVGRHPFDGITDPWYN  
PGDVKGSDTNIQQLEQERYSWIKAPRWRGNAMEVGLPARTLIAYHKGDAATVESVDRMMSALNPLSGIQSTLG  
RILCRAHEAQWAAGKLYFFDKLMTNLKGNLATASTEKWEPATWPTECRGVGFTEAPRGALGHWAAIRDGKIDL  
YQCVVPTTWNASPRDPKGQIGAYEAALMNTKMAIPEQPLEILRTLHSFDPCLACSTHVLGDDGSELISVQVR

### HyaC

MQQKSDNVVSHYVFEAPVRIWHWLTVLCMAVLMVTGYFIGKPLPSVSGEATYLFYMGYIRLIHFSAGMVFTVLL  
MRIYWAFVGNRYSRELFIVPVWRKSWWQGVWYERWYLF LAKRPSADIGHNPIAQAAMFGYFLMSVFMIIITGFAL  
YSEHSQYAI FAPFRYVVEFFYWTGGNSMDIHSWHRLGMWLGAFVIGHVYMALREDIMSDDTVISTMVNGYRSHK  
FGKISNKERS

### HyaABCDEF operon

**TAT:** light orange

**HyaA:** orange

**HyaB:** yellow

**HyaC:** light red

**HyaD:** light grey

**HyaE:** green

**HyaF:** grey

ATGAATAACGAGGAAACATTTTTACCAGGCCATGCGGCGTCAGGGCGTTACCCGGCGCAGCTTTCTCAAATATTGT  
AGTCTGGCTGCCACGTCGCTGGGATTAGGCGCGGAATGCCACCAAAGATTGCCTGGGCG-CTGGAGAACAAACC  
GCGCATTCCGGTGGTATGGATCCACGGTCTGGAATGCACCTGCTGTACCGAATCTTTTATCCGCTCCGCTCACCC  
ACTGGCGAAGGACGTCATCCTTTCCCTGATTTCCCTCGATTACGACGATACTTTGATGGCTGCCGCCGGAACCCA  
GGCGGAAGAAGTCTTTGAAGACATCATCACGCAATACAATGGCAAAATATATCCTCGCAGTAGAAGGTAATCCGCC  
GCTGGGCGAGCAGGGGATGTTCTGTATCAGCAGCGGTGACCGTTTATTTGAGAAACTCAAACGTGCCGCTGCCGG  
AGCCAGCGGATTATCGCCTGGGGAACCTGCGCGTCTTGGGGCTGCGTGCAGGCCGCGGACCCAATCCGACGCA  
GGCAACGCCTATCGACAAAGTCATCACCGACAAACCCATTTATCAAAGTACCTGGCTGCCCGCCGATCCCGGATGT  
GATGAGCGCCATCATTACTTACATGGTGACCTTTGATCGCTTGCCAGATGTCGACAGAATGGGCCGTCCGCTGAT  
GTTCTATGGTCAGCGAATCCACGATAAATGCTATCGCCGCGCCACTTCGACCGCGAGAGTTCGTCCAGAGTTG  
GGATGATGACGCTGCCCGCAAAGGTTACTGCCTGTACAAAATGGGCTGCAAAGGGCTACCACCTATAACGCCTG

TTCTCCACACGCTGGAATGATGGCGTTTCTTTCCCAATCCAGTCTGGTCACGGCTGCCTGGGCTGTGCGGAAAA  
TGGTTTTCTGGGATCGCGGTTCGTTCTACAGCCGCGTGGTTCGATATTCGCAAATGGGTACTCATTCACCGCCGA  
TACCGTTCGGTTTAAACCGCGCTTGGCGTGGTGGCAGCGGCTGTTGGTGTGCACGCAGTCGCCAGCGCCGTTGACCA  
GCGCAGACGTATAACAGCAACCTACAGAAACCGAACATCAGCCAGGCAATGAGGATAAAACAGGCATGACT  
CAGTACGAAACTCAGGGATACACCATCAATAATGCCGGACGCCGCTGGTGGTGCACCCGATTACGCGCATCGAA  
GGCCACATGCGCTGCGAAGTGAATATTAACGATCAGAATGTGATCACCAATGCCGTCTCTGCGGCACCATCTTT  
CGCGGGCTGGAGATCATCCTACAAGGGCGCGACCCGCGCGATGCGCGGGCGTTCGTTGAACGTATCTGCGGCGTC  
TGTACTGGCGTACACGCCCTGGCTTCGGTTACGCCATCGAAGATGCTATCGGTATTAAGTGCCTGGACAACGCC  
AATATCATCCGCAACATTATGCTGGCAACGCTCTGGTGCCACGATCATCTGGTGCACCTTCTATCAGCTTGGCGGG  
ATGGACTGGATCGATGTGTTAGATGCGCTGAAAGCCGACCCGCGGAAAACTCCGAACTGGCGCAAAGTCTCTCC  
TCTTGGCCGAAATCATCCCCTGGCTATTTCTTCGAGCTACAAAACCGCTGAAAAAATTTGTTGAAGGCGGGCAG  
TTGGGGATCTTCCGCAATGGCTACTGGGGGCACCCGAGTACAAAACGCGCCAGAAAGCTAACCTGATGGGCTTT  
GCCACTATCTCGAAGCTCTCGATTTCCAGCGTGAAATTGTCAAAAATCCACGCGTCTTTGGCGGTAAAAACCCG  
CATCCAAACTGGATTGTCGGCGGGATGCCTTGCGCCATCAACATTGACGAAAGCGGCGCGGTGGGGCAGTCAAT  
ATGGAACGCTGAACCTGGTGCAGTCAATTATCACCCGCACGGCGGACTTCATTAACAACGTGATGATCCCCGAC  
GCCTTAGCCATCGGTGAGTTCAACAAAACCGTGAGGCGAAAACTGGCAGTGGTCTTTCTGATAAAATGCGTCTCAGC  
TACGGCGCATTCCCGGATATTGCCAACGACTTTGGCGAGAAAAAGTCTGCTGATGCCTGGCGGCGCGGTGATTAAC  
GGCGACTTCAACAATGTGCTGCCAGTGGATTGGTTGATCCGCAGCAGGTGCAGGAGTTTGTGACCACGCCCTGG  
TATCGATATCCCAACGATCAGGTGGGGCGTATCCGTTGATGGCATCACCGACCCGTGGTACAACCCCGGCGAT  
GTCAAAGGCAGCGATACCAACATTCAGCAGCTGAATGAACAGGAACGCTACTCGTGGATCAAAGCGCCACGCTGG  
CGCGGTAACGCGATGGAAGTGGGGCCGCTGGCGCGCACGTTAATCGCTTATCACAAGGCGATGCTGCGACCGTT  
GAGTCGGTTCGATCGCATGATGTGCGCGTTGAACCTGCCGCTTTCCGGTATCCAGTCAACGTTAGGCCGCATTTTG  
TGCCGCGCGCACGAAGCGCAGTGGGCCGAGGTAAGTTGCAGTATTTCTTCGACAAGCTGATGACCAACCTGAAA  
AACGGCAATCTCGCCACTGCTTCCACGGAAAAATGGGAACCTGCAACCTGGCCGACAGAGTCCCGTGGTGTGCGT  
TTTACCGAAGCGCCGCGCGGGCGTTAGGCCACTGGGCCGCAATTCGCGATGGCAAGATTGATCTTACCAGTGC  
GTGGTCCCGACCCAGTGAACCGCCAGCCCGCGGATCCCAAAGGGCAGATTGGCGCTTATGAAGCGGCGCTGATG  
AACACAAAATGGCCTCCCCGAGCAACCCGCTGGAGATCTGCGTACTGTCACAGCTTTGACCCGTGCCCTGCC  
TGTTCAACACACGCTGCTGGGCGACGAGCTGATCTCCGTCGAGGTGCGTTAAACAGCGAAGGAGAAATC  
ATCATGCAACAGAAAAGCGACAACGTTGTGACCCACTATGCTTTTGAAGCGCCAGTGCGCATCTGGCACTGGTTG  
ACGGTGTATGATGCGGCTGTTGATGGTACCAGGATACTTTATCGGCAAGCCGCTACCTTCCGTGACGCGCGAG  
GCGACGTATCTGTTCTATATGGGCTACATCAGGTTAATTCACCTTCAGCGCCGGGATGGTTTTTACCCTGGTTTTG  
CTGATGCGGATCTACTGGGCTTTTGTGGCAATCGATACTCCCGGAGCTGTTTATCGTGCCGGTATGGCGTAAA  
AGCTGGTGGCAGGGCGTGGTATGAAATCCGCTGGTATCTGTTTCTGGCAAAACGTCGAGTGCCGATATAGGC  
CATAATCCCATCGCCAGGCGGCGATGTTCCGCTATTTCTGATGTCGGTCTTTATGATCATCACTGGTTTTTGGC  
CTGTACAGCGAACACAGCCAGTACGCTATTTTTGCGCCGTTCCGTTATGTGGTGAATTTTTCTACTGGACGGGT  
GGCAACTCAATGGACATTCACAGCTGGCATCGGCTGGGGATGTGGCTGATTGGCGGTTTTGTGATCGGTGATGTC  
TACATGGCGCTGCGTGAAGACATCATGTCCGACGACACGGTATCTCCACCATGGTCAACGGCTACCGTAGCCAC  
AAATTTGGCAAAATAAGTAACAAGGAGCGTTTCATGAGCGAGCAACCGCTGGTGGTGCATGGGGCTGGGCAACCTGC  
TGTGGGCCGATGAAGGCTTCGGCGTGCGGGTGGCGGAACGGCTGTATGCCCATTACCCTGGCCCGAGTATGTGG  
AGATTGTGATGCGGCTACTCAGGGACTGAACTTGTGGGGTATGTGAAAGCGCCAGCCATCTGTTGATTCCTCG  
ATGCCATTGACTACGGGCTGGAACCTGGAACGCTGCGAACCTATGCCGGAGAACGCATTCGGGCTTATCTCAGCG  
CGAAGAAAATGAGCCTGCATCAGAACAGTTTCTCCGAAAGTGTGGCGCTGGCGGATATCCGCGGACATCTGCCAG  
CACATATTGCCCTCGTGGTCTGCAACCCGCAATGCTCGACGACTACGGCGGTAGCCTGAGCGAACTGGCACGGG  
AGCAACTGCCCGCTGCGGAACAGGGCGGCGCTGGCGCAGCTTGTGCTGGTGGGAATTTGTGCCGAACCCGGCTAATG  
AATCGCGCTGTCTCAATTATGACTGTCTGTGATGGAATAATACGAAGGCGTTGCTTGCGCCAGTACCGGATGA  
CACAGGAGGAGCAGGGATGAGCAACGACACGCCAATTTGATGCGTTGTGGCAACGAATGCTGGCGCGCGGCTGGAC  
GCCAGTCAAGTGAATCCCGTCTTACGACTGGCTTACGCAAGCGCCAGACGGCGTGGTGTATTAAGCAGTGACCC  
GAAACGCACGCCAGAGGTGAGCGATAATCCGGTAATGATTGGCGAATTAAGTGCAGGTTTTCCCGACTATACATG  
GCAGGTGGCGATTGCTGACCTTGAAGCAGAGCGAAGCCATCGGCGATCGCTTTGGCGTCTTTCCGCTTTCTGCCAC  
TTTGTGTTTACCAGCGGAAACTATCGCGGCGTGTGAATGGTATTCACCCGTGGGCGGAACTGATAAACTGAT  
GCGCGGGCTTGTGCAACCGCAGCAGGAGCGTGCCTCATGAGCGAACTTTTTTCCATCTGCTGGGGCCAGGAACG  
CAACCGAACGATGACAGTTTTACGATGAATCCACTGCCGATCACCTGTCAGGTGAATGATGAACCGAGTATGGCG  
GCCCTGGAGCAATGTGCTCACAGCCCGCAGGTGATTGCGCTGTTAAACGAGTTACAACATCAACTAAGCGAACGC  
CAACCGCCGTTGGGCGAGGTGCTGGCAGTGCATCTGTTAAATCTCAACGCCGACGATCGTCACTTTATCAATACG  
CTTCTCGGGGAAGGGGAAGTGTGAGTGCAGTGCAGTTCAGCAGGCTGACGACAGTGAAGTGAATAACAGGAGGCGATC  
TTCTGCGGATTATGGCGGGTGCAGCAGCGTCCGCGGCAAAAAGTTGCTGGAGGACAACTGGAGGCTGGCTGCGCG  
CCGCTGGCGTTGTGGCAGGCGGCAACGCAAAAATCTCTTGGCGACAGATTCGCTGTTACCAGCCGCCATTTGATGGC  
CTGATGAATGGCCTACCGTTGGCGCATGAGTTACTGGCACATGTACGTAACCCCGACGCGCAGCCGCACAGCAT  
AATCTGACGCAATTACCCATCAGCGAGGCTGATCGGCTTTTTCTCTCACGCTCTGTGGGCGGGAAATATTCAG  
ATTCGTACCATTGGCTATGGCGAGAGCTATATCAACGCCACGGGGTTACGCCATGTCTGGCATTACGCTGTACG

GACACCTTAAAAGGCCCGTTACTGGAAAGTTATGAAATCTGCCAATACCGGAAGTGGTGCTGGCAGCGCCAGAA  
GATTTGGTCGACTCTGCGCAGCGGCTTAGCGAGGTATGTCAGTGGCTGGCGGAAGCTGCACCGACG**TAA**

>pEcgRNA\_insertion Sequence

cgRNAccdbknrwardcaCCccTTgatGaaCTGaaCCGCTCGCCGCAGCCGAaCGACCGAGCG  
CAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCT  
TACGCATCTGTGCGgTatTTCACACCGCATATGCTGGATCCTTGACAGCTAGCTCA  
GTCCTAGGTATAATACTAGTTGAGACCCAGCAGTGAAAGATAAATGATCTGAGA  
CCCAGGTCGACCGACAGCCTTCCAAATGTTCTTCTCAAACGGAATCGTCGTATCC  
AGCCTACTCGCTATTGTCCTCAATGCCGTATTAATCATAAAAAGAAATAAGAAaA  
AAGAGGTGCGAGCCTCTTTTTTGTGTGACAAAATAAAAACATCTACCTATTCATA  
TACGCTAGTGTGCATAGTCCTGAAAATCATCTGCATCAAGAACAATTCACAACCTC  
TTATACTTTTCTCTTACAAGTCGTTTCGGCTTCATCTGGATTTTCAGCCTCTATACTT  
ACTAAACGTGATAAAGTTTCTGTAATTTCTACTGTATCGACCTGCAGACTGGCTG  
TGTATAAGGGAGCCTGACATTTATATTCCCAGAACATCAGGTTAATGGCGTTTT  
TGATGTCATTTTCGCGGTGGCTGAGATCAGCCACTTCTTCCCCGATAACGGAGAC  
CGGCACACTGGCCATATCGGTGGTCATCATGCGCCAGCTTTCATCCCCGATATGC  
ACCACCGGGTAAAGTTCACGGGAGACTTTATCTGACAGCAGACGTGCACTGGCC  
AGGGGGATCACCATCCGTCGCCCGGGCGTGTCAATAATATCACTCTGTAGGTAAT  
GACTCCAACCTTATTGATAGTGTTTTATGTTTCAGATAATGCCCGATGACTTTGTCAT  
GCAGCTCCACCGATTTTGAGAACGACAGCGACTTCCGTCCCAGCCGTGCCAGGTG  
CTGCCTCAGATTCAGGTTATGCCGCTCAATTCGCTGCGTATATCGCTTGCTGATTA  
CGTGCAGCTTTCCCTTCAGGCGGGATTTCATACAGCGGCCAGCCATCCGTCATCCA  
TATCACCACGTCAAAGGGTGACAGCAGGCTCATAAgACGCCCCAGCGTCGCCAT  
AGTGC GTTCACCGAATACGTGCGCAACAACCGTCTTCCGGAGACTGTCATACGCG  
TAAAACAGCcAgcgCTGGCgCGATTTAgCCcgAaatagCCCCacTGTTTCGTCCATTTCCcgc  
AgAcAAtgaagtccCTGCCCGGttgatGcccaaGGTtcCCaatgggggcCtGgatTTTTTtaAgggacaaA  
AAaccgggtgaagccaaccCcccaaaGGGGGgggtgttcggggccccccccctctcggggccttcccgggt

## **Additional Section:**

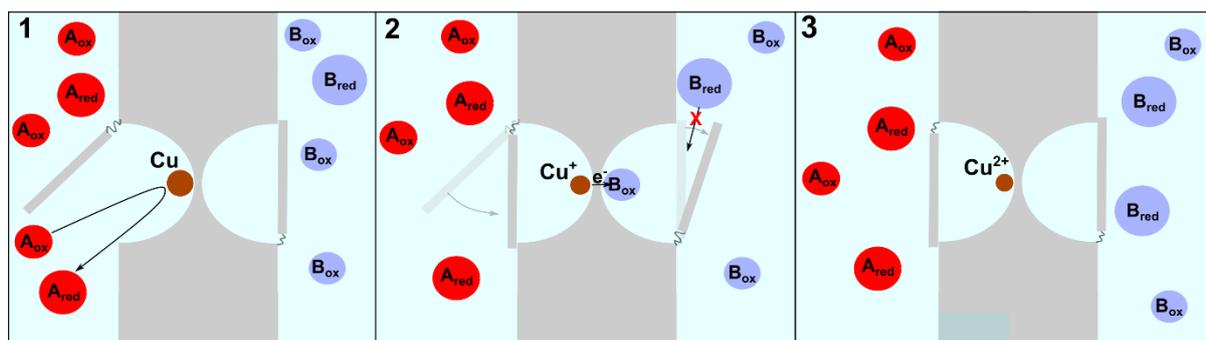
### **Can artificial bifurcation become a reality?**

Even though *T. maritima* hydrogenase is possibly not bifurcating, other enzymes are proven to be. Bifurcation requires by definition the presence of thermodynamic coupling which has important advantages in terms of energy efficiency. In fact, an exergonic reaction could be coupled with an endergonic one with the excess energy used to drive the endergonic reaction.

Thermodynamic couplings of reactions can be simply achieved by heating up an endergonic reaction using the heat produced by the combustion of methane or coal (exergonic reaction) or in more ‘fancy’ cases by using bifurcation.

Bifurcation is the thermodynamic coupling of reactions through electron splitting (detailed definition in Chapter 3.1.7) and so far only enzymes have been shown to be capable of it. But, could an artificial system (meant as non-biological) be built to perform bifurcation? If, yes, how much enzymes can teach us? I will answer these questions.

One of the simplest conceivable artificial bifurcation mechanisms would be that of an inverted potential species (such as copper) that in the semi-reduced form triggers a nano valve-like mechanism (Fig.A.2). However, even this conceptually simple mechanism involves dynamic movements of parts (e.g. a nano valve) that act as a logical gate (if *condition* then *open/close*) and currently, this is technologically too difficult to achieve.



**Figure A.2.** Potential artificial bifurcation mechanism. **1.** Fully reduced copper atoms absorbed in cavities of a separator material (e.g. a polymer) react with high potential one-electron acceptors molecules ( $A_{red}$ ) and by donating an electron become  $Cu^+$  ions. Immediately the nano valve (left) closes to avoid another  $A_{ox}$  molecule reacting with  $Cu^+$ . **2.** The nano valve on the right opens and the low potential one-electron acceptor ( $B_{ox}$ ) receives an electron from  $Cu^+$ . The nano valve (right) has a mechanism (e.g. only opens partially sterically excluding  $B_{red}$ ) to prevent  $B_{red}$  from entering the cavity and reducing  $Cu^+$  to  $Cu$ . **3.**  $B_{ox}$  after receiving an electron from  $Cu^+$  is released and the nano valve on the right closes. When all the  $Cu$  atoms in the cavities react following steps 1,2, and 3 a higher percentage of  $B_{red}$  molecules will be present in the solution (right) compared to the uncoupled reaction ( $Cu + 2B_{ox} \rightleftharpoons Cu^{2+} + 2B_{red}$ ).

Only enzymes have the capability of possessing such dynamic control over electron flow and any non-biological derived system would need a significantly different operating mechanism. Studying in great detail how bifurcating enzymes work is unlikely to provide the solution to an artificial bifurcating system (see Chapter 1.6.3 discussing biomimetics). Furthermore, artificial bifurcation should not even be attempted until an economically viable use for it has been found (research should start with the problem, not with the solution!).

# 8. References

All website references were last accessed on 02/08/2023.

1. <https://www.genome.jp/pathway/eco01100>.
2. D. R. Lovley, Microbial nanowires. *Current Biology* **32**, 97–115 (2022).
3. M. Ingelman, V. Bianchi, H. Eklund, The three-dimensional structure of flavodoxin reductase from *Escherichia coli* at 1.7 Å resolution. *J Mol Biol* **268**, 147–157 (1997).
4. M. Jormakka, S. Törnroth, B. Byrne, S. Iwata, Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. *Science* **295**, 1863–1868 (2002).
5. R. Baradaran, J. M. Berrisford, G. S. Minhas, L. A. Sazanov, Crystal structure of the entire respiratory complex I. *Nature* **494**, 443–448 (2013).
6. P. Mitchell, Coupling of phosphorylation to electron and hydrogen transfer by a chemi-osmotic type of mechanism. *Nature* **191**, 144–148 (1961).
7. R. A. Marcus, N. Sutin, Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **811**, 265–322 (1985).
8. C. C. Moser, T. A. Farid, S. E. Chobot, P. L. Dutton, Electron tunneling chains of mitochondria. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1757**, 1096–1109 (2006).
9. G. L. Closs, J. R. Miller, Intramolecular Long-Distance Electron Transfer in Organic Molecules. *Science* **240**, 440–447 (1988).
10. C. C. Moser, J. M. Keske, K. Warncke, R. S. Farid, P. L. Dutton, Nature of biological electron transfer. *Nature* **355**, 796–802 (1992).
11. C. C. Page, C. C. Moser, X. Chen, P. L. Dutton, Natural engineering principles of electron tunnelling in biological oxidation–reduction. *Nature* **402**, 47–52 (1999).
12. C. C. Moser, P. L. Dutton, Engineering protein structure for electron transfer function in photosynthetic reaction centers. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1101**, 171–176 (1992).
13. A. R. Crofts, S. Rose, Marcus treatment of endergonic reactions: a commentary. *Biochim Biophys Acta* **1767**, 1228–1232 (2007).
14. J. J. Hopfield, U. Tab, Electron Transfer Between Biological Molecules by Thermally Activated Tunneling. *Proceedings of the National Academy of Sciences* **71**, 3640–3644 (1974).
15. D. Devault, Quantum mechanical tunnelling in biological systems. *Q Rev Biophys* **13**, 387–564 (1980).
16. J. Liu, S. Chakraborty, P. Hosseinzadeh, Y. Yu, S. Tian, I. Petrik, A. Bhagi, Y. Lu, Metalloproteins containing cytochrome, iron-sulfur, or copper redox centers. *Chem Rev* **114**, 4366–4369 (2014).
17. R. Shi, W. Hou, Z. Q. Wang, X. Xu, Biogenesis of Iron-Sulfur Clusters and Their Role in DNA Metabolism. *Front Cell Dev Biol* **9** (2021).

18. Thermal-FluidsPedia | Heat of Combustion | Thermal-Fluids Central.  
[https://www.thermalfluidscentral.org/encyclopedia/index.php/Heat\\_of\\_Combustion](https://www.thermalfluidscentral.org/encyclopedia/index.php/Heat_of_Combustion).
19. S. P. J. Albracht, Hydrogen Metabolism: Structure and Function of Hydrogenases. *The Molecular Basis of Bacterial Metabolism*, 40–51 (1990).
20. C. Greening, P. Constant, K. Hards, S. E. Morales, J. G. Oakeshott, R. J. Russell, M. C. Taylor, M. Berney, R. Conrad, G. M. Cook, Atmospheric hydrogen scavenging: from enzymes to ecosystems. *Appl Environ Microbiol* **81**, 1190–1199 (2015).
21. R. Grinter, A. Kropp, H. Venugopal, M. Senger, J. Badley, P. R. Cabotaje, R. Jia, Z. Duan, P. Huang, S. T. Stripp, C. K. Barlow, M. Belousoff, H. S. Shafaat, G. M. Cook, R. B. Schittenhelm, K. A. Vincent, S. Khalid, G. Berggren, C. Greening, Structural basis for bacterial energy extraction from atmospheric hydrogen. *Nature* **615**, 541–547 (2023).
22. P. Constant, S. P. Chowdhury, L. Hesse, J. Pratscher, R. Conrad, Genome data mining and soil survey for the novel group 5 [NiFe]-hydrogenase to explore the diversity and ecological importance of presumptive high-affinity H<sub>2</sub>-oxidizing bacteria. *Appl Environ Microbiol* **77**, 6027–6035 (2011).
23. By Smokefoot - Own work, CC BY-SA 4.0,  
<https://commons.wikimedia.org/w/index.php?curid=59599838>.  
<https://en.wikipedia.org/wiki/Hydrogenase#/media/File:ActiveSitesCorrected.png>.
24. C. Ding, K. L. Yang, J. He, Biological and fermentative production of hydrogen. *Handbook of Biofuels Production: Processes and Technologies: Second Edition*, 303–333 (2016).
25. P. C. Hallenbeck, Photofermentative Biohydrogen Production. *Biohydrogen*, 145–159 (2013).
26. C. Greening, A. Biswas, C. R. Carere, C. J. Jackson, M. C. Taylor, M. B. Stott, G. M. Cook, S. E. Morales, Genomic and metagenomic surveys of hydrogenase distribution indicate H<sub>2</sub> is a widely utilised energy source for microbial growth and survival. *The ISME Journal* **10**:3, 761–777 (2015).
27. P. M. Vignais, B. Billoud, Occurrence, classification, and biological function of hydrogenases: an overview. *Chem Rev* **107**, 4206–4272 (2007).
28. A. K. Jones, E. Sillery, S. P. J. Albracht, F. A. Armstrong, Direct comparison of the electrocatalytic oxidation of hydrogen by an enzyme and a platinum catalyst. *Chemical Communications* **2**, 866–867 (2002).
29. Hydrogen Production and Uses - World Nuclear Association. <https://www.world-nuclear.org/information-library/energy-and-the-environment/hydrogen-production-and-uses.aspx>.
30. Renewable Energy Agency, Green hydrogen: A guide to policy making. (2020).  
<https://www.irena.org/publications/2020/Nov/Green-hydrogen#:~:text=This%20report%20from%20the%20International,the%20national%20or%20Regional%20level>.
31. C. Liu, B. C. Colón, M. Ziesack, P. A. Silver, D. G. Nocera, Water splitting-biosynthetic system with CO<sub>2</sub> reduction efficiencies exceeding photosynthesis. *Science* **352**, 1210–1213 (2016).

32. J. Yu, Fixation of carbon dioxide by a hydrogen-oxidizing bacterium for value-added products. *World J Microbiol Biotechnol* **34**, 1–7 (2018).
33. Hydrogen Production Costs 2021. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1011506/Hydrogen\\_Production\\_Costs\\_2021.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1011506/Hydrogen_Production_Costs_2021.pdf) (2021).
34. Hydrogen production costs using natural gas in selected regions, 2018 – Charts – Data & Statistics - IEA. <https://www.iea.org/data-and-statistics/charts/hydrogen-production-costs-using-natural-gas-in-selected-regions-2018-2>.
35. Hydrogen Basics - Production. <http://www.fsec.ucf.edu/en/consumer/hydrogen/basics/production.htm>.
36. J. E. Kim, E. J. Kim, H. Chen, C. H. Wu, M. W. W. Adams, Y. H. P. Zhang, Advanced water splitting for green hydrogen gas production through complete oxidation of starch by in vitro metabolic engineering. *Metab Eng* **44**, 246–252 (2017).
37. J. A. Rollin, J. M. Del Campo, S. Myung, F. Sun, C. You, A. Bakovic, R. Castro, S. K. Chandrayan, C. H. Wu, M. W. W. Adams, R. S. Senger, Y. H. P. Zhang, High-yield hydrogen production from biomass by in vitro metabolic engineering: Mixed sugars coutilization and kinetic modeling. *Proc Natl Acad Sci U S A* **112**, 4964–4969 (2015).
38. Y. Liu, S. Webb, P. Moreno-García, A. Kulkarni, P. Maroni, P. Broekmann, R. D. Milton, Facile Functionalization of Carbon Electrodes for Efficient Electroenzymatic Hydrogen Production. *JACS Au* **3**, 124–130 (2023).
39. J. C. Ruth, R. D. Milton, W. Gu, A. M. Spormann, Enhanced Electrosynthetic Hydrogen Evolution by Hydrogenases Embedded in a Redox-Active Hydrogel. *Chemistry* **26**, 7323–7329 (2020).
40. E. J. Kim, J. E. Kim, Y. H. P. J. Zhang, Ultra-rapid rates of water splitting for biohydrogen gas production through in vitro artificial enzymatic pathways. *Energy Environ Sci* **11**, 2064–2072 (2018).
41. X. Ye, Y. Wang, R. C. Hopkins, M. W. W. Adams, B. R. Evans, J. R. Mielenz, Y. H. P. Zhang, Spontaneous high-yield production of hydrogen from cellulosic materials and water catalyzed by enzyme cocktails. *ChemSusChem* **2**, 149–152 (2009).
42. J. S. Martín Del Campo, J. Rollin, S. Myung, Y. Chun, S. Chandrayan, R. Patiño, M. W. Adams, Y. H. P. Zhang, High-yield production of dihydrogen from xylose by using a synthetic enzyme cascade in a cell-free system. *Angew Chem Int Ed Engl* **52**, 4587–4590 (2013).
43. S. Myung, J. Rollin, C. You, F. Sun, S. Chandrayan, M. W. W. Adams, Y. H. P. Zhang, In vitro metabolic engineering of hydrogen production at theoretical yield from sucrose. *Metab Eng* **24**, 70–77 (2014).
44. H. M. A. Moustafa, E. J. Kim, Z. Zhu, C. H. Wu, T. I. Zaghoul, M. W. W. Adams, Y. H. P. Zhang, Water Splitting for High-Yield Hydrogen Production Energized by Biomass Xylooligosaccharides Catalyzed by an Enzyme Cocktail. *ChemCatChem* **8**, 2898–2902 (2016).
45. Y. H. P. Zhang, A sweet out-of-the-box solution to the hydrogen economy: is the sugar-powered car science fiction? *Energy Environ Sci* **2**, 272–282 (2009).

46. R. K. Thauer, K. Jungermann, K. Decker, Energy conservation in chemotrophic anaerobic bacteria. *Bacteriol Rev* **41**, 100–180 (1977).
47. Y. H. P. Zhang, B. R. Evans, J. R. Mielenz, R. C. Hopkins, M. W. W. Adams, High-Yield Hydrogen Production from Starch and Water by a Synthetic Enzymatic Pathway. *PLoS One* **2**, e456 (2007).
48. Y. H. P. Zhang, J. Sun, J. J. Zhong, Biofuel production by in vitro synthetic enzymatic pathway biotransformation. *Curr Opin Biotechnol* **21**, 663–669 (2010).
49. M. B. Wolfe, Sweet Hydrogen: High-Yield Production of Hydrogen from Biomass Sugars Catalyzed by in vitro Synthetic Biosystems. *U.S. Department of Energy Office of Scientific and Technical Information*, doi: 10.2172/1482506 (2018).
50. R. Wu, C. Ma, Z. Zhu, Enzymatic electrosynthesis as an emerging electrochemical synthesis platform. *Curr Opin Electrochem* **19**, 1–7 (2020).
51. J. C. Ruth, F. M. Schwarz, V. Müller, A. M. Spormann, Enzymatic hydrogen electrosynthesis at enhanced current density using a redox polymer. *Catalysts* **11**, 1197 (2021).
52. R. D. Milton, S. Abdellaoui, N. Khadka, D. R. Dean, D. Leech, L. C. Seefeldt, S. D. Minter, Nitrogenase bioelectrocatalysis: heterogeneous ammonia and hydrogen production by MoFe protein. *Energy Environ Sci* **9**, 2550–2554 (2016).
53. H. Li, D. Buesen, S. Dementin, C. Léger, V. Fourmond, N. Plumeré, Complete Protection of O<sub>2</sub>-Sensitive Catalysts in Thin Films. *J Am Chem Soc* **141**, 16734–16742 (2019).
54. A. A. Oughli, S. Hardt, O. Rüdiger, J. A. Birrell, N. Plumeré, Reactivation of sulfide-protected [FeFe] hydrogenase in a redox-active hydrogel. *Chemical Communications* **56**, 9958–9961 (2020).
55. Economics — SGH2 Energy. <https://www.sgh2energy.com/economics>.
56. M. Seibert, S. Lien, P. F. Weaver, A. F. Janzen, Photobiological production of hydrogen and electricity. *Solar Energy Conversion II*, 273–292 (1981).
57. A. Pandey, V. V. Mohan, J. S. Chang, P. C. Hallenbeck, C. Larroche, Biomass, Biofuels, Biochemicals: Biohydrogen, Second Edition. *Biomass, Biofuels, Biochemicals: Biohydrogen, Second Edition*, 1–525 (2019).
58. S. Fouchard, A. Hemschemeier, A. Caruana, J. Pruvost, J. Legrand, T. Happe, G. Peltier, L. Cournac, Autotrophic and mixotrophic hydrogen photoproduction in sulfur-deprived chlamydomonas cells. *Appl Environ Microbiol* **71**, 6199–6205 (2005).
59. K.-Y. Show, Y. Yan, M. Ling, G. Ye, T. Li, D.-J. Lee, Hydrogen production from algal biomass-Advances, challenges and prospects. *Bioresource Technology* **257**, 290-300 (2018).
60. M. Oey, A. L. Sawyer, I. L. Ross, B. Hankamer, Challenges and opportunities for hydrogen production from microalgae. *Plant Biotechnol J* **14**, 1487–1499 (2016).
61. M. H. Huesemann, T. S. Hausmann, B. M. Carter, J. J. Gerschler, J. R. Benemann, Hydrogen generation through indirect biophotolysis in batch cultures of the nonheterocystous nitrogen-fixing cyanobacterium *Plectonema boryanum*. *Appl Biochem Biotechnol* **162**, 208–220 (2010).

62. C. N. C. Hitam, A. A. Jalil, A review on biohydrogen production through photo-fermentation of lignocellulosic biomass. *Biomass Convers Biorefin* **13**, 8465–8483 (2023).
63. J. D. Holladay, J. Hu, D. L. King, Y. Wang, An overview of hydrogen production technologies. *Catal Today* **139**, 244–260 (2009).
64. S. Dahiya, S. Chatterjee, O. Sarkar, S. V. Mohan, Renewable hydrogen production by dark-fermentation: Current status, challenges and perspectives. *Bioresour Technol* **321**, 124354 (2021).
65. E. Gürtekin, Biological Hydrogen Production Methods.
66. L. Treu, S. Campanaro, P. G. Kougias, C. Sartori, I. Bassani, I. Angelidaki, Hydrogen-Fueled Microbial Pathways in Biogas Upgrading Systems Revealed by Genome-Centric Metagenomics. *Front Microbiol* **9** (2018).
67. B. Stoppacher, T. Sterniczky, S. Bock, V. Hacker, On-site production of high-purity hydrogen from raw biogas with fixed-bed chemical looping. *Energy Convers Manag* **268**, 115971 (2022).
68. E. Stevens, M. L. Marco, Bacterial extracellular electron transfer in plant and animal ecosystems. *FEMS Microbiol Rev* **47** (2023).
69. A. Kadier, Y. Simayi, P. Abdeshahian, N. F. Azman, K. Chandrasekhar, M. S. Kalil, A comprehensive review of microbial electrolysis cells (MEC) reactor designs and configurations for sustainable hydrogen gas production. *Alexandria Engineering Journal* **55**, 427–443 (2016).
70. R. Rousseau, L. Etcheverry, E. Roubaud, R. Basséguy, M. L. Délia, A. Bergel, Microbial electrolysis cell (MEC): Strengths, weaknesses and research needs from electrochemical engineering standpoint. *Appl Energy* **257**, 113938 (2020).
71. Comparing Fuel Cell Technologies - GenCell - Fuel Cell Generators.  
<https://www.gencellenergy.com/news/comparing-fuel-cell-technologies/>.
72. A. Kampker, H. Heimes, M. Kehrer, S. Hagedorn, P. Reims, O. Kaul, Fuel cell system production cost modeling and analysis. *Energy Reports* **9**, 248–255 (2023).
73. B. Yang, X. Yu, J. Hou, Z. Xiang, Secondary reduction strategy synthesis of Pt–Co nanoparticle catalysts towards boosting the activity of proton exchange membrane fuel cells. *Particuology* **79**, 18–26 (2023).
74. Alternative Fuels Data Center: Hydrogen Benefits and Considerations.  
[https://afdc.energy.gov/fuels/hydrogen\\_benefits.html](https://afdc.energy.gov/fuels/hydrogen_benefits.html).
75. A. Mehmood, M. Gong, F. Jaouen, A. Roy, A. Zitolo, A. Khan, M. T. Sougrati, M. Primbs, A. M. Bonastre, D. Fongalland, G. Drazic, P. Strasser, A. Kucernak, High loading of single atomic iron sites in Fe–NC oxygen reduction catalysts for proton exchange membrane fuel cells. *Nature Catalysis* **2022 5:4 5**, 311–323 (2022).
76. A. Morozan, B. Josselme, S. Palacin, Low-platinum and platinum-free catalysts for the oxygen reduction reaction at fuel cell cathodes. *Energy Environ Sci* **4**, 1238–1254 (2011).
77. E. Lojou, Hydrogenases as catalysts for fuel cells: Strategies for efficient immobilization at electrode interfaces. *Electrochim Acta* **56**, 10385–10397 (2011).

78. Anglo American positions for fuel cell boost to platinum. <https://www.ft.com/content/757f907a-ebde-11e8-8180-9cf212677a57>.
79. Fuel cell electric vehicles and platinum demand - CME Group. <https://www.cmegroup.com/articles/2022/wpic-fuel-cell-electric-vehicles-and-platinum-demand.html>.
80. Manufacturing Cost Analysis of PEM Fuel Cell Systems for 5- and 10-kW Backup Power Applications. (2016). <https://www.energy.gov/eere/fuelcells/articles/manufacturing-cost-analysis-pem-fuel-cell-systems-5-and-10-kw-backup-power>
81. S. Cherevko, G. P. Keeley, S. Geiger, A. R. Zeradjanin, N. Hodnik, N. Kulyk, K. J. J. Mayrhofer, Dissolution of Platinum in the Operational Range of Fuel Cells. *ChemElectroChem* **2**, 1471–1478 (2015).
82. P. C. Okonkwo, O. O. Ige, E. M. Barhoumi, P. C. Uzoma, W. Emori, A. Benamor, A. M. Abdullah, Platinum degradation mechanisms in proton exchange membrane fuel cell (PEMFC) system: A review. *Int J Hydrogen Energy* **46**, 15850–15865 (2021).
83. J. W. Tye, M. B. Hall, M. Y. Darensbourg, Better than platinum? Fuel cells energized by enzymes. *Proc Natl Acad Sci U S A* **102**, 16911–16912 (2005).
84. Stationary PEM fuel cells with lifetimes beyond five years <https://cordis.europa.eu/project/id/256721/reporting>.
85. Why hydrogen fuel cell cars can't compete with electric cars. <https://energypost.eu/hydrogen-fuel-cell-cars-competitive-hydrogen-fuel-cell-expert/>.
86. Fuel Cells | Department of Energy. <https://www.energy.gov/eere/fuelcells/fuel-cells>.
87. K. A. Vincent, J. A. Cracknell, J. R. Clark, M. Ludwig, O. Lenz, B. Friedrich, F. A. Armstrong, Electricity from low-level H<sub>2</sub> in still air – an ultimate test for an oxygen tolerant hydrogenase. *Chemical Communications*, 5033–5035 (2006).
88. K. A. Vincent, J. A. Cracknell, O. Lenz, I. Zebger, B. Friedrich, F. A. Armstrong, Electrocatalytic hydrogen oxidation by an enzyme at high carbon monoxide or oxygen levels. *Proc Natl Acad Sci U S A* **102**, 16951–16954 (2005).
89. New Enzyme-Based Fuel Cell Produces Electricity from Hydrogen in Plain Air - Green Car Congress. [https://www.greencarcongress.com/2007/03/new\\_enzymebased.html](https://www.greencarcongress.com/2007/03/new_enzymebased.html).
90. L. Marchetti, M. Levine, Biomimetic catalysis. *ACS Catal* **1**, 1090–1118 (2011).
91. F. Liew, A. M. Henstra, M. Köpke, K. Winzer, S. D. Simpson, N. P. Minton, Metabolic engineering of *Clostridium autoethanogenum* for selective alcohol production. *Metab Eng* **40**, 104–114 (2017).
92. S. L. Benoit, R. J. Maier, R. G. Sawers, C. Greening, Molecular Hydrogen Metabolism: a Widespread Trait of Pathogenic Bacteria and Protists. *Microbiol Mol Biol Rev* **84** (2020).
93. BLAST: Basic Local Alignment Search Tool. <https://blast.ncbi.nlm.nih.gov/Blast.cgi>.
94. A. R. Coates, G. Halls, Y. Hu, Novel classes of antibiotics or more of the same? *Br J Pharmacol* **163**, 184 (2011).

95. Why is it so hard to develop new antibiotics? <https://wellcome.org/news/why-is-it-so-hard-develop-new-antibiotics>.
96. A. K. Shah, A. A. Shah, A. K. Shah, A. A. Shah, Responsible antibiotic therapy simplified. *Karnataka Paediatric Journal* **35**, 29–38 (2020).
97. J. D. Westbrook, S. K. Burley, How Structural Biologists and the Protein Data Bank Contributed to Recent US FDA New Drug Approvals. *Structure* **27**, 211 (2019).
98. PDB Statistics. <https://www.rcsb.org/stats/>.
99. EMDB < Statistics. <https://www.ebi.ac.uk/emdb/statistics>.
100. W. Kühlbrandt, The Resolution Revolution. *Science (1979)* **343**, 1443–1444 (2014).
101. M. Beckers, D. Mann, C. Sachse, Structural interpretation of cryo-EM image reconstructions. *Prog Biophys Mol Biol* **160**, 26–36 (2021).
102. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Atomic-resolution protein structure determination by cryo-EM. *Nature* **587**, 157–161 (2020).
103. PDB Statistics: PDB Data Distribution by Resolution. <https://www.rcsb.org/stats/distribution-resolution>.
104. K. R. Acharya, M. D. Lloyd, The advantages and limitations of protein crystal structures. *Trends Pharmacol Sci* **26**, 10–14 (2005).
105. K. Wentinck, C. Gogou, D. H. Meijer, Putting on molecular weight: Enabling cryo-EM structure determination of sub-100-kDa proteins. *Curr Res Struct Biol* **4**, 332–337 (2022).
106. J. Hutchings, G. Zanetti, Correction: Fine details in complex environments: The power of cryo-electron tomography. *Biochem Soc Trans* **46**, 1771 (2018).
107. S. Dunst, P. Tomancak, Imaging Flies by Fluorescence Microscopy: Principles, Technologies, and Applications. *Genetics* **211**, 15–34 (2019).
108. P. A. Penczek, Resolution measures in molecular electron microscopy. *Methods Enzymol* **482**, 73–100 (2010).
109. Wave–Particle Duality: De Broglie, Einstein, and Schrödinger. *Critical Appraisal of Physical Science as a Human Enterprise*, 159–165 (2009).
110. J. Dubochet, A. W. McDowell, Vitrification of pure water for electron microscopy. *J Microsc* **124**, 3–4 (1981).
111. A. J. Noble, V. P. Dandey, H. Wei, J. Brasch, J. Chase, P. Acharya, Y. Z. Tan, Z. Zhang, L. Y. Kim, G. Scapin, M. Rapp, E. T. Eng, W. J. Rice, A. Cheng, C. J. Negro, L. Shapiro, P. D. Kwong, D. Jeruzalmi, A. des Georges, C. S. Potter, B. Carragher, Routine single particle CryoEM sample and grid characterization by tomography. *Elife* **7** (2018).
112. C. Wu, H. Shi, D. Zhu, K. Fan, X. Zhang, Low-cooling-rate freezing in biomolecular cryo-electron microscopy for recovery of initial frames. *QRB Discov* **2**, e11 (2021).
113. G. Weissenberger, R. J. M. Henderikx, P. J. Peters, Understanding the invisible hands of sample preparation for cryo-EM. *Nature Methods* **18**, 463–471 (2021).

114. R. B. G. Ravelli, F. J. T. Nijpels, R. J. M. Henderikx, G. Weissenberger, S. Thewessem, A. Gijsbers, B. W. A. M. M. Beulen, C. López-Iglesias, P. J. Peters, Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification. *Nature Communications* **11**, 1–9 (2020).
115. J. L. Rubinstein, H. Guo, Z. A. Ripstein, A. Haydaroglu, A. Au, C. M. Yip, J. M. D. Trani, S. Benlekbira, T. Kwoka, Shake-it-off: a simple ultrasonic cryo-EM specimen-preparation device. *Acta Crystallogr D Struct Biol* **75**, 1063–1070 (2019).
116. M. E. Mäeots, B. Lee, A. Nans, S. G. Jeong, M. M. N. Esfahani, S. Ding, D. J. Smith, C. S. Lee, S. S. Lee, M. Peter, R. I. Enchev, Modular microfluidics enables kinetic insight from time-resolved cryo-EM. *Nature Communications* **11**, 1–14 (2020).
117. C. J. Russo, L. A. Passmore, Ultrastable gold substrates: Properties of a support for high-resolution electron cryomicroscopy of biological specimens. *J Struct Biol* **193**, 33–44 (2016).
118. R. Danev, M. Belousoff, Y. L. Liang, X. Zhang, F. Eisenstein, D. Wootten, P. M. Sexton, Routine sub-2.5 Å cryo-EM structure determination of GPCRs. *Nature Communications* **12**, 1–10 (2021).
119. T. R. D. Costa, A. Ignatiou, E. V. Orlova, Structural Analysis of Protein Complexes by Cryo Electron Microscopy. *Methods Mol Biol* **1615**, 377–413 (2017).
120. J. R. Meyerson, P. Rao, J. Kumar, S. Chittori, S. Banerjee, J. Pierson, M. L. Mayer, S. Subramaniam, Self-assembled monolayers improve protein distribution on holey carbon cryo-EM supports. *Sci Rep* **4**, 7084 (2014).
121. Z. Ma, L. Niu, W. Jiang, al -, S. Malunavar, X. Wang, B. D. A Levin, Direct detectors and their applications in electron microscopy for materials science. *Journal of Physics: Materials* **4**, 042005 (2021).
122. A. R. Faruqi, Direct Electron Detectors for Electron Microscopy. *Advances in Imaging and Electron Physics* **145**, 55–93 (2007).
123. G. McMullan, A. R. Faruqi, R. Henderson, Direct Electron Detectors. *Methods Enzymol* **579**, 1–17 (2016).
124. D. Bhella, Cryo-electron microscopy: an introduction to the technique, and considerations when working to establish a national facility. *Biophys Rev* **11**, 515–519 (2019).
125. B. Himes, N. Grigorieff, Cryo-TEM simulations of amorphous radiation-sensitive samples using multislice wave propagation. *IUCrJ* **8**, 943–953 (2021).
126. C. Cassidy, A. Dhar, T. Shintake, Determination of the mean inner potential of cadmium telluride via electron holography. *Appl Phys Lett* **110**, 163503 (2017).
127. X. Zhang, Y. Oshima, Reproducible measurement of the mean inner potential of gold nanoparticles using TIE method. *Microscopy Society of America*, poster session (2013).
128. L. E. Franken, K. Grünewald, E. J. Boekema, M. C. A. Stuart, L. E. Franken, K. Grünewald, E. J. Boekema, M. C. A. Stuart, A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. *Small* **16**, 1906198 (2020).
129. Image frequency spectrum analysis. [https://www.djmannion.net/img\\_freq\\_web/](https://www.djmannion.net/img_freq_web/).

130. N. Hampp, C. Bräuchle, Bacteriorhodopsin and its Functional Variants: Potential Applications in Modern Optics. *Photochromism: Molecules and Systems*, 954–975 (2003).
131. D. Střelák, J. Filipovič, A. Jiménez-Moreno, J. M. Carazo, C. Ó. S. Sorzano, FlexAlign: An Accurate and Fast Algorithm for Movie Alignment in Cryo-Electron Microscopy. *Electronics 2020, Vol. 9, Page 1040* **9**, 1040 (2020).
132. M. Carroni, H. R. Saibil, Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods* **95**, 78–85 (2016).
133. A. F. Brilot, J. Z. Chen, A. Cheng, J. Pan, S. C. Harrison, C. S. Potter, B. Carragher, R. Henderson, N. Grigorieff, Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* **177**, 630–637 (2012).
134. K. Naydenova, P. Jia, C. J. Russo, Cryo-EM with sub-1 Å specimen movement. *Science* **370**, 223–226 (2020).
135. S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, D. A. Agard, MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nature Methods* **14**, 331–332 (2017).
136. A. Rohou, N. Grigorieff, CTFIND4: Fast and accurate defocus estimation from electron micrographs. *J Struct Biol* **192**, 216 (2015).
137. E. V. Orlova, H. R. Saibil, Structural analysis of macromolecular assemblies by electron microscopy. *Chem Rev* **111**, 7710–7748 (2011).
138. D. Woolford, B. Hankamer, G. Ericksson, The Laplacian of Gaussian and arbitrary z-crossings approach applied to automated single particle reconstruction. *J Struct Biol* **159**, 122–134 (2007).
139. X. Wang, Y. Lu, J. Liu, A Fast Image Alignment Approach for 2D Classification of Cryo-EM Images Using Spectral Clustering. *Current Issues in Molecular Biology* **43**, 1652–1668 (2021).
140. X. Agirrezabala, J. Frank, From DNA to proteins via the ribosome: structural insights into the workings of the translation machinery. *Hum Genomics* **4**, 226–237 (2010).
141. K. M. Yip, N. Fischer, E. Paknia, A. Chari, H. Stark, Breaking the next Cryo-EM resolution barrier – Atomic resolution determination of proteins! *bioRxiv* (2020).
142. Plane with ENU embedded axes.  
<https://commons.wikimedia.org/w/index.php?curid=80781452>.
143. P. B. Rosenthal, R. Henderson, Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-particle Electron Cryomicroscopy. *J Mol Biol* **333**, 721–745 (2003).
144. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845–858 (2015).
145. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.

- W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
146. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486–501 (2010).
147. D. Liebschner, P. V. Afonine, M. L. Baker, G. Bunkoczi, V. B. Chen, T. I. Croll, B. Hintze, L. W. Hung, S. Jain, A. J. McCoy, N. W. Moriarty, R. D. Oeffner, B. K. Poon, M. G. Prisant, R. J. Read, J. S. Richardson, D. C. Richardson, M. D. Sammito, O. V. Sobolev, D. H. Stockwell, T. C. Terwilliger, A. G. Urzhumtsev, L. L. Videau, C. J. Williams, P. D. Adams, Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr D Struct Biol* **75**, 861–877 (2019).
148. T. Moriya, M. Saur, M. Stabrin, F. Merino, H. Voicu, Z. Huang, P. A. Penczek, S. Raunser, C. Gatsogiannis, High-resolution Single Particle Analysis from Electron Cryo-microscopy Images Using SPHIRE. *J Vis Exp* **2017** (2017).
149. S. Choudhuri, Additional Bioinformatic Analyses Involving Protein Sequences. *Bioinformatics for Beginners*, 183–207 (2014).
150. N. Chongdar, K. Pawlak, O. Rüdiger, E. J. Reijerse, P. Rodríguez-Maciá, W. Lubitz, J. A. Birrell, H. Ogata, Spectroscopic and biochemical insight into an electron-bifurcating [FeFe] hydrogenase. *Journal of Biological Inorganic Chemistry* **25**, 135–149 (2020).
151. J. M. Kuchenreuther, C. S. Grady-Smith, A. S. Bingham, S. J. George, S. P. Cramer, J. R. Swartz, High-yield expression of heterologous [FeFe] hydrogenases in *Escherichia coli*. *PLoS One* **5** (2010).
152. G. Berggren, A. Adamska, C. Lambertz, T. R. Simmons, J. Esselborn, M. Atta, S. Gambarelli, J. M. Mouesca, E. Reijerse, W. Lubitz, T. Happe, V. Artero, M. Fontecave, Biomimetic assembly and activation of [FeFe]-hydrogenases. *Nature* **499**, 66–69 (2013).
153. J. Esselborn, C. Lambertz, A. Adamska-Venkatesh, T. Simmons, G. Berggren, J. Noth, J. Siebel, A. Hemschemeier, V. Artero, E. Reijerse, M. Fontecave, W. Lubitz, T. Happe, Spontaneous activation of [FeFe]-hydrogenases by an inorganic [2Fe] active site mimic. *Nat Chem Biol* **9**, 607–609 (2013).
154. J. Esselborn, N. Muraki, K. Klein, V. Engelbrecht, N. Metzler-Nolte, U. P. Apfel, E. Hofmann, G. Kurisu, T. Happe, A structural view of synthetic cofactor integration into [FeFe]-hydrogenases. *Chem Sci* **7**, 959–968 (2016).
155. H. Guo, E. Franken, Y. Deng, S. Benlekbir, G. S. Lezcano, B. Janssen, L. Yu, Z. A. Ripstein, Y. Z. Tan, J. L. Rubinstein, Electron-event representation data enable efficient cryoEM file storage with full preservation of spatial and temporal resolution. *IUCrJ* **7**, 860–869 (2020).
156. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. H. Hagen, E. Lindahl, S. H. W. Scheres, New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7** (2018).
157. J. Zivanov, T. Nakane, S. H. W. Scheres, Estimation of high-order aberrations and anisotropic magnification from cryo-EM data sets in RELION-3.1. *IUCrJ* **7**, 253–267 (2020).
158. J. Zivanov, T. Nakane, S. H. W. Scheres, A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).

159. S. L. Ilca, A. Kotecha, X. Sun, M. M. Poranen, D. I. Stuart, J. T. Huiskonen, Localized reconstruction of subunits from electron cryomicroscopy images of macromolecular complexes. *Nature Communications* **6**, 1–8 (2015).
160. E. F. Pettersen, T. D. Goddard, C. C. Huang, E. C. Meng, G. S. Couch, T. I. Croll, J. H. Morris, T. E. Ferrin, UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70–82 (2021).
161. N. W. Moriarty, P. D. Adams, Iron-sulfur clusters have no right angles. *Acta Crystallogr D Struct Biol* **75**, 16–20 (2019).
162. M. Nouailler, X. Morelli, O. Bornet, B. Chetrit, Z. Dermoun, F. Guerlesquin, Solution structure of HndAc: A thioredoxin-like domain involved in the NADP-reducing hydrogenase complex. *Protein Sci* **15**, 1369 (2006).
163. Q. Li, B. Sun, J. Chen, Y. Zhang, Y. Jiang, S. Yang, A modified pCas/pTargetF system for CRISPR-Cas9-assisted genome editing in *Escherichia coli*. *Acta Biochim Biophys Sin (Shanghai)* **53**, 620–627 (2020).
164. PCR Primer Design Tips - Behind the Bench.  
<https://www.thermofisher.com/blog/behindthebench/pcr-primer-design-tips>.
165. L. Flanagan, Hydrogen as fuel for enteric bacteria: biochemistry of the membrane bound nife hydrogenase. (2016).
166. Marion M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* **72**, 248–254 (1976).
167. B. J. S. C. Olson, J. Markwell, Assays for determination of protein concentration. *Curr Protoc Protein Sci* **Chapter 3** (2007).
168. P. K. Smith, R. I. Krohn, G. T. Hermanson, A. K. Mallia, F. H. Gartner, M. D. Provenzano, E. K. Fujimoto, N. M. Goeke, B. J. Olson, D. C. Klenk, Measurement of protein using bicinchoninic acid. *Anal Biochem* **150**, 76–85 (1985).
169. J. E. Noble, M. J. A. Bailey, Quantitation of protein. *Methods Enzymol* **463**, 73–95 (2009).
170. S. L. D. Wrathall, B. Procacci, M. Horch, E. Saxton, C. Furlan, J. Walton, Y. Rippers, J. N. Blaza, G. M. Greetham, M. Towrie, A. W. Parker, J. Lynam, A. Parkin, N. T. Hunt, Ultrafast 2D-IR spectroscopy of [NiFe] hydrogenase from *E. coli* reveals the role of the protein scaffold in controlling the active site environment. *Physical Chemistry Chemical Physics* **24**, 24767–24783 (2022).
171. C. Furlan, N. Chongdar, P. Gupta, W. Lubitz, H. Ogata, J. N. Blaza, J. A. Birrell, Structural insight on the mechanism of an electron-bifurcating [FeFe] hydrogenase. *Elife* **11** (2022).
172. Y. Qi, X. Shanguan, J. He, L. Chen, J. Jin, Y. Liu, G. Qiu, R. Yu, J. Li, W. Zeng, L. Shen, X. Wu, Expression, purification, characterization and direct electrochemistry of two HiPIPs from *Acidithiobacillus caldus* SM-1. *Anal Biochem* **650**, 114724 (2022).
173. I. B. Trindade, M. Invernici, F. Cantini, R. O. Louro, M. Piccioli, <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N assignment of the paramagnetic high potential iron–sulfur protein (HiPIP) PioC from *Rhodospseudomonas palustris* TIE-1. *Biomol NMR Assign* **14**, 211 (2020).

174. J. Meyer, Iron-sulfur protein folds, iron-sulfur chemistry, and evolution. *Journal of Biological Inorganic Chemistry* **13**, 157–170 (2008).
175. T. Ohnishi, Iron–sulfur clusters/semiquinones in Complex I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1364**, 186–206 (1998).
176. J. A. Birrell, K. Morina, H. R. Bridges, T. Friedrich, J. Hirst, Investigating the function of [2Fe-2S] cluster N1a, the off-pathway cluster in complex I, by manipulating its reduction potential. *Biochem J* **456**, 139–146 (2013).
177. L. Valer, D. Rossetto, T. Parkkila, L. Sebastianelli, G. Guella, A. L. Hendricks, J. A. Cowan, L. Sang, S. S. Mansy, Histidine Ligated Iron-Sulfur Peptides. *ChemBioChem* **23**, e202200202 (2022).
178. C. Madden, M. D. Vaughn, I. Díez-Pérez, K. A. Brown, P. W. King, D. Gust, A. L. Moore, T. A. Moore, Catalytic turnover of [FeFe]-hydrogenase based on single-molecule imaging. *J Am Chem Soc* **134**, 1577–1582 (2012).
179. M. T. Stiebritz, M. Reiher, Hydrogenases and oxygen. *Chem Sci* **3**, 1739–1751 (2012).
180. K. A. Vincent, A. Parkin, F. A. Armstrong, Investigating and exploiting the electrocatalytic properties of hydrogenases. *Chem Rev* **107**, 4366–4413 (2007).
181. Y. Nicolet, C. Piras, P. Legrand, C. E. Hatchikian, J. C. Fontecilla-Camps, *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* **7**, 13–23 (1999).
182. J. C. Fontecilla-Camps, A. Volbeda, C. Cavazza, Y. Nicolet, Structure/function relationships of [NiFe]- and [FeFe]-hydrogenases. *Chem Rev* **107**, 4273–4303 (2007).
183. B. J. Lemon, J. W. Peters, Binding of exogenously added carbon monoxide at the active site of the iron-only hydrogenase (Cpl) from *Clostridium pasteurianum*. *Biochemistry* **38**, 12969–12973 (1999).
184. Y. L. Li, Y. Wu, J. Wei, J. Wei, B. Xie, L. K. Zou, J. Cheng, Z. Wang, J. He, M. L. Wu, P. H. Zhao, Investigations on the synthesis, structural characterization and electrochemical properties of diiron azadithiolate complexes and phosphine-substituted derivatives. *Polyhedron* **135**, 231–236 (2017).
185. A. Barth, Infrared spectroscopy of proteins. *Biochim Biophys Acta* **1767**, 1073–1101 (2007).
186. A. Adamska-Venkatesh, D. Krawietz, J. Siebel, K. Weber, T. Happe, E. Reijerse, W. Lubitz, New redox states observed in [FeFe] hydrogenases reveal redox coupling within the H-cluster. *J Am Chem Soc* **136**, 11339–11346 (2014).
187. M. W. Ratzloff, M. B. Wilker, D. W. Mulder, C. E. Lubner, H. Hamby, K. A. Brown, G. Dukovic, P. W. King, Activation Thermodynamics and H/D Kinetic Isotope Effect of the Hox to HredH+ Transition in [FeFe] Hydrogenase. *J Am Chem Soc* **139**, 12879–12882 (2017).
188. W. Lubitz, H. Ogata, O. Rüdiger, E. Reijerse, Hydrogenases. *Chem Rev* **114**, 4081–4148 (2014).
189. A. Silakov, C. Kamp, E. Reijerse, T. Happe, W. Lubitz, Spectroelectrochemical characterization of the active site of the [FeFe] hydrogenase HydA1 from *Chlamydomonas reinhardtii*. *Biochemistry* **48**, 7780–7786 (2009).

190. S. Gao, B. Gu, X. Jiao, Y. Sun, X. Zu, F. Yang, W. Zhu, C. Wang, Z. Feng, B. Ye, Y. Xie, Highly Efficient and Exceptionally Durable CO<sub>2</sub> Photoreduction to Methanol over Freestanding Defective Single-Unit-Cell Bismuth Vanadate Layers. *J Am Chem Soc* **139**, 3438–3445 (2017).
191. A. Adamska, A. Silakov, C. Lambertz, O. Rüdiger, T. Happe, E. Reijerse, W. Lubitz, Identification and characterization of the “super-reduced” state of the H-cluster in [FeFe] hydrogenase: a new building block for the catalytic cycle? *Angew Chem Int Ed Engl* **51**, 11458–11462 (2012).
192. M. Winkler, J. Esselborn, T. Happe, Molecular basis of [FeFe]-hydrogenase function: An insight into the complex interplay between protein and catalytic cofactor. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1827**, 974–985 (2013).
193. B. M. Zeldes, M. W. Keller, A. J. Loder, C. T. Straub, M. W. W. Adams, R. M. Kelly, Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. *Front. Microbiol.* **6** (2015). <https://doi.org/10.3389/fmicb.2015.01209>.
194. A. Juszcak, S. Aono, M. W. W. Adams, The extremely thermophilic eubacterium, *Thermotoga maritima*, contains a novel iron-hydrogenase whose cellular activity is dependent upon tungsten. *Journal of Biological Chemistry* **266**, 13834–13841 (1991).
195. R. Singh, D. White, Y. Demirel, R. Kelly, K. Noll, P. Blum, Uncoupling fermentative synthesis of molecular hydrogen from biomass formation in *Thermotoga maritima*. *Appl Environ Microbiol* **84**, 998–1016 (2018).
196. R. Huber, T. A. Langworthy, H. König, M. Thomm, C. R. Woese, U. B. Sleytr, K. O. Stetter, *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch Microbiol* **144**, 324–333 (1986).
197. C. Schröder, M. Selig, P. Schönheit, Glucose fermentation to acetate, CO<sub>2</sub> and H<sub>2</sub> in the anaerobic hyperthermophilic eubacterium *Thermotoga maritima*: involvement of the Embden-Meyerhof pathway. *Arch Microbiol* **161**, 460–470 (1994).
198. M. F. J. M. Verhagen, T. W. O’Rourke, A. L. Menon, M. W. W. Adams, Fe-only hydrogenase from *Thermotoga maritima*. *Methods Enzymol* **331**, 216–226 (2001).
199. H. Sticht, G. Wildegger, D. Bentrop, B. Darimont, R. Sterner, P. Rösch, An NMR-derived model for the solution structure of oxidized *Thermotoga maritima* 1[Fe<sub>4</sub>-S<sub>4</sub>] ferredoxin. *Eur J Biochem* **237**, 726–735 (1996).
200. G. J. Schut, M. W. W. Adams, The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J Bacteriol* **191**, 4451–4457 (2009).
201. A. Flamholz, E. Noor, A. Bar-Even, W. Liebermeister, R. Milo, Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci U S A* **110**, 10039–10044 (2013).
202. N. Chongdar, P. Rodríguez-Maciá, E. J. Reijerse, W. Lubitz, H. Ogata, J. A. Birrell, Redox tuning of the H-cluster by second coordination sphere amino acids in the sensory [FeFe] hydrogenase from *Thermotoga maritima*. *Chem Sci* **14**, 3682–3692 (2023).
203. M. F. J. M. Verhagen, T. O’Rourke, M. W. W. Adams, The hyperthermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization. *Biochim Biophys Acta* **1412**, 212–229 (1999).

204. M. F. J. M. Verhagen, T. W. O'Rourke, A. L. Menon, M. W. W. Adams, Heterologous expression and properties of the  $\gamma$ -subunit of the Fe-only hydrogenase from *Thermotoga maritima*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1505**, 209–219 (2001).
205. L. Gentry-Dye, Iron coordination and protein-protein interactions of the protein frataxin. (2014).
206. W. Buckel, R. K. Thauer, Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation. *Biochim Biophys Acta* **1827**, 94–113 (2012).
207. A. Katsyv, A. Kumar, P. Saura, M. C. Pöverlein, S. A. Freibert, S. T. Stripp, S. Jain, A. P. Gamiz-Hernandez, V. R. I. Kaila, V. Müller, J. M. Schuller, Molecular Basis of the Electron Bifurcation Mechanism in the [FeFe]-Hydrogenase Complex HydABC. *J Am Chem Soc* **145**, 5696–5709 (2023).
208. Y. Zheng, J. Kahnt, I. H. Kwon, R. I. Mackie, R. K. Thauer, Hydrogen formation and its regulation in *Ruminococcus albus*: involvement of an electron-bifurcating [FeFe]-hydrogenase, of a non-electron-bifurcating [FeFe]-hydrogenase, and of a putative hydrogen-sensing [FeFe]-hydrogenase. *J Bacteriol* **196**, 3840–3852 (2014).
209. S. Wang, H. Huang, J. Kahnt, R. K. Thauer, A reversible electron-bifurcating ferredoxin- and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*. *J Bacteriol* **195**, 1267–1275 (2013).
210. A. Kpebe, M. Benvenuti, C. Guendon, A. Rebai, V. Fernandez, S. Le Laz, E. Etienne, B. Guigliarelli, G. García-Molina, A. L. de Lacey, C. Baffert, M. Brugna, A new mechanistic model for an O<sub>2</sub>-protected electron-bifurcating hydrogenase, Hnd from *Desulfovibrio fructosovorans*. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1859**, 1302–1312 (2018).
211. X. Feng, G. J. Schut, D. K. Haja, M. W. W. Adams, H. Li, Structure and electron transfer pathways of an electron-bifurcating NiFe-hydrogenase. *Sci Adv* **8**, 7546 (2022).
212. G. J. Schut, E. S. Boyd, J. W. Peters, M. W. W. Adams, The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications. *FEMS Microbiol Rev* **37**, 182–203 (2013).
213. A. M. G. Costas, S. Poudel, A. F. Miller, G. J. Schut, R. N. Ledbetter, K. R. Fixen, L. C. Seefeldt, M. W. W. Adams, C. S. Harwood, E. S. Boyd, J. W. Peters, Defining Electron Bifurcation in the Electron-Transferring Flavoprotein Family. *J Bacteriol* **199** (2017).
214. C. E. Wise, A. E. Ledinina, J. L. Yuly, J. H. Artz, C. E. Lubner, The role of thermodynamic features on the functional activity of electron bifurcating enzymes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1862**, 148377 (2021).
215. P. Rich, Chemiosmotic coupling: The cost of living. *Nature* **421**, 583 (2003).
216. D. H. Evans, One-electron and two-electron transfers in electrochemistry and homogeneous solution reactions. *Chem Rev* **108**, 2113–2144 (2008).
217. M. Hans, E. Bill, I. Cirpus, A. J. Pierik, M. Hetzel, D. Alber, W. Buckel, Adenosine Triphosphate-Induced Electron Transfer in 2-Hydroxyglutaryl-CoA Dehydratase from *Acidaminococcus fermentans*<sup>†</sup>. *Biochemistry* **41**, 5873–5882 (2002).

218. W. Buckel, Energy Conservation in Fermentations of Anaerobic Bacteria. *Front Microbiol* **12**, 703525 (2021).
219. R. F. Anderson, Energetics of the one-electron reduction steps of riboflavin, FMN and FAD to their fully reduced forms. *Biochim Biophys Acta* **722**, 158–162 (1983).
220. M. Martinez, Standard Potentials in Aqueous Solutions.  
[https://www.academia.edu/41125380/Standard\\_Potentials\\_in\\_Aqueous\\_Solutions](https://www.academia.edu/41125380/Standard_Potentials_in_Aqueous_Solutions)
221. W. Buckel, R. K. Thauer, Flavin-Based Electron Bifurcation, A New Mechanism of Biological Energy Coupling. *Chem Rev* **118**, 3862–3886 (2018).
222. F. Kremp, J. Roth, V. Müller, The Sporomusa type Nfn is a novel type of electron-bifurcating transhydrogenase that links the redox pools in acetogenic bacteria. *Scientific Reports* **2020 10:1** **10**, 1–14 (2020).
223. R. F. Anderson, Energetics of the one-electron steps in the NAD<sup>+</sup>/NADH redox couple. *Biochim Biophys Acta* **590**, 277–281 (1980).
224. W. Buckel, R. K. Thauer, Flavin-Based Electron Bifurcation, Ferredoxin, Flavodoxin, and Anaerobic Respiration With Protons (Ech) or NAD<sup>+</sup> (Rnf) as Electron Acceptors: A Historical Review. *Front Microbiol* **9** (2018).
225. V. D. Sled, N. I. Rudnitsky, T. Ohnishi, Y. Hatefi, Thermodynamic Analysis of Flavin in Mitochondrial NADH:Ubiquinone Oxidoreductase (Complex I). *Biochemistry* **33**, 10069–10075 (1994).
226. F. Lee, R. And, E. G. Ball, Oxidation-reduction potentials of the cytochrome c system. *Journal of Biological Chemistry* **182**, 17-28 (1950).
227. E. T. Smith, Y. Mao, B. A. Feinberg, J. M. Tomich, T. Iwamoto, J. H. Richards, A totally synthetic histidine-2 ferredoxin: thermal stability and redox properties. *Biochemistry* **30**, 11669–11676 (1991).
228. L. Bergdoll, F. Ten Brink, W. Nitschke, D. Picot, F. Baymann, From low- to high-potential bioenergetic chains: Thermodynamic constraints of Q-cycle function. *Biochim Biophys Acta* **1857**, 1569–1579 (2016).
229. P. Mitchell, Protonmotive redox mechanism of the cytochrome b-c1 complex in the respiratory chain: Protonmotive ubiquinone cycle. *FEBS Lett* **56**, 1–6 (1975).
230. F. Li, J. Hinderberger, H. Seedorf, J. Zhang, W. Buckel, R. K. Thauer, Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*. *J Bacteriol* **190**, 843–850 (2007).
231. S. Wang, H. Huang, J. Moll, R. K. Thauer, NADP<sup>+</sup> reduction with reduced ferredoxin and NADP<sup>+</sup> reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*. *J Bacteriol* **192**, 5115–5123 (2010).
232. W. Nitschke, M. J. Russell, Redox bifurcations: Mechanisms and importance to life now, and at its origin. *BioEssays* **34**, 106–109 (2012).
233. M. K. F. Wikström, J. A. Berden, Oxidoreduction of cytochrome b in the presence of antimycin. *Biochim Biophys Acta* **283**, 403–420 (1972).

234. U. Brandt, Bifurcated ubihydroquinone oxidation in the cytochrome bc 1 complex by protonated charge transfer. *FEBS Lett* **387**, 1–6 (1996).
235. P. R. Rich, Electron and proton transfers through quinones and cytochrome bc complexes. *Biochim Biophys Acta* **768**, 53–79 (1984).
236. N. P. Chowdhury, A. M. Mowafy, J. K. Demmer, V. Upadhyay, S. Koelzer, E. Jayamani, J. Kahnt, M. Hornung, U. Demmer, U. Ermler, W. Buckel, Studies on the mechanism of electron bifurcation catalyzed by electron transferring flavoprotein (Etf) and butyryl-CoA dehydrogenase (Bcd) of *Acidaminococcus fermentans*. *Journal of Biological Chemistry* **289**, 5145–5157 (2014).
237. J. K. Demmer, N. Pal Chowdhury, T. Selmer, U. Ermler, W. Buckel, The semiquinone swing in the bifurcating electron transferring flavoprotein/butyryl-CoA dehydrogenase complex from *Clostridium difficile*. *Nature Communications* **2017 8:1 8**, 1–10 (2017).
238. B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, J. D. Rabinowitz, Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nature Chemical Biology* **5**, 593–599 (2009).
239. J. Liang, H. Huang, S. Wang, Distribution, evolution, catalytic mechanism, and physiological functions of the flavin-based electron-bifurcating NADH-dependent reduced ferredoxin: NADP<sup>+</sup> oxidoreductase. *Front Microbiol* **10**, 427828 (2019).
240. J. K. Demmer, H. Huang, S. Wang, U. Demmer, R. K. Thauer, U. Ermler, Insights into flavin-based electron bifurcation via the nadh-dependent reduced ferredoxin:nadp oxidoreductase structure. *Journal of Biological Chemistry* **290**, 21985–21995 (2015).
241. K. D. Swanson, M. W. Ratzloff, D. W. Mulder, J. H. Artz, S. Ghose, A. Hoffman, S. White, O. A. Zadvornyy, J. B. Broderick, B. Bothner, P. W. King, J. W. Peters, [FeFe]-hydrogenase oxygen inactivation is initiated at the H cluster 2Fe subcluster. *J Am Chem Soc* **137**, 1809–1816 (2015).
242. C. Zhu, J. S. Francisco, Production of hydrogen peroxide enabled by microdroplets. *Proc Natl Acad Sci U S A* **116**, 19222–19224 (2019).
243. C. C. Winterbourn, Toxicity of iron and hydrogen peroxide: the Fenton reaction. *Toxicol Lett* **82–83**, 969–974 (1995).
244. Ranjana sarma, Stepwise [FeFe]-hydrogenase H-cluster assembly revealed in the structure of HydA (DeltaEFG). *Nature* **465**, 248–251 (2010).
245. J. Gutiérrez-Fernández, K. Kaszuba, G. S. Minhas, R. Baradaran, M. Tambalo, D. T. Gallagher, L. A. Sazanov, Key role of quinone in the mechanism of respiratory complex I. *Nature Communications* **11**, 1–17 (2020).
246. J. H. Artz, D. W. Mulder, M. W. Ratzloff, C. E. Lubner, O. A. Zadvornyy, A. X. Levan, S. G. Williams, M. W. W. Adams, A. K. Jones, P. W. King, J. W. Peters, Reduction Potentials of [FeFe]-Hydrogenase Accessory Iron-Sulfur Clusters Provide Insights into the Energetics of Proton Reduction Catalysis. *J Am Chem Soc* **139**, 9544–9550 (2017).
247. C. Gauquelin, C. Baffert, P. Richaud, E. Kamionka, E. Etienne, D. Guieysse, L. Girbal, V. Fourmond, I. André, B. Guigliarelli, C. Léger, P. Soucaille, I. Meynial-Salles, Roles of the F-domain in [FeFe] hydrogenase. *Biochim Biophys Acta Bioenerg* **1859**, 69–77 (2018).

248. D. J. Goetschius, H. Lee, S. Hafenstein, CryoEM reconstruction approaches to resolve asymmetric features. *Adv Virus Res* **105**, 73–91 (2019).
249. N. A. Losey, F. Mus, J. W. Peters, H. M. Le, M. J. McInerney, Syntrophomonas wolfei Uses an NADH-Dependent, Ferredoxin-Independent [FeFe]-Hydrogenase To Reoxidize NADH. *Appl Environ Microbiol* **83**, e01335-17 (2017).
250. N. A. Losey, S. Poudel, E. S. Boyd, M. J. McInerney, The Beta Subunit of Non-bifurcating NADH-Dependent [FeFe]-Hydrogenases Differs From Those of Multimeric Electron-Bifurcating [FeFe]-Hydrogenases. *Front Microbiol* **11** (2020).
251. S. G. Bratsch, Standard Electrode Potentials and Temperature Coefficients in Water at 298.15 K. *J Phys Chem Ref Data* **18**, 1–21 (1989).
252. A. J. deBethune, T. S. Licht, N. Swendeman, The Temperature Coefficients of Electrode Potentials: The Isothermal and Thermal Coefficients—The Standard Ionic Entropy of Electrochemical Transport of the Hydrogen Ion. *J Electrochem Soc* **106**, 616 (1959).
253. K. Schuchmann, V. Müller, A bacterial electron-bifurcating hydrogenase. *J Biol Chem* **287**, 31165–31171 (2012).
254. E. T. Smith, B. A. Feinberg, Redox properties of several bacterial ferredoxins using square wave voltammetry. *J Biol Chem* **265**, 14371–14376 (1990).
255. R. C. Prince, M. W. Adams, Oxidation-reduction properties of the two Fe<sub>4</sub>S<sub>4</sub> clusters in Clostridium pasteurianum ferredoxin. *Journal of Biological Chemistry* **262**, 5125–5128 (1987).
256. A. K. Kaster, J. Moll, K. Parey, R. K. Thauer, Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc Natl Acad Sci U S A* **108**, 2981–2986 (2011).
257. A. Katsyv, S. Jain, M. Basen, V. Müller, Electron carriers involved in autotrophic and heterotrophic acetogenesis in the thermophilic bacterium Thermoanaerobacter kivui. *Extremophiles* **25**, 513–526 (2021).
258. J. E. Schreier, A. L. Reysenbach, Deep-Sea Ridges, Microbiology. *Encyclopedia of Ocean Sciences*, 291–298 (2019).
259. D. McLaggan, J. Stephen, I. R. Booth, Chapter 5 Regulation of cytoplasmic pH in bacteria. *Principles of Medical Biology* **9**, 65–77 (1998).
260. J. Jeong, N. Cho, D. Jung, D. Bang, Genome-scale genetic engineering in Escherichia coli. *Biotechnol Adv* **31**, 804–810 (2013).
261. T. M. Wannier, A. Nyerges, H. M. Kuchwara, M. Czikkely, D. Balogh, G. T. Filsinger, N. C. Borders, C. J. Gregg, M. J. Lajoie, X. Rios, C. Pál, G. M. Church, Improved bacterial recombineering by parallelized protein discovery. *Proc Natl Acad Sci U S A* **117**, 13689–13698 (2020).
262. R. D. Arroyo-Olarte, R. Bravo Rodríguez, E. Morales-Ríos, Genome Editing in Bacteria: CRISPR-Cas and Beyond. *Microorganisms 2021, Vol. 9, Page 844* **9**, 844 (2021).

263. N. S. McCarty, A. E. Graham, L. Studená, R. Ledesma-Amaro, Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nature Communications* 2020 11:1 **11**, 1–13 (2020).
264. Lambda Red: A Homologous Recombination-based Technique for Genetic Engineering. <https://blog.addgene.org/lambda-red-a-homologous-recombination-based-technique-for-genetic-engineering>.
265. N. Costantino, D. L. Court, Enhanced levels of  $\lambda$  Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100**, 15748 (2003).
266. Y. Li, Z. Lin, C. Huang, Y. Zhang, Z. Wang, Y. Jie Tang, T. Chen, X. Zhao, Metabolic engineering of *Escherichia coli* using CRISPR-Cas9 mediated genome editing. *Metab Eng* **31**, 13–21 (2015).
267. Cas9 Lambda Red and gRNA Plasmids For CRISPR-Mediated Recombineering in *E. coli*. <https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/216/834/crisprbacdbul.pdf>
268. W. Jiang, D. Bikard, D. Cox, F. Zhang, L. A. Marraffini, RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* 2013 31:3 **31**, 233–239 (2013).
269. GRNA-Cas9 - CRISPR gene editing - Wikipedia, Mariuswalter - [https://en.wikipedia.org/wiki/CRISPR\\_gene\\_editing](https://en.wikipedia.org/wiki/CRISPR_gene_editing), CC BY-SA 4.0.
270. D. Gleditsch, P. Pausch, H. Müller-Esparza, A. Özcan, X. Guo, G. Bange, L. Randau, PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol* **16**, 504–517 (2019).
271. W. Wei, P. Sun, Z. Li, K. Song, W. Su, B. Wang, Y. Liu, J. Zhao, A surface-display biohybrid approach to light-driven hydrogen production in air. *Sci Adv* **4** (2018).
272. J. Y. H. Kim, B. H. Jo, H. J. Cha, Production of biohydrogen by recombinant expression of [NiFe]-hydrogenase 1 in *Escherichia coli*. *Microb Cell Fact* **9**, 1–10 (2010).
273. Part:BBa K1958000 - parts.igem.org. [http://parts.igem.org/Part:BBa\\_K1958000](http://parts.igem.org/Part:BBa_K1958000).
274. A. Volbeda, C. Darnault, A. Parkin, F. Sargent, F. A. Armstrong, J. C. Fontecilla-Camps, Crystal structure of the O<sub>2</sub>-tolerant membrane-bound hydrogenase 1 from *Escherichia coli* in complex with its cognate cytochrome b. *Structure* **21**, 184–190 (2012).
275. A. Volbeda, J. M. Mouesca, C. Darnault, M. M. Roessler, A. Parkin, F. A. Armstrong, J. C. Fontecilla-Camps, X-ray structural, functional and computational studies of the O<sub>2</sub>-sensitive *E. coli* hydrogenase-1 C19G variant reveal an unusual [4Fe-4S] cluster. *Chem Commun (Camb)* **54**, 7175–7178 (2018).
276. O. Kokhan, V. P. Shinkarev, C. A. Wraight, Binding of imidazole to the heme of cytochrome c1 and inhibition of the bc1 complex from *Rhodobacter sphaeroides*: I. Equilibrium and modeling studies. *J Biol Chem* **285**, 22513–22521 (2010).
277. B. Soboh, L. Adrian, S. T. Stripp, An in vitro reconstitution system to monitor iron transfer to the active site during the maturation of [NiFe]-hydrogenase. *Journal of Biological Chemistry* **298**, 102291 (2022).

278. M. Huber, G. Faure, S. Laass, E. Kolbe, K. Seitz, C. Wehrheim, Y. I. Wolf, E. V. Koonin, J. Soppa, Translational coupling via termination-reinitiation in archaea and bacteria. *Nature Communications* **10**, 1–11 (2019).
279. B. C. Berks, F. Sargent, T. Palmer, The Tat protein export pathway. *Mol Microbiol* **35**, 260–274 (2000).
280. L. Bowman, T. Palmer, F. Sargent, A regulatory domain controls the transport activity of a twin-arginine signal peptide. *FEBS Lett* **587**, 3365–3370 (2013).
281. X. Chen, J. L. Zaro, W. C. Shen, Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357–1369 (2013).
282. J. C. Ayala, E. Pimienta, C. Rodríguez, J. Anné, C. Vallín, M. T. Milanés, E. King-Batsios, K. Huygen, L. Van Mellaert, Use of Strep-tag II for rapid detection and purification of Mycobacterium tuberculosis recombinant antigens secreted by Streptomyces lividans. *J Microbiol Methods* **94**, 192–198 (2013).
283. T. G. M. Schmidt, J. Koepke, R. Frank, A. Skerra, Molecular interaction between the Strep-tag affinity peptide and its cognate target, streptavidin. *J Mol Biol* **255**, 753–766 (1996).
284. X. T. Li, L. C. Thomason, J. A. Sawitzke, N. Costantino, D. L. Court, Positive and negative selection using the tetA-sacB cassette: recombineering and P1 transduction in *Escherichia coli*. *Nucleic Acids Res* **41** (2013).
285. M. Ehrenberg, Hypothesis: hypersensitive plasmid copy number control for ColE1. *Biophys J* **70**, 135–145 (1996).
286. B. Gibson, D. J. Wilson, E. Feil, A. Eyre-Walker, The distribution of bacterial doubling times in the wild. *Proc Biol Sci* **285** (2018).
287. M. G. Thompson, N. Sedaghatian, J. F. Barajas, M. Wehrs, C. B. Bailey, N. Kaplan, N. J. Hillson, A. Mukhopadhyay, J. D. Keasling, Isolation and characterization of novel mutations in the pSC101 origin that increase copy number. *Scientific Reports* **2018 8:1 8**, 1–11 (2018).
288. C. DP, The fermentation pathways of *Escherichia coli*. *FEMS Microbiol Rev* **5**, 223–234 (1989).
289. S. T. Odonkor, J. K. Ampofo, *Escherichia coli* as an indicator of bacteriological quality of water: an overview. *Microbiol Res (Pavia)* **4**, 2 (2013).
290. F. Sargent, The Model [NiFe]-Hydrogenases of *Escherichia coli*. *Adv Microb Physiol* **68**, 433–507 (2016).
291. K. Trchounian, A. Trchounian, Hydrogenase 2 is most and hydrogenase 1 is less responsible for H<sub>2</sub> production by *Escherichia coli* under glycerol fermentation at neutral and slightly alkaline pH. *Int J Hydrogen Energy* **34**, 8839–8845 (2009).
292. K. Trchounian, V. Sanchez-Torres, T. K. Wood, A. Trchounian, *Escherichia coli* hydrogenase activity and H<sub>2</sub> production under glycerol fermentation at a low pH. *Int J Hydrogen Energy* **36**, 4323–4331 (2011).
293. L. Brøndsted, T. Atlung, Effect of growth conditions on expression of the acid phosphatase (cyx-appA) operon and the *appY* gene, which encodes a transcriptional activator of *Escherichia coli*. *J Bacteriol* **178**, 1556–1564 (1996).

294. L. Brondsted, T. Atlung, Anaerobic regulation of the hydrogenase 1 (*hya*) operon of *Escherichia coli*. *J Bacteriol* **176**, 5423–5428 (1994).
295. P. W. King, A. E. Przybyla, Response of *hya* expression to external pH in *Escherichia coli*. *J Bacteriol* **181**, 5250–5256 (1999).
296. P. I. Nickel, M. J. Pettinari, M. C. Ramírez, M. A. Galvagno, B. S. Méndez, *Escherichia coli arca* mutants: metabolic profile characterization of microaerobic cultures using glycerol as a carbon source. *J Mol Microbiol Biotechnol* **15**, 48–54 (2008).
297. N. Derdouri, N. Ginet, Y. Denis, M. Ansaldi, A. Battesti, Pleiotropic effects on *E. coli* physiology of the AraC-like regulator from prophage origin, AppY. *bioRxiv*, 2022.02.28.482378 (2022).
298. J. R. Guest, Anaerobic growth of *Escherichia coli* K12 with fumarate as terminal electron acceptor. Genetic studies with menaquinone and fluoroacetate-resistant mutants. *J Gen Microbiol* **115**, 259–271 (1979).
299. J. S. McDowall, M. C. Hjersing, T. Palmer, F. Sargent, Dissection and engineering of the *Escherichia coli* formate hydrogenlyase complex. *FEBS Lett* **589**, 3141–3147 (2015).
300. C. Pinske, R. G. Sawers, The importance of iron in the biosynthesis and assembly of [NiFe]-hydrogenases. *Biomol Concepts* **5**, 55–70 (2014).
301. R. M. Evans, P. A. Ash, S. E. Beaton, E. J. Brooke, K. A. Vincent, S. B. Carr, F. A. Armstrong, Mechanistic Exploitation of a Self-Repairing, Blocked Proton Transfer Pathway in an O<sub>2</sub>-Tolerant [NiFe]-Hydrogenase. *J Am Chem Soc* **140**, 10208–10220 (2018).
302. L. Zhang, G. Morello, S. B. Carr, F. A. Armstrong, Aerobic Photocatalytic H<sub>2</sub> Production by a [NiFe] Hydrogenase Engineered to Place a Silver Nanocluster in the Electron Relay. *J Am Chem Soc* **142**, 12699–12707 (2020).
303. R. M. Evans, E. J. Brooke, S. A. M. Wehlin, E. Nomerotskaia, F. Sargent, S. B. Carr, S. E. V. Phillips, F. A. Armstrong, Mechanism of hydrogen activation by [NiFe] hydrogenases. *Nat Chem Biol* **12**, 46–50 (2016).
304. E. J. Brooke, R. M. Evans, S. T. A. Islam, G. M. Roberts, S. A. M. Wehlin, S. B. Carr, S. E. V. Phillips, F. A. Armstrong, Importance of the active site “canopy” residues in an O<sub>2</sub>-tolerant [NiFe]-hydrogenase. *Biochemistry* **56**, 132–142 (2017).
305. A. Volbeda, P. Amara, C. Darnault, J. M. Mouesca, A. Parkin, M. M. Roessler, F. A. Armstrong, J. C. Fontecilla-Camps, X-ray crystallographic and computational studies of the O<sub>2</sub>-tolerant [NiFe]-hydrogenase 1 from *Escherichia coli*. *Proc Natl Acad Sci USA* **109**, 5305–5310 (2012).
306. S. Hartwig, C. Pinske, R. G. Sawers, Chromogenic assessment of the three molybdo-selenoprotein formate dehydrogenases in *Escherichia coli*. *Biochem Biophys Rep* **1**, 62–67 (2015).
307. M. J. Lukey, A. Parkin, M. M. Roessler, B. J. Murphy, J. Harmer, T. Palmer, F. Sargent, F. A. Armstrong, How *Escherichia coli* is equipped to oxidize hydrogen under different redox conditions. *J Biol Chem* **285**, 3928–3938 (2010).
308. C. Pinske, M. Jaroschinsky, F. Sargent, G. Sawers, Zymographic differentiation of [NiFe]-Hydrogenases 1, 2 and 3 of *Escherichia coli* K-12. *BMC Microbiol* **12**, 1–12 (2012).

309. R. A. Rothery, F. Blasco, A. Magalon, J. H. Weiner, The diheme cytochrome b subunit (NarI) of *Escherichia coli* nitrate reductase A (NarGHI): structure, function, and interaction with quinols. *J Mol Microbiol Biotechnol* **3**, 273–283 (2001).
310. R. G. SAWERS, D. H. BOXER, Purification and properties of membrane-bound hydrogenase isoenzyme 1 from anaerobically grown *Escherichia coli* K12. *Eur J Biochem* **156**, 265–275 (1986).
311. J. A. Cracknell, K. A. Vincent, M. Ludwig, O. Lenz, B. Friedrich, F. A. Armstrong, Enzymatic oxidation of H<sub>2</sub> in atmospheric O<sub>2</sub>: the electrochemistry of energy generation from trace H<sub>2</sub> by aerobic microorganisms. *J Am Chem Soc* **130**, 424–425 (2007).
312. L. A. Flanagan, A. Parkin, Electrochemical insights into the mechanism of NiFe membrane-bound hydrogenases. *Biochem Soc Trans* **44**, 315–328 (2016).
313. P. Wulff, C. Thomas, F. Sargent, F. A. Armstrong, How the oxygen tolerance of a [NiFe]-hydrogenase depends on quaternary structure. *Journal of Biological Inorganic Chemistry* **21**, 121–134 (2016).
314. P. Wulff, C. C. Day, F. Sargent, F. A. Armstrong, How oxygen reacts with oxygen-tolerant respiratory [NiFe]-hydrogenases. *Proc Natl Acad Sci U S A* **111**, 6606–6611 (2014).
315. S. E. Beaton, R. M. Evans, A. J. Finney, C. M. Lamont, F. A. Armstrong, F. Sargent, S. B. Carr, The structure of hydrogenase-2 from *Escherichia coli*: implications for H<sub>2</sub>-driven proton pumping. *Biochem J* **475**, 1353–1370 (2018).
316. L. Bowman, L. Flanagan, P. K. Fyfe, A. Parkin, W. N. Hunter, F. Sargent, How the structure of the large subunit controls function in an oxygen-tolerant [NiFe]-hydrogenase. *Biochemical Journal* **458**, 449 (2014).
317. J. Fritsch, P. Scheerer, S. Frielingsdorf, S. Kroschinsky, B. Friedrich, O. Lenz, C. M. T. Spahn, The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. *Nature* **2011 479:7372** **479**, 249–252 (2011).
318. Y. Shomura, K. S. Yoon, H. Nishihara, Y. Higuchi, Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase. *Nature* **479**, 253–256 (2011).
319. H. Ogata, P. Kellers, W. Lubitz, The Crystal Structure of the [NiFe] Hydrogenase from the Photosynthetic Bacterium *Allochromatium vinosum*: Characterization of the Oxidized Enzyme (Ni-A State). *J Mol Biol* **402**, 428–444 (2010).
320. M. C. Marques, C. Tapia, O. Gutiérrez-Sanz, A. R. Ramos, K. L. Keller, J. D. Wall, A. L. De Lacey, P. M. Matias, I. A. C. Pereira, The direct role of selenocysteine in [NiFeSe] hydrogenase maturation and catalysis. *Nat Chem Biol* **13**, 544–550 (2017).
321. E. Siebert, A. Schmidt, S. Frielingsdorf, J. Kalms, U. Kuhlmann, O. Lenz, P. Scheerer, I. Zebger, P. Hildebrandt, Resonance Raman spectroscopic analysis of the iron–sulfur cluster redox chain of the *Ralstonia eutropha* membrane-bound [NiFe]-hydrogenase. *Journal of Raman Spectroscopy* **52**, 2621–2632 (2021).
322. O. Lenz, L. Lauterbach, S. Frielingsdorf, O<sub>2</sub>-tolerant [NiFe]-hydrogenases of *Ralstonia eutropha* H16: Physiology, molecular biology, purification, and biochemical analysis. *Methods Enzymol* **613**, 117–151 (2018).

323. S. Frielingsdorf, T. Schubert, A. Pohlmann, O. Lenz, B. Friedrich, A trimeric supercomplex of the oxygen-tolerant membrane-bound [NiFe]-hydrogenase from *Ralstonia eutropha* H16. *Biochemistry* **50**, 10836–10843 (2011).
324. D. Kampjut, J. Steiner, L. A. Sazanov, Cryo-EM grid optimization for membrane proteins. *iScience* **24** (2021).
325. X. Yao, X. Fan, N. Yan, Cryo-EM analysis of a membrane protein embedded in the liposome. *Proc Natl Acad Sci U S A* **117**, 18497–18503 (2020).
326. J. Thoma, B. M. Burmann, Fake It ‘Till You Make It—The Pursuit of Suitable Membrane Mimetics for Membrane Protein Biophysics. *International Journal of Molecular Sciences* **2021**, Vol. 22, Page 50 **22**, 50 (2020).
327. L. M. Armijo, S. J. Wawrzyniec, M. Kopciuch, Y. I. Brandt, A. C. Rivera, N. J. Withers, N. C. Cook, D. L. Huber, T. C. Monson, H. D. C. Smyth, M. Osiński, Antibacterial activity of iron oxide, iron nitride, and tobramycin conjugated nanoparticles against *Pseudomonas aeruginosa* biofilms. *J Nanobiotechnology* **18** (2020).
328. S. Friedenauer, H. H. Berlet, Sensitivity and variability of the Bradford protein assay in the presence of detergents. *Anal Biochem* **178**, 263–268 (1989).
329. B. Li, D. Zhu, H. Shi, X. Zhang, Effect of charge on protein preferred orientation at the air–water interface in cryo-electron microscopy. *J Struct Biol* **213**, 107783 (2021).
330. J. Chen, A. J. Noble, J. Y. Kang, S. A. Darst, Eliminating effects of particle adsorption to the air/water interface in single-particle cryo-electron microscopy: Bacterial RNA polymerase and CHAPSO. *J Struct Biol X* **1**, 100005 (2019).
331. S. Chen, J. Li, K. R. Vinothkumar, R. Henderson, Interaction of human erythrocyte catalase with air–water interface in cryoEM. *Microscopy* **71**, i51–i59 (2022).
332. S. Wrathall, Exploring active site structural dynamics in NiFe-hydrogenases with ultrafast pump-probe and two-dimensional infrared spectroscopy. (2022).
333. J. N. Blaza, H. R. Bridges, D. Aragão, E. A. Dunn, A. Heikal, G. M. Cook, Y. Nakatani, J. Hirst, The mechanism of catalysis by type-II NADH:quinone oxidoreductases. *Sci Rep* **7** (2017).
334. A. J. Y. Jones, J. N. Blaza, H. R. Bridges, B. May, A. L. Moore, J. Hirst, A Self-Assembled Respiratory Chain that Catalyzes NADH Oxidation by Ubiquinone-10 Cycling between Complex I and the Alternative Oxidase. *Angew Chem Int Ed Engl* **55**, 728–731 (2016).
335. J. N. Blaza, K. R. Vinothkumar, J. Hirst, Structure of the Deactive State of Mammalian Respiratory Complex I. *Structure(London, England:1993)* **26**, 312 (2018).
336. J. A. Lebitania, N. Inada, M. Morimoto, J. You, M. Shahiduzzaman, T. Taima, K. Hirata, T. Fukuma, A. Ohta, T. Asakawa, H. Asakawa, Local Cross-Coupling Activity of Azide-Hexa(ethylene glycol)-Terminated Self-Assembled Monolayers Investigated by Atomic Force Microscopy. *Langmuir* **37**, 14688–14696 (2021).
337. D. P. Klebl, Y. Wang, F. Sobott, R. F. Thompson, S. P. Muench, It started with a Cys: Spontaneous cysteine modification during cryo-EM grid preparation. *Front Mol Biosci* **9** (2022).

338. S. Aiyer, T. S. Strutzenberg, M. E. Bowman, J. P. Noel, D. Lyumkis, Single-Particle Cryo-EM Data Collection with Stage Tilt using Leginon. *J Vis Exp* **2022** (2022).
339. A. Flayhan, H. D. T. Mertens, Y. Ural-Blimke, M. Martinez Molledo, D. I. Svergun, C. Löw, Saposin Lipid Nanoparticles: A Highly Versatile and Modular Tool for Membrane Protein Research. *Structure* **26**, 345 (2018).
340. M. L. Carlson, J. W. Young, Z. Zhao, L. Fabre, D. Jun, J. Li, J. Li, H. S. Dhupar, I. Wason, A. T. Mills, J. T. Beatty, J. S. Klassen, I. Rouiller, F. Duong, The Peptidisc, a simple method for stabilizing membrane proteins in detergent-free solution. *Elife* **7** (2018).
341. D. J. K. Swainsbury, S. Scheidelaar, N. Foster, R. van Grondelle, J. A. Killian, M. R. Jones, The effectiveness of styrene-maleic acid (SMA) copolymers for solubilisation of integral membrane proteins from SMA-accessible and SMA-resistant membranes. *Biochim Biophys Acta* **1859**, 2133 (2017).
342. S. Inouye, M. Inouye, Up-promoter mutations in the *lpp* gene of *Escherichia coli*. *Nucleic Acids Res* **13** (1985).
343. L. R. Lynd, G. T. Beckham, A. M. Guss, L. N. Jayakody, E. M. Karp, C. Maranas, R. L. McCormick, D. Amador-Noguez, Y. J. Bomble, B. H. Davison, C. Foster, M. E. Himmel, E. K. Holwerda, M. S. Laser, C. Y. Ng, D. G. Olson, Y. Román-Leshkov, C. T. Trinh, G. A. Tuskan, V. Upadhayay, D. R. Vardon, L. Wang, C. E. Wyman, Toward low-cost biological and hybrid biological/catalytic conversion of cellulosic biomass to fuels. *Energy Environ Sci* **15**, 938–990 (2022).
344. R. Gonzalez, R. Phillips, D. Saloni, H. Jameel, R. Abt, A. Pirraglia, J. Wright, Biomass to energy costs. *Bioresources* **6**, 2954–2976 (2011).
345. T. A. Rogers, A. S. Bommarius, Utilizing Simple Biochemical Measurements to Predict Lifetime Output of Biocatalysts in Continuous Isothermal Processes. *Chem Eng Sci* **65**, 2118–2124 (2010).
346. J. Brask, C. C. Fufhsang, O. Kirk, Cost of enzymes, *Novoenzymes A/S* (2016). <https://studylib.net/doc/7258453/cost-of-enzymes>.
347. High-fructose corn syrups (HFCS). <https://water.lsbu.ac.uk/water/enztech/hfcs.html>.
348. M. F. (Martin F.) Chaplin, C. Bucke, Enzyme technology. 264 (1990).
349. X. Wang, T. Saba, H. H. P. Yiu, R. F. Howe, J. A. Anderson, J. Shi, Cofactor NAD(P)H Regeneration Inspired by Heterogeneous Pathways. *Chem* **2**, 621–654 (2017).
350. A. Angelastro, W. M. Dawson, L. Y. P. Luk, R. K. Allemann, A Versatile Disulfide-Driven Recycling System for NADP<sup>+</sup> with High Cofactor Turnover Number. *ACS Catal* **7**, 1025–1029 (2017).
351. H. Y. Jia, M. H. Zong, G. W. Zheng, N. Li, Myoglobin-Catalyzed Efficient in Situ Regeneration of NAD(P)<sup>+</sup> and Their Synthetic Biomimetic for Dehydrogenase-Mediated Oxidations. *ACS Catal* **9**, 2196–2202 (2019).
352. Y. Hong, A. S. Nizami, M. Pour Bafrani, B. A. Saville, H. L. Maclean, Impact of cellulase production on environmental and financial metrics for lignocellulosic ethanol. *Biofuels, Bioproducts and Biorefining* **7**, 303–313 (2013).