



**HAL**  
open science

## The dyad of the Y-junction- and a flavin module unites diverse redox enzymes

Kilian Zuchan, Frauke Baymann, Carole Baffert, Myriam Brugna, Wolfgang Nitschke

### ► To cite this version:

Kilian Zuchan, Frauke Baymann, Carole Baffert, Myriam Brugna, Wolfgang Nitschke. The dyad of the Y-junction- and a flavin module unites diverse redox enzymes. *Biochimica biophysica acta (BBA) - Bioenergetics*, 2021, 1862 (6), pp.148401. 10.1016/j.bbabbio.2021.148401 . hal-03169066

**HAL Id: hal-03169066**

**<https://hal.science/hal-03169066v1>**

Submitted on 15 Mar 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The dyad of the Y-junction- and a flavin module unites diverse redox enzymes

Kilian Zuchan<sup>a</sup>, Frauke Baymann<sup>a</sup>, Carole Baffert<sup>a</sup>, Myriam Brugna<sup>a,\*</sup>, Wolfgang Nitschke<sup>a</sup>

<sup>a</sup>*Aix Marseille Univ, CNRS, BIP, 31 Chemin Joseph Aiguier, 13402, Marseille Cedex 09, France*

---

## Abstract

The concomitant presence of two distinctive polypeptide modules, which we have chosen to denominate as the “Y-junction” and the “flavin” module, is observed in 3D structures of enzymes as functionally diverse as complex I, NAD(P)-dependent [NiFe]-hydrogenases and NAD(P)-dependent formate dehydrogenases. Amino acid sequence conservation furthermore suggests that both modules are also part of NAD(P)-dependent [FeFe]-hydrogenases for which no 3D structure model is available yet. The flavin module harbours the site of interaction with the substrate NAD(P) which exchanges two electrons with a strictly conserved flavin moiety. The Y-junction module typically contains four iron-sulphur centres arranged to form a Y-shaped electron transfer conduit and mediates electron transfer between the flavin module and the catalytic units of the respective enzymes. The Y-junction module represents an electron transfer hub with three potential electron entry/exit sites. The pattern of specific redox centres present both in the Y-junction and the flavin module is correlated to present knowledge of these enzymes’ functional properties. We have searched publicly accessible genomes for gene clusters containing both the Y-junction and the flavin module to assemble a comprehensive picture of the diversity of enzymes harbouring this dyad of modules and to reconstruct their phylogenetic relationships. These analyses indicate the presence of the dyad already in the last universal common ancestor and the emergence of complex I’s EFG-module out of a subgroup of NAD(P)-dependent formate dehydrogenases.

*Keywords:* hydrogenase, complex I, formate dehydrogenase, electron bifurcation, evolution, flavin

---

---

\*Corresponding author

*Email address:* [mbrugna@imm.cnrs.fr](mailto:mbrugna@imm.cnrs.fr) (Myriam Brugna)

## Highlights

- a dyad of protein modules confers NAD(P)-dependence to otherwise unrelated enzymes
- the “flavin module” accomplishes the interfacing of 2- and 1-electron redox reactions
- the “Y-junction module” permits branched electron transfer
- the dyad of modules likely makes energy converting electron bifurcation possible
- phylogenetic analyses of the two modules inform on the evolution of complex I EFG-module

## Abbreviations

**NAD(P)** nicotinamide adenine dinucleotide (phosphate)

**FMN** flavin mononucleotide

***bis*PGD** *bis* pyranopterin guanosine dinucleotide

**ETF** electron transfer flavoprotein

**Nfn** NADH-dependent reduced ferredoxin:NADP<sup>+</sup> oxidoreductase

## 1. Introduction

Enzyme subunits (or N- and C-terminal subunit-extensions thereof) rich in iron-sulphur centres are found in a plethora of redox enzymes. Such [Fe-S]-cluster-bearing units frequently are variations on one of the widespread structural themes referred to as dicubane- (or bacterial [8Fe-8S]-), tetracubane- ([16Fe-16S]-, [1]) or plant-type ([2Fe-2S]-) ferredoxins or feature more divergent architectures (*e.g.* the proximal cluster of the O<sub>2</sub>-tolerant [NiFe]-hydrogenases [2]). They are in general considered to play the role of simple electron transfer relays. The [FeFe]-hydrogenase from *Clostridium (C.) pasteurianum* contains such an iron-sulphur-centre rich domain as an N-terminal extension to the actual catalytic domain [3, 4]. The high-resolution structure of this enzyme showed that the set of [Fe-S]-clusters in this domain are arranged in a Y-shaped geometry [3]. Intriguingly, [Fe-S]-cluster-domains with strikingly similar molecular architecture have turned up in several 3D structure models of very diverse enzymes, such as complex I [5] or specific types of [NiFe]-hydrogenases [6] and most recently in a subgroup of formate dehydrogenases [7]. The simultaneous occurrence of this peculiar structural motif in so diverse enzymes begs the question of its functional significance. To advance our understanding of its role we have searched genome databases for further homologs. The results described below strongly suggest a role for this domain which significantly exceeds that of a simple electron wire building block. As emerging from our survey and already from the smaller sample of available 3D-structures, this [Fe-S]-domain occurs in most enzymes together with another functional unit which uses a flavin moiety and one or two iron sulphur centres to interface 1- and 2-electron redox events. Interestingly, many enzymes containing those two structural units were described as performing electron bifurcation [8, 9, 10, 11, 12, 13, 14, 15]. The analysis of the pattern of occurrence of these modules in distinct enzymes, their likely molecular make-up and their phylogenetic grouping provides insights into their functional roles and suggests unexpected evolutionary relationships.

## 2. Methods

### 2.1. Database searches

The protein sequences of the Y-junction module from the structurally known proteins for complex I (NuoG/Nqo3 of *Thermus thermophilus* HB8: WP\_011227701.1), the bidirectional [NiFe]-hydrogenase (HoxU of *Hydrogenophilus thermoluteolus* TH-1: WP\_119335399.1) and [FeFe]-hydrogenase (CpI of *Clostridium pasteurianum* DSM 525: WP\_004455619.1) were used as BLASTp and tBLASTn queries under default settings (BLOSUM62, gap opening penalty=11, gap extension penalty=1, word size=6) to identify homologs in organisms with publicly available complete RefSeq genomes from NCBI until May 2020. Homologous sequences from phylogenetically diverse, representative sequences with e-values below  $e^{-10}$  and operon structures that contain a neighboring flavin-binding protein to ensure a functioning of the protein complex as described in the literature were selected [5, 6, 8].

Trimeric and tetrameric [FeFe]-hydrogenases were selected, when all proteins could be found in one operon, optionally followed by genes coding for hydrogenase maturase proteins (*hydE*, *hydF*, *hydG*) [16]. [NiFe]-hydrogenases were selected when all proteins could be found in one operon optionally with hydrogenase maturase (*hoxW*) factors present alongside [17]. Complex I sequences were selected based on the presence of all 14 genes specifying the minimal structural framework of the mitochondrial type enzyme (*nuoA-nuoN*) encoded in the same operon (to the exception of sequences from the family of Aquificales, where the genes are scattered over the genome, the gene products, however, were shown to assemble [18]). Trimeric and tetrameric formate dehydrogenases were selected provided all proteins could be found in one operon and the catalytic subunit contained a cysteine or selenocysteine at the molybdopterin/tungstopterin binding site, optionally also containing molybdopterin maturation factors (*moaA*, *mobB*, *fdhD*) alongside [19]. Supplementary Table 1 contains a list of specific binding/ligating motifs for each of these cofactors, on which the presence or absence of that cofactor was determined. From this initial dataset genes of biochemically characterised representatives from a wide phylogenetic variety that also fulfilled this criterion were selected and used as queries in a second round of data mining [7, 20, 21, 22, 23, 24].

### 2.2. Structure-informed protein alignment and phylogenetic analysis

Structure image preparation was done in the open-source version of PyMOL (v1.8.4, Schrödinger, LLC). Structural alignment of the Y-junction modules (n=160) were initially performed by the align function in PyMOL and Swiss PDB Viewer, successively constructed using MAFFT-DASH by employing the iterative refinement algorithm, E-INS-i at default settings (BLOSUM62, gap opening penalty=1.53, offset=0; as shown in Supplementary Fig. 1) [25, 26]. For tree reconstruction the alignment of the Y-junction module was trimmed 15 aa before the first coordinating cysteine residue and 15 aa after the last coordinating cysteine residue of the highly conserved iron-sulphur clusters resulting in contiguous sequence stretches of 186-228 amino acids. The alignment of the sequences from the Y-junction modules from representative sequences for which 3D structure models have been determined is shown in Supplementary Fig. 1.

Structure-informed alignments of the RefSeq sequences from the flavin-binding subunit and NuoE/Nqo1 homologs (n=154, 6 [FeFe]-hydrogenases without flavin module excluded) were constructed based on the structural alignments described above and subsequently manually adjusted. Homologous sequences corresponding to NuoE/Nqo1 and NuoF/Nqo2 of the above mentioned structurally known proteins were separately aligned using MAFFT-DASH and trimmed to the last conserved residues. Due to the overall agreement in the architecture of the trees in these two sets of aligned sequences (see results section), the trimmed alignments were concatenated (as shown in Supplementary Fig. 2).

Neighbor-joining clustering of both, the alignments of the Y-junction module (n=160) and the flavin module (n=154) was conducted in ClustalX2.1 under correction for multiple substitutions by including all gaps. Bootstrap values correspond to the frequency of occurrence of nodes in 1000 bootstrap-replicates. Tree representation for the figures were prepared in ITOL v5 and manually annotated [27].

### 3. Results and Discussion

#### 3.1. An [Fe-S]-cluster-bearing module with a conserved structural layout is common to [FeFe]-hydrogenases, NAD-dependent complex I, NAD-dependent [NiFe]-hydrogenases and NAD(P)-dependent formate dehydrogenases

The minimal core of [FeFe]-hydrogenases as exemplified by the enzyme from *Chlamydomonas (C.) reinhardtii* corresponds to a protein harbouring only the catalytic H-cluster, which is formed by a [4Fe-4S]-cluster together with a 2Fe subcluster, the latter containing non-protein ligands such as azadithiolate, carbon monoxide and cyanide groups (PDB: 3LX4, *nota bene*: this structure lacks the 2Fe subcluster [28]). In numerous prokaryotic [FeFe]-hydrogenases, however, this catalytic domain is supplemented by additional N-terminal domains carrying between two (PDB: 1HFE [29]) and four (PDB: 6N59 [4]) iron-sulphur centres which serve as electron wires connecting the H-cluster to physiological electron donors and acceptors. The earliest crystal structure of an [FeFe]-hydrogenase, *i.e.* that of the *Clostridium (C.) pasteurianum* enzyme (Fig. 1C, in blue), revealed that the four iron-sulphur clusters present in this enzyme form a puzzling Y-shaped, *i.e.* branched chain of redox centres (Fig. 1A and 1B, [3]) rather than a linear one as usually observed in electron wire modules [3, 30, 31]. This geometric arrangement suggested the possible existence of two distinct patches on the enzyme surface where electrons may enter or exit. Mutagenesis approaches designed to probe the pertinence of either of the two possible exit/entry pathways found that they both could mediate electron transfer between the ferredoxin and the H-cluster corroborating the idea that the central iron-sulphur cluster acts as a branching point for electron transfer within this module [32]. The polypeptide harbouring these four electron transfer clusters forms a compact (N-terminal) subdomain within the *C. pasteurianum* enzyme. In line with the notion of a Y-shaped arrangement of cofactors discussed for the *C. pasteurianum* hydrogenase, we will in the following refer to this electron transfer unit as the “Y-junction module”.

A surprisingly similar structural unit (Fig. 1E) was early on identified to form the N-terminal domain of the NuoG (Nqo3) subunit of the NADH:quinone oxidoreductase or NAD-dependent version of complex I (Fig. 1F, in red; for a discussion of the intricacies of the complex I nomenclature, see the Supplementary section 1) [33]. The C-terminal part of NuoG, however, is fully unrelated to [FeFe]-hydrogenases but belongs to the superfamily of Mo/W-*bis*PGD enzymes (*alias* the DMSO-reductase superfamily) [34, 35]. Fig. 1B and 1E illustrate the close similarity of the polypeptide fold and iron-sulphur cluster positioning between the module in [FeFe]-hydrogenases and in NAD-dependent complex I, respectively [36]. One iron-sulphur cluster is missing in the latter enzyme, the protein chain harbouring this cluster in the clostridial [FeFe]-hydrogenase, however, is still present. In NAD-dependent complex I the remaining clusters are part of the electron transfer pathway from the NADH-oxidising site to the quinone-reducing one (Fig. 1D).

In the crystal structure of the NAD-reducing [NiFe]-hydrogenase (frequently referred to as “Hox”-type hydrogenase) from *Hydrogenophilus (H.) thermoluteolus* a separate protein subunit bearing [Fe-S]-clusters was found which strongly resembles the Y-junction module (Fig. 1I, in cyan). Like in the NAD-dependent complex I, this module lacks the “C-terminal”

(see section 3.1.1) iron-sulphur cluster as depicted in Fig. 1H. Again, the 3D architectures of the [NiFe]-hydrogenase’s catalytic HoxH and HoxY subunits are unrelated to that of [FeFe]-hydrogenases, but share structural features with the NuoD (Nqo4) and NuoB (Nqo6) proteins of NAD-dependent complex I, respectively, although their redox substrates are completely different [6]. The physiological electron transfer pathway in this class of [NiFe]-hydrogenases is shown in Fig. 1G.

**Figure 1 Electron-transfer pathways through the Y-junction modules from different enzymes. here, (1.5 or 2-column fitting image)**

Finally, the recently published Cryo-EM structure of the NAD-dependent formate dehydrogenase from *Rhodobacter (R.) capsulatus* again features, as anticipated from sequence similarities, the presence of the Y-junction module, in this case containing the full set of iron-sulphur clusters just as in the *C. pasteurianum* enzyme [7]. Following the scheme of the previous three cases mentioned above, Fig. 1 shows the electron transfer (Fig. 1J), the structure of the Y-junction module (Fig. 1K) and its positioning within the parent enzyme (Fig. 1L) in purple. The color coding for the specific enzymes as introduced in Figure 1 right column will be used throughout the manuscript.

Most interestingly, the structure of the entire NAD-dependent formate dehydrogenase from *R. capsulatus* also informs on the mutual spatial arrangement of individual monomers in the physiological dimeric state of the enzyme. This enzyme indeed forms a homodimer of heterotetramers (and was crystalized in that state) wherein the interaction between individual enzymes occurs mainly *via* the surface of the Y-junction module. This arrangement positions the two histidine-coordinated clusters in sufficiently close proximity (edge-to-edge distance of about 10 Å [7]) to allow facile and rapid inter-monomer electron transfer [37]. We have indicated the position of the histidine cluster in the second half of the functional dimer as A4’ in Fig. 1J. Such an organization of potentially cross-talking electron transfer chains is reminiscent of what is observed in the Rieske/cytb complexes (where inter-enzyme electron transfer may occur through the  $b_L$  hemes at an edge-to-edge distance of about 11 Å [38, 39]) or heterodisulfide reductases (over a distance of again about 11 Å [40]). However, while in these latter cases, inter-monomer communication may be afforded by redox centres which are part of the intrinsic electron transfer chain of each monomer, the formate dehydrogenase enzyme from *R. capsulatus* recruits the “out-of-line” clusters to electrically connect the two individual monomers. It is noteworthy that cross-dimer electron transfer in the formate dehydrogenase from *R. capsulatus* only is a theoretical possibility (albeit extremely likely) and no experimental evidence for this phenomenon has been obtained so far. For the Rieske/cytb complex, by contrast, inter-monomer electron transfer has been shown to occur, albeit at slow rates [39]. A physiological organisation as a dimer was observed in several further NAD(P)-dependent formate dehydrogenases and, intriguingly (as further discussed below) also in electron bifurcating NAD(P)-dependent [FeFe]-hydrogenases [8, 13, 10].

### 3.1.1. Structural features of the Y-junction module

As illustrated in Fig. 2A, the entire Y-junction module (always shown in green shades hereafter) is itself composed from three distinct submodules, that is, an N-terminal plant-type [2Fe-2S]-ferredoxin domain (in olive) followed by a median domain (in mint) ligating a single [4Fe-4S]-cluster and finally a C-terminal bacterial-ferredoxin-type domain (in lime) featuring 2 cubane [4Fe-4S]-centres one of which may be absent in certain cases as mentioned above and described in more detail below [41, 34]. The cubane cluster harboured by the median domain features a histidine ligand in addition to the three canonical cysteine ligands (in the following denoted as "histidine-[4Fe-4S]") in all enzymes shown in Fig. 1. Interesting structural and sequence features of this ligating histidine are discussed in the Supplementary section 2 (see also Supplementary Fig. 3). For the sake of annotating the phylogenetic trees which we will present below (section 3.5), we have represented each of the four individual iron-sulphur clusters by a specific symbol and the resulting electron transfer pathways in an iconographic way as defined in Fig. 2B. This depiction is used throughout the article in order to more abstractly present the genomic organisation and structural layouts of the above mentioned protein complexes.

The histidine-[4Fe-4S]-cluster and the C-terminal [4Fe-4S]-centre were found to be missing in certain enzymes. However, most representatives of all enzyme families dealt with in this work contain all four clusters. This and the fact that all four structures shown in Fig. 1 contain the full complement of the polypeptide chain with or without harbouring all four clusters seen in the *C. pasteurianum* structure leads us to conclude that the four-cluster domain represents the archetypal version of the module and that either the histidine-ligated or the C-terminal [4Fe-4S] cluster have been lost in certain cases.

**Figure 2 Domain-architecture of the Y-junction module. here, (1 or 1.5-column fitting image)**

A modular make-up of electron transfer proteins has been pointed out repeatedly over the last two decades [1, 35, 42, 43, 44] and it is by itself therefore not astonishing to find an electron transfer module built-up from more basic components. The strictly conserved succession of subdomains and the astonishingly similar global architecture of the module, however, render the possibility that this structure emerged four times independently in the above considered enzymes from the described more basic building blocks extremely unlikely. The Y-junction module therefore almost certainly originated only once and was then docked *en bloc* onto different catalytic protein units.

### 3.1.2. The Y-junction module at the crossroads of electron transfer chains

The variability of electron transfer pathways through the 4-cluster domain seen in the left column of Fig. 1 exceeds even that identified for the *C. pasteurianum* [FeFe]-hydrogenase. Redox communication with catalytic subunits are seen to occur either *via* the C-terminal [4Fe-4S]-cluster (Fig. 1A and 1J) or the histidine cluster (Fig. 1D and 1G). Both the histidine cluster and the [2Fe-2S]-centre can interact with soluble electron carriers such as ferredoxins and likely also flavodoxins (Fig. 1A) [32]. Alternatively, the histidine cluster is seen to

mediate inter-monomer communication (Fig. 1J) and the [2Fe-2S]-centre can talk to an additional flavin-bearing subunit (Fig. 1D, G and J; see below). The comparison of the cases depicted in the left column of Fig. 1 thus shows that the 4-cluster domain fully merits to be called the Y-junction module.

### 3.1.3. Inter-enzyme similarities exceed the Y-junction module and include a flavin-bearing protein module

While similarity between the *Clostridium* hydrogenase and the other three systems is limited to the Y-junction module, common elements of NAD-dependent complex I, NAD(P)-dependent [NiFe]-hydrogenases and NAD(P)-dependent formate dehydrogenases furthermore comprise a structural module binding a flavin together with two additional iron-sulphur centres, more specifically a [2Fe-2S] and a [4Fe-4S] cluster. In the NAD-dependent [NiFe]-hydrogenase from *H. thermoluteolus*, this unit is contained within a single polypeptide, the HoxF protein, while in the NAD-dependent complex I and NAD(P)-dependent formate dehydrogenase of Fig. 1 it is made up from two subunits (NuoE and NuoF for complex I) which structurally correspond to N- and C-terminal domains of the HoxF protein, respectively (Fig. 3). The N-terminal part of HoxF thus contains the [2Fe-2S] cluster domain (without the Fe-S cluster in this particular structure) and the C-terminal domain corresponds to the flavin/[4Fe-4S]-bearing subunit of the other enzymes. Whereas this [4Fe-4S]-cluster is present in all systems and connects the flavin to the Y-junction module, the [2Fe-2S]-cluster is found in many, but not all representatives of the diverse enzyme families as will be elaborated below. The polypeptide chain harbouring the cluster in the other discussed cases, by contrast, is present, again arguing for a selective loss of this Fe-S cluster in certain enzyme representatives.

This structural unit couples the 2-electron redox conversion of NAD(P)<sup>+</sup>/NAD(P)H to electron transfer chains which perform only 1-electron redox reactions. It therefore serves the function of a 2- to 1-electron gate. Since the functionally pivotal cofactor in this unit is a flavin, we will refer to it as the “flavin-module”.

**Fig.3 Structural comparison of the “flavin module” from different complexes. here, (1 or 1.5-column fitting image, the three images can be changed to horizontal alignment)**

### 3.2. The link to NAD(P)-dependent, electron bifurcating [FeFe]-hydrogenases

The larger segment containing both the flavin- and the Y-junction-module is suggested *via* sequence similarities to also be present in NAD(P)-dependent, electron bifurcating [FeFe]-hydrogenases [8, 16]. These enzymes are generally classed into tri- and tetrameric representatives, while rare monomeric and dimeric enzymes due to subunit fusion exist as well [24](see below). Two of the common subunits constitute the above described flavin module while the third is closely related to the *C. pasteurianum* [FeFe]-hydrogenase. Similar to the *C. pasteurianum* enzyme, the hydrogenase catalytic subunit of these enzymes features the Y-junction module as its N-terminal part followed by the core domain of [FeFe]-hydrogenases [16]. The Y-junction module therefore exists in all four enzymes shown in

Fig. 1. These similarities indicate that bifurcating [FeFe]-type hydrogenases employ structural components of NAD-dependent complex I, NAD(P)-dependent [NiFe]-hydrogenases and NAD(P)-dependent formate dehydrogenases on one side and of the *C. pasteurianum* [FeFe] hydrogenase on the other side. The structural element common to both sides is the Y-junction module and it can therefore serve as a cornerstone for building a homology model of the entire bifurcating [FeFe]-hydrogenase by merging the structural features depicted in Fig. 1 and Fig. 3. Such a model using the NuoEFG (Nqo213) part of complex I as a template has recently been presented by Chongdar *et al.* (2019) for the trimeric hydrogenase of *Thermotoga (T.) maritima* [45]. Still, judging from sequence comparisons, this structural segment of complex I differs from both trimeric and tetrameric electron bifurcating [FeFe]-hydrogenases in several aspects (see Fig. 4). (i) The C-terminal cluster of the Y-junction module is generally absent in complex I. (ii) An N-terminal extension in the flavin/[4Fe-4S] subunit of the flavin module is absent in complex I and also in the structures of the NAD(P)-dependent [NiFe]-hydrogenases. This domain, found to be present in most [FeFe]-hydrogenases and formate dehydrogenases, was even proposed to ligate an additional iron-sulphur cluster [15]. In between the genes encoding the [2Fe-2S]-ligating and the flavin/[4Fe-4S]-bearing subunits and in rare cases in other gene cluster positions, a gene coding for an additional subunit is present in tetrameric [FeFe]-hydrogenases and certain formate dehydrogenases [16]. Sequence analysis indicates homology to the [2Fe-2S]-domain and, more generally, to thioredoxin-like [2Fe-2S]-ferredoxins [46] but the [2Fe-2S] cofactor likely is absent (for a more detailed discussion, see section 3.4.1). (iii) The flavin/[4Fe-4S]-carrying protein features an additional [8Fe-8S]-ferredoxin-like domain as C-terminal extension. Again, neither complex I nor NAD(P)-dependent [NiFe]-hydrogenases possess this domain (see below and Fig. 4).

Since Chongdar *et al.* (2019) proposed their homology model, the above discussed (and shown in Fig. 1) 3D-structures of NAD(P)-dependent formate dehydrogenases have come out [7, 47]. These structures remedy two of the three above mentioned shortcomings entailed by using complex I or NAD(P)-dependent [NiFe]-hydrogenases as templates since both formate dehydrogenase structures contain all four iron sulphur cluster of the Y-junction module and furthermore possess the equivalent of the small subunit/domain in the flavin module. Unfortunately, the C-terminal [8Fe-8S]-extension does not occur in the NAD-dependent formate dehydrogenase enzymes for which structures are now available, although other representatives of this enzyme family contain this domain (see section 3.5.4). While positioning at molecular detail of the C-terminal [8Fe-8S] domain in the flavin/[4Fe-4S] subunit is thus not yet possible, a rough estimate of its approximate configuration with respect to the core of the enzyme formed by the three main subunits can be derived from the spatial localisation of the C-terminus of the flavin/[4Fe-4S]-subunit. Since the [8Fe-8S]-domain represents the C-terminal extension of this subunit, it must be positioned close to where the flavin/[4Fe-4S]-subunits of the enzymes of Fig. 3 have their C-termini.

We have in this work surveyed prokaryotic genomes for homologs of the characterised bifurcating [FeFe]-hydrogenases (to be detailed in section 3.5.2) and the retrieved data set suggests that the structural unit encompassing the NAD-redox-converting flavin module and the Y-junction module is indeed common to NAD-dependent complex I, NAD(P)-dependent

[NiFe]-hydrogenases, NAD(P)-dependent formate dehydrogenases and ferredoxin-dependent (also called electron bifurcating) [FeFe]-hydrogenases.

### 3.3. NAD(P)-dependent [FeFe]-hydrogenases and NAD(P)-dependent formate dehydrogenases: two manifestations of the same electron bifurcating enzyme?

As discussed in section 3.2 and further detailed below (section 3.5 and Fig. 4), the molecular make-up of NAD(P)-dependent [FeFe]-hydrogenases most closely resembles that of NAD(P)-dependent formate dehydrogenases. These two enzymes families share the flavin and the Y-junction module and communicate reducing equivalents (*via* the C-terminal [4Fe-4S]-cluster of the Y-junction module) to and from a catalytic unit which redox-converts either the CO<sub>2</sub>/formate or the H<sup>+</sup>/H<sub>2</sub> couple. We emphasise that the apparent standard electrochemical potentials of these two redox couples are virtually indistinguishable (-414 mV vs. -420 mV at pH 7, respectively) suggesting a facile interchangeability without needing to adapt the redox properties of the cofactors in the Y-junction- and the flavin-module [48].

The similarities in functional properties between these two enzymes are indeed striking. In 2013, the Thauer group reported the biochemical and functional characterisation of a formate dehydrogenase purified from *Gottschalkia (G.) acidurici* which couples NAD<sup>+</sup>-reduction by electrons derived from the oxidation of formate to the concomitant reduction of ferredoxin [11]. This reaction scheme closely resembles the process carried out by electron bifurcating [FeFe]-hydrogenases. It is noteworthy that other representatives of this family have been characterised previously, such as the enzymes from *Methylosinus trichosporium* [49], *Cupriavidus necator* [41] and *Rhodobacter capsulatus* [50] but their potential ferredoxin-dependent reactions weren't analysed, even if Jollie and Lipscomb (1991) [49] found characteristics which in retrospect may be interpreted to suggest electron bifurcating properties of the enzyme. For the two NAD(P)-dependent formate dehydrogenases for which 3D structure models recently became available [7, 51], the implication of ferredoxins hasn't been studied in detail. We deem worth noting that the coupling of NAD(P)- and ferredoxin-redox conversions in the formate dehydrogenase from *G. acidurici* was observed as early as 1972 [52], that is, almost contemporary with the development of the Q-cycle scheme by Peter Mitchell, which rationalised the redox processes in Rieske/cytb complexes through electron bifurcation [53, 54]. However, the conceptual association of the reactions in NAD(P)-dependent [FeFe]-hydrogenases and formate dehydrogenases to that of the Rieske/cytb complex was only made four decades later [36, 55, 56].

The pioneering work on the *G. acidurici* enzyme [52, 11] therefore revealed that selected representatives from the NAD(P)-dependent formate dehydrogenase family can indeed couple ferredoxin reductions/oxidations to their overall redox reactions just as ferredoxin-dependent [FeFe]-hydrogenases do. Both the [FeFe]-hydrogenase [57, 24] and the formate dehydrogenase families [50], however, also seem to include cases which redox convert NAD(P) without implicating redox reactions of ferredoxins.

### 3.4. Functional significance of the Y-junction - flavin module dyad

#### 3.4.1. On the number and roles of the iron-sulphur centres in the flavin module

The 3D-structure model of the flavin module is known for representatives from three out of the four described families. The spatial positioning of the flavin moiety and the cubane iron-sulphur cluster located in between the flavin and the Y-junction module is well-conserved between all these three cases (Fig. 3). However, the [2Fe-2S]-cluster in NuoE (Nqo2) of NAD-dependent complex I-type enzymes (denoted N1a) positioned on the opposite side of the flavin [33], is also seen in the structures of NAD(P)-dependant formate dehydrogenases but is absent in the NAD-reducing [NiFe]-hydrogenase from *H. thermoluteolus*. This latter structure nonetheless exhibits the protein fold ligating this cluster in complex I and formate dehydrogenases (see Fig. 3). However, this cluster is present in phylogenetically close sister enzymes of the *H. thermoluteolus* one as we will show below (section 3.5.3). Indeed, in all four enzyme-families, this [2Fe-2S]-cluster is present in the majority of considered cases as suggested by the conservation of its respective cysteine ligands (Supplementary Tab. 1).

In many NAD-dependent complex I-enzymes studied so far, the N1a cluster is seen to become transiently reduced upon addition of NADH [58]. Deleting the N1a cluster results in an inactive enzyme [59], showing that the presence of N1a and N3 clusters is essential for the function of that complex. A similar observation has been made for the homologous [2Fe-2S]-cluster-bearing HoxE subunit in the [NiFe]-hydrogenase of *Synechocystis sp. PCC 6803*. Deleting this subunit from the complex causes all NAD-related functionality to cease [60]. The reduction of complex I by NADH (just like that of cyanobacterial [NiFe]-hydrogenase) therefore very likely involves an initial bifurcated 2-electron transfer from the flavin to its flanking clusters and the electron “parked” on the [2Fe-2S] centre eventually follows the path of the first electron through the Y-junction module towards the quinone reduction site [61]. As we have argued before, this reaction scheme may allow NAD-dependent complex I to optimize electron transfer from the 2-electron compound NADH towards the 1-electron redox conduit constituted solely from iron-sulphur clusters by striking a compromise between the conflicting requirements of 2- and 1-electron transfer events [62].

Many flavin modules in all families except NAD-dependent complex I harbour one to several additional iron-sulphur centres, the function of which is not known so far. Several of these additional centres are most likely positioned close to the [2Fe-2S]-cluster. It seems inevitable to us that for those cases the scenario of “parking” the second electron involved in flavin redox reactions cannot be applicable. Rather, these clusters almost certainly provide a conduit for shuttling the electron to/from other electron acceptors such as for example soluble ferredoxins [8].

We therefore consider that presence/absence of the [2Fe-2S]-cluster next to the flavin as well as of an additional electron transfer pathway in contact with the [2Fe-2S]-cluster is of prominent importance for the catalytic and kinetic abilities of the respective enzymes in communicating with soluble electron shuttles.

In many hydrogenases and formate dehydrogenases, an additional domain is fused to the N-terminus of the flavin/[4Fe-4S] subunit as is the case in the two formate dehydrogenase enzymes for which a structure is available (Fig. 3). Tetrameric versions of NAD(P)-dependent

[FeFe]-hydrogenases (see section 3.2) and of NAD(P)-dependent formate dehydrogenases contain a further small subunit. In most cases, the gene encoding this protein is flanked upstream by that coding for the [2Fe-2S]-bearing subunit of the flavin-module and downstream by that encoding the flavin/[4Fe-4S] subunit. The gene cluster disposition of these cases and those discussed below are summarised in Fig. 4. As noticed previously [46, 16], sequence comparisons indicate that the small subunit/N-terminal-domain and the [2Fe-2S]-binding domain most likely arose from gene duplications and that they therefore are likely to exhibit a common fold. Meyer *et al.* (2011) pointed out that this fold corresponds to the structure of thioredoxin-like [2Fe-2S]-ferredoxins [46]. The recently obtained structures validate these sequence-based predictions (see Fig. 3). The presence of cysteine residues in the small subunit/N-terminal-domain has led to the proposal that this subunit may add a further iron-sulphur-centre to the cluster inventory of the flavin module [16, 63]. Our multiple alignments of a large sample of small subunit/N-terminal-domain sequences in both hydrogenases and formate dehydrogenases showed that the first cysteine ligating the cluster in thioredoxin-like ferredoxins and in the [2Fe-2S]-domains of the flavin module is absent in these polypeptides. The second one is only sporadically conserved and only the third and fourth cysteines are present in appropriate positions in the majority of sequences from NAD(P)-dependent [FeFe]-hydrogenases. Since only two potential cluster ligands are conserved and since furthermore mutagenesis studies on thioredoxin-like ferredoxins [64, 46] have demonstrated the first cysteine (absent in these sequences) to be indispensable for stabilising an iron-sulphur cluster, we tend to consider it highly unlikely that these domains contribute a further cluster to the inventory of redox centres in the flavin module. In the homologous subunits/domains of NAD(P)-dependent formate dehydrogenases (such as for example the enzyme depicted in Figures 1J-L and 3C), cysteine residues are fully absent furthering our doubts on the existence of additional [Fe-S]-clusters afforded by these polypeptide chains.

Remarkably, in a specific group of formate dehydrogenases, the gene encoding the thioredoxin-like protein and flanked in the gene cluster by ORFs for the [2Fe-2S]-subunit (upstream) and the flavin/[4Fe-4S]-protein (downstream) is replaced by an unrelated gene the product of which is located in the enzyme on the surface of the catalytic subunit [7], that is, in a spatial position completely different from that of the thioredoxin-related small subunit (Fig. 4). We will come back to this observation below (section 3.5.4).

In a few tetrameric versions of NAD(P)-dependent [FeFe]-hydrogenases, a gene coding for a thioredoxin-like protein is additionally found at the end of the gene cluster downstream of the ORF encoding the catalytic subunit. This gene can also be found fused to the 5'-end of the catalytic subunits' gene as is for example the case for the biochemically characterised enzyme from *Thermotoga maritima* [65]. In contrast to what is the case for the above discussed subunit/domain, the sequences of the thioredoxin-like proteins/domains encoded at the end of the gene cluster contain all four canonical cysteines of thioredoxin-like [2Fe-2S]-ferredoxins [46] and they are generally assumed to indeed provide an additional redox centre to the enzyme [65, 45] (see Fig. 4).

**Fig.4 Operon structures here, (1.5 or 2-column fitting image)**

### 3.4.2. Branched electron transfer, electron bifurcation and energy-converting electron bifurcation: a specification of meanings

Branching of electron transfer occurs when single reducing equivalents can take either of two distinct paths leading away from the branchpoint redox centre. Which pathway or to which extent each of the two pathways is taken will depend on relative redox potentials of the involved redox centres, their distances and/or on congesting either pathway by downstream kinetic limitations. The inverse reaction would in this definition be the “merging” of distinct electron transfer chains and the branchpoint redox centre would become the merging cofactor. Electron transfer reactions through this module are exergonic and all individual redox transitions are 1-electron reactions. This type of processes obviously is what the Y-junction module can mediate, since the edge-to-edge distances between each of the surface-oriented clusters, *i.e.* the [2Fe-2S], the histidine-[4Fe-4S] and the C-terminal [4Fe-4S], to the central [4Fe-4S] are around 7 Å (Fig. 2). Such a branching behavior of the Y-junction module has indeed been demonstrated *via* mutational studies for the [FeFe]-hydrogenase from *C. pasteurianum* [32].

Bifurcation in colloquial language may appear quite synonymous to branching. In the field of bioenergetics, however, the term has originally been introduced to describe the quinol oxidation reaction in mitochondrial complex III (a member of the Rieske/cytb family of enzymes) by Peter Mitchell [53, 54] and in particular to rationalise the phenomenon that low potential hemes were seen to be reduced by quinones in a reaction triggered by the addition of oxidants (the so-called “oxidant-induced reduction” [66]). This concept was later expanded by Buckel and Thauer [36, 55, 56, 67, 68] towards flavin-containing enzymes in the aim to explain the observed reduction of low potential ferredoxins concomitant with and coupled to the reduction of oxidants such as NAD(P) or crotonyl-CoA. The phenomenon of the reduction of low potential acceptors is presently understood on the basis of an interfacing of 2-electron redox compounds (such as quinones or flavins) with two distinct 1-electron redox chains. It is this interfacing process which we regard as the actual electron bifurcation/confurcation reaction. The immediate redox partners of the bifurcating centre within these two distinct 1-electron redox chains, however, can be (and mostly are) intra-enzyme redox centres, rather than ferredoxins or flavodoxins.

Quinones and flavins are known to the organic chemists since almost a century as displaying intriguingly versatile redox properties [69, 70] and in particular variable strengths of what is called redox cooperativity between their individual 1-electron redox transitions (for a more detailed discussion, see [62]). While intuitively the loss of the first electron from a fully reduced 2-electron compound should disadvantage departure of the second one (due to loss of electrostatic repulsion), redox cooperativity denotes cases when, by contrast, the first redox transition favours occurrence of the second one. Redox cooperativity thus results in lowering of the difference in redox potentials of the two individual 1-electron transitions and can even lead to so-called “inverted” redox potentials where the first redox transition generates a highly reactive semi-reduced state and thereby triggers the immediate occurrence of the second one, yielding an apparent concomitant 2-electron redox event [71]. The strong reactivity of the semi-reduced state and appropriate tuning of electrochemical potentials of all redox centres involved in the “ $2e^- / 2 \times 1e^-$ ”-interfacing reaction allows the

reduction of low potential acceptors, the hallmark phenomenon of electron bifurcation, to naturally emerge from the energetic landscape of the process [72, 73]. The reaction schemes observed in the Rieske/cytb complexes, in Electron Transfer Flavoproteins (ETFs) and in the NADH-dependent reduced ferredoxin:NADP<sup>+</sup> oxidoreductase (Nfn) enzyme nicely fit this electrochemical scheme [36, 55, 56, 62, 73].

However, as we have argued before [74, 62], the extent of redox inequivalence (resulting from varying strength of redox cooperativity of the 2-electron compound) was found to be variable and likely represents a continuum reaching up to a virtual equivalence as may for example be the case in certain types of complex I (discussed in Baymann *et al.* (2018) [62]). Electron bifurcation defined as  $2e^- / 2 \times 1e^-$ -interfacing thus is not equivalent to the process of reducing low potential acceptors, since the two electrons could be forwarded to two roughly equipotential 1-electron acceptors. Ferredoxin-dependent electron bifurcation is therefore a special case of the wider scheme of electron bifurcation.

To integrate the empirically observed redox reactions performed by the NAD(P)-dependent [FeFe]-hydrogenases and formate dehydrogenases dealt with in this work into the above described conceptual framework, Buckel and Thauer have realised that the most straightforward way implied assuming the presence of two distinct flavin moieties in these enzymes [36]. One of these flavins would represent the  $2e^- / 2 \times 1e^-$  bifurcating site while the second one would consecutively accumulate two electrons to eventually be able to reduce NAD(P) in a 2-electron redox reaction. Unfortunately, the ensemble of structural information accumulated over the last few years and summarised in section 3.1 renders the two-flavin model extremely unlikely. The previous mechanistic rationalisation of what is observed experimentally has thereby lost its basis. To remedy this problem, several alternative mechanistic models have been proposed. For example, Peters *et al.* (2018) [75] considered the possibility that electron bifurcation actually occurs on the H-cluster of the hydrogenase catalytic subunit while Chongdar *et al.* (2019) [45] proposed the Y-junction module to perform this function.

The definitions detailed above make it clear that the Y-junction module in the bifurcating [FeFe]-hydrogenases cannot perform electron bifurcation as defined by NAD-dependent ferredoxin reduction but only electron branching. Such a branching pathway has been proposed to result in redox equilibration with the ferredoxin pool in an attempt to rationalise the observed substoichiometric reduction of ferredoxin by hydrogen [8]. The recent structure model of NAD-dependent formate dehydrogenase from *R. capsulatus*, however, showed that the branching in the Y-junction module might connect individual enzymes of the functional dimer. Assuming that the dimeric forms of electron bifurcating, NAD(P)-dependent [FeFe]-hydrogenases resemble their formate dehydrogenase sister enzymes, an equilibration with the ferredoxin pool through the Y-junction module's histidine cluster is thus precluded. We are also reluctant to consider the H-cluster as the electron bifurcating site. Since NAD(P)-dependent formate dehydrogenases were shown to operate pretty much the same way as the [FeFe]-hydrogenases do, this would imply that the molybdopterin co-factor performs the same redox reaction as the H-cluster is proposed to do in this scheme. While molybdopterins have been shown to feature the crucial electrochemical properties allowing also quinones and flavins to bifurcate electrons [76], it seems difficult to imagine the reduction of soluble ferredoxins at the molybdopterin site in the light of the known

3D-structures of this class of enzymes [77, 7]. To our minds the only potentially electron bifurcating site needs to be the flavin within the flavin module. We admit that an adequate mechanistic theory so far is sorely missing. However, we suspect that the single flavin must somehow integrate the distinct redox functions of the two electrochemically differing flavins hypothesised previously. To accomplish this feat, the electrochemical properties of this single flavin may be more malleable and tunable by external parameters than previously thought. Recent studies pinpointing sequence stretches in the flavin module which distinguish ferredoxin-dependent [FeFe]-hydrogenases from ferredoxin-independent ones [24] or unraveling conformational changes within this module which potentially alter the energetic landscape of the entire bifurcating reaction [78] may provide first clues into how the single flavin can be coaxed into acting in two electrochemically different ways to fulfill the roles of both previously proposed flavin entities.

It is worth restating that the above definition of electron bifurcation thus differs from that frequently used in the literature on [FeFe]-hydrogenases and formate dehydrogenases which calls an enzyme from this group electron bifurcating/confurcating if it reduces/oxidises ferredoxins while it redox-connects the NAD(P)H/NAD(P)<sup>+</sup> couple to the H<sup>+</sup>/H<sub>2</sub> or CO<sub>2</sub>/formate couples. The ferredoxin-reducing process, however, appears as a special case in our definition of electron bifurcation which, by contrast, also considers situations where reducing equivalents on a 2-electron compound are bi-/confurcated to/from two individual 1-electron centres and none of the two electrons goes to or derives from a ferredoxin. Rather than “bifurcating” or “non-bifurcating” [24] we suggest to refer to these enzymes as “ferredoxin-dependent” or “ferredoxin-independent”. Distinguishing the bifurcating 2e<sup>-</sup> / 2 x 1e<sup>-</sup>-reaction from the generation of reduced low potential acceptors allows to integrate the cases of the ferredoxin-independent [FeFe]-hydrogenases [57] and formate dehydrogenases (which may both operate in a way related to complex I, that is, bifurcate electrons towards two distinct 1-electron chains, however without reducing ferredoxins). This distinction may also help to provide insights for understanding the observation of less-than-strict coupling of ferredoxin reduction in certain [FeFe]-hydrogenases [8, 13, 12].

### *3.5. Phylogenetic relationships between and within the four families of NAD(P)-dependent enzymes and mapping of structural/functional particularities onto the tree*

We have used sequences of the Y-junction module as queries against complete RefSeq genomes available in NCBI databases to search for the presence of gene clusters containing this module. In the majority of cases where the sequence of the Y-junction module was detected, the surrounding gene cluster (see Fig. 4) also featured ORFs coding for the flavin module and the individual clusters could be attributed to one of the four enzyme families introduced above plus an additional enzyme family to be discussed below (section 3.5.6) *via* their dissimilar catalytic subunits/domains. However, significant variability in the detailed make-up of these gene clusters was observed resulting in corresponding variability in the structural layout of these individual enzymes (see Figures 4 and 5a, 5b). Rare examples of gene clusters which include the Y-junction module but not the flavin module were excluded from the present analysis which focuses on the NAD(P)-dependent enzymes. These rare cases likely correspond to so far unknown enzymes.

Phylogenies of both the Y-junction and the flavin modules were reconstructed and are shown in Figures 5a and 5b, respectively. Sequences of the Y-junction module are contiguous with no significant indels (the only exception shall be discussed in section 3.5.6) and contain amino acid residues ligating between three and four iron-sulphur clusters. The 12-16 conserved residues are evenly distributed over the entire sequence and therefore substantially facilitate the alignment. The sequence stretch considered for tree reconstruction amounts to 200 amino acid residues on average (alignment of the Y-junction module can be found in Supplementary Figure 1). In addition to these functionally conserved residues, sequences of the Y-junction module exhibit several stretches of high conservation. The flavin module, by contrast, is frequently split into two to three subunits requiring concatenation for tree reconstruction. We restricted the concatenation procedure to the two ubiquitously present domains as discussed above and highlighted in Figure 3 (section 3.1.3 and Supplementary Fig. 2). The concatenated sequence of the flavin module contains on the order of on average 510 amino acid residues. Neither the Y-junction nor the flavin module therefore correspond to very large molecules and the number of phylogenetically informative sequence sites therefore is limited. This resulted in low bootstrap values (visualized by grey shaded branches in the phylogenetic trees in Fig. 5a and 5b) for some of the lowest nodes connecting the individual enzyme families. These low bootstrap values preclude firm inferences on the sequence of evolutionary events during emergence of the entire superfamily of these NAD(P)-dependent enzymes. However, individual clades were strongly bootstrap-supported.

**Fig. 5A Tree Y-junction module and Fig. 5B Tree flavin subunit here, (whole page each)**

The phylogenetic tree of the Y-junction module shown in Fig. 5a clearly distinguishes the four groups of enzymes discussed above, that is, the NAD(P)-dependent subfamilies of [FeFe]-hydrogenases (on a dark blue background shading), [NiFe]-hydrogenases (cyan background), formate dehydrogenases (purple background) and complexes I (red background). The molecular layout of the enzymes and the presence or absence of specific redox cofactors were deduced from the organisation of the observed gene clusters and the presence/absence of conserved motifs (as shown in Supplementary Table 1). The conserved general outline of the gene clusters in the various families is schematically presented in Fig. 4.

The schematic pictograms drawn within the boxes of Fig. 5a indicate the respective minimal cofactor composition for each of the four main enzyme families. When individual clusters within a family subtree share common setups with additional redox centres, these arrangements are shown in smaller pictograms drawn at the representative leaves of the tree in Fig. 5a and 5b. For branches which don't feature pictograms, their specific additional cofactor complement varies too much to be explicitly displayed without overloading the figure. Fully-annotated trees can be found in the Supplementary Figures 4a and 4b, for Fig. 5a and 5b, respectively. The exact layout of each enzyme considered in this study is specified in the Supplementary Table 3.

### 3.5.1. Did the Y-junction- and the flavin-module coevolve?

A detailed inspection of the well-defined clades in Figures 5a and 5b shows that the phylogenies reconstructed from the Y-junction- and the flavin module feature very similar clusterings. However, the general layout of these trees, that is, the ordering of low-branching nodes, differs substantially between Fig. 5a and Fig. 5b. This may in part be due to poor reliability of the deep branchings mentioned above. By contrast, several discrepancies between the two trees cannot be rationalised by low bootstrap support. The tree of the flavin module in fact shows a pronounced and strongly bootstrap-supported (99.6%) cleavage into two subtrees which notably tears apart the two individual clades of [NiFe]-hydrogenases (denoted as Group I and II, respectively, see section 3.5.3) as well as numerous [FeFe]-hydrogenases and formate dehydrogenases, all of which are grouped together in the tree of the Y-junction module.

A second indisputable discrepancy consists in the fact that in the tree of the flavin module, formate dehydrogenases are interspersed within [FeFe]-hydrogenases and *vice versa*. Even in the absence of any tree reconstruction, the multiple alignments by themselves already show that the respective sequences are much more closely related to those of neighbouring branches corresponding to dissimilar enzymes than to those of their genuine family.

We consider that the following observations show a lead to a possible and quite parsimonious rationalisation of the above described discrepancies: (a) The Y-junction module almost perfectly clusters individual enzyme families. (b) The Y-junction module is fused to the catalytic domain in all representatives of NAD(P)-dependent [FeFe]-hydrogenases and formate dehydrogenases hindering shuffling of genes encoding these two protein domains between different gene clusters in the same species or even inter-species. (c) As already touched upon in section 3.3., from an electrochemical point of view the CO<sub>2</sub>/formate and the H<sup>+</sup>/H<sub>2</sub> couples are basically indistinguishable. Interchanging flavin modules between these two families therefore does not require redox adaptations and resulting hybrids would be instantly fully functional. (d) The split in the flavin tree of Figure 5b reflects a dichotomy which has recently been proposed by Losey *et al.* (2020) to be linked to the ability/disability of [FeFe]-hydrogenases and formate dehydrogenases to exchange electrons derived from the flavin-based bifurcation/confurcation reactions with the pool of low potential ferredoxins/flavodoxins as defined in section 3.4.2 [24].

The ensemble of these findings suggest the following scenario: The Y-junction module almost strictly coevolved with the catalytic domains of the respective enzyme families rationalising the at first sight puzzling observation that a tree based solely on sequences of the Y-junction module manages to reproduce the grouping of the diverse enzyme families. The flavin module, by contrast, can be interchanged between (particularly [FeFe]-hydrogenase and formate dehydrogenase) families which explains the occurrence of hybrid enzymes. The equivalence of the formate and hydrogen redox reactions furthermore may rationalise the emergence of more exotic cases such as that characterised in *Clostridium autoethanogenum* [14, 48] which merges [FeFe]-hydrogenase, formate dehydrogenase, the Y-junction and the flavin module into a super-enzyme (see Supplementary Section 3). Finally, if the proposal by Losey *et al.* (2020) [24] that the major dichotomy of the flavin module tree corresponds to its ability/disability to reduce low potential ferredoxins/flavodoxins (see sec-

tion 3.4.2), should weather the challenges of an increasing number of characterised enzymes, this would point towards very early divergence of the flavin module into two groups involving or excluding ferredoxin/flavodoxin redox reactions in/from the bifurcation/confurcation process. According to the metabolic needs of each organism it may then have plugged either of these functionally different modules onto their respective enzymes.

The molecular layout of the two clades (denoted as Group I and Group II) of [NiFe]-hydrogenases appears to be in line with this hypothesis. While these two clades appear as close sisters in the tree of the Y-junction module, they are pulled apart into the distinct regions of the flavin tree. The 3D-structure of the characterised enzyme (Fig. 3B) as well as the predicted molecular makeup of the other representatives in this clade (Group I in Fig. 5a) show that the flavin module lacks the [2Fe-2S] centre, a redox cofactor which most likely is indispensable for electron transfer between the flavin and soluble ferredoxins/flavodoxins. Characterised [NiFe]-hydrogenases from this clade do not exchange electrons with the low potential soluble carriers. In line with these structural considerations, these enzymes are found within the ferredoxin-independent part (according to [24]) of the tree. The representatives of the Group II clade (located within the presumed ferredoxin/flavodoxin-dependent part of the flavin tree), by contrast, have the full complement of redox centres also found in ferredoxin-dependent [FeFe]-hydrogenases and formate dehydrogenases at their disposition. Biochemically studied cases from Group II have been shown to exchange electrons both with NAD(P)<sup>+</sup>/NAD(P)H and ferredoxins or flavodoxins, although a potential coupling of these two reactions could not be conclusively assessed so far [80, 81]. We will present more speculative assessments of the very early evolutionary pathways of the enzyme families towards the end of this article. In the following we will first discuss both the fine structure of the tree and the pertinence of the observed structural and functional idiosyncrasies.

### 3.5.2. NAD(P)-dependent [FeFe]-hydrogenases

#### 3.5.2.1. Phylogenetic distribution and molecular layout

As suggested by the tree of the Y-junction module in Fig. 5a, the NAD(P)-dependent [FeFe]-hydrogenases form a compact clade with several deep-branching subclades. The gene cluster compositions of the enzymes representing this clade strongly resemble each other (Fig. 4) to the exception of the subclade clustering around the *C. pasteurianum* enzyme (see 3.5.2.2.). The clade of NAD(P)-dependent [FeFe]-hydrogenases comprise sequences encoded by genes which are annotated in their respective genomes as either “hnd”, “hyd”, “hyt” or “hym”. These dissimilar naming schemes do not coincide with phylogenetic clustering but are purely historical. The alphabetical numbering of subunits, despite substantial homologies and strongly conserved succession in gene clusters (Fig. 4), furthermore differs between the Hnd-, Hyd- etc naming schemes. To avoid confusion when discussing the molecular makeup of enzymes in this clade, we have adopted the Hnd-type nomenclature throughout and provide in Supplementary Table 2, the correlation between all these naming schemes. The introduction of the term Hnd was originally motivated by the fact that these hydrogenases are nad(p)-de-dependent and, in the light of the fact that all cases treated in this work contain the flavin module, we consider this nomenclature as the most appropriate.

So far in prokaryotes, Hnd-type enzymes are exclusively found in Bacteria. They share this trait with their NAD(P)-independent relatives further corroborating the picture that [FeFe]-hydrogenases are absent from the archaeal domain [82].

Hnd-type enzymes generally occur as tetrameric, trimeric or in rare occasions as dimeric versions. In the ciliate *N. ovalis*, a eukaryote, even a monomeric Hnd has been found [83, 84]. It has been proposed previously that the small HndB subunit of the tetrameric form corresponds to an N-terminal domain of the flavin containing HndC subunit in trimeric manifestations of the enzyme and that therefore the molecular layouts of tri- and tetrameric versions are virtually the same [12, 16]. Mono- and dimeric forms represent more extensive fusions of subunits [24]. Our genome survey corroborates this view since in all trimeric cases we retrieved, the N-terminal domain of the HndC subunit, when it exists, shows intriguing sequence similarities with the HndB proteins (Such homology is illustrated by same color and shade in Figure 4). HndB, just as the [2Fe-2S]-domain of HndA, have been predicted to feature the fold of Trx-like [2Fe-2S]-ferredoxins as discussed in section 3.4.1.

We furthermore found four trimeric sequences (also discussed in section 3.4.1), with a domain homologous to HndB fused to the C-terminal end of the catalytic [FeFe]-hydrogenase subunit HndD (see Fig. 4). A clustering of these four cases is observed both in the Y-junction- and the flavin-module-tree. No clustering into tri- and tetrameric clades can be observed on these trees (in accordance with [16]). While the lower nodes on the tree aren't supported by high bootstrap values, higher level branchings come close to 100% bootstrap support and in several cases such nodes separate tri- from tetrameric representatives. We therefore conclude that the gene fusion/fission of the DNA-stretch coding for the HndB subunit has occurred several times independently during [FeFe]-hydrogenases' evolutionary history from their emergence to the present day.

### **3.5.2.2. The *C. pasteurianum*-type monomeric, NAD(P)-independent enzymes possibly derive from NAD(P)-dependent hydrogenase ancestors**

As mentioned above we restricted this analysis to enzymes that contain both the Y-junction and the flavine module. As the only exception to this rule we included the monomeric enzymes resembling the *C. pasteurianum* [FeFe]-hydrogenase (denoted by a dark-grey background within the tree of [FeFe]-Hydrogenase in Fig. 5a). Since this entire work was initially triggered by the observation of a structural module conserved between the *C. pasteurianum* (*i.e.* NAD(P)-independent!) [FeFe]-hydrogenase and the three NAD(P)-dependent enzymes shown in Fig. 1, we found it indispensable to include the *C. pasteurianum* sequence into our phylogenetic analysis (as it also guided the structure model informed alignments, Supplementary Figure 1). To increase the reliability of nodes, we searched genomes for further cases related to the *C. pasteurianum* enzyme, that is, featuring the Y-junction module as N-terminal (F-) domain with a [FeFe]-catalytic (H-) domain but lacking genes coding for the flavin module. The resulting tree topology draws an intriguing evolutionary picture. All *C. pasteurianum*-type enzymes cluster together and emerge as a subclade from within the tree of NAD(P)-dependent [FeFe]-hydrogenase (Fig. 5a). The most parsimonious interpretation of this topology is that *C. pasteurianum*-type enzymes

derive from NAD(P)-dependent precursors which still contained the flavin module. The Y-shaped form of the electron entry/exit pathways to and from the catalytic H-cluster thus would find its *raison d'être* as an evolutionary relic from an NAD(P)-dependent past. This of course does not necessarily mean that the two distinct electron transfer pathways aren't physiologically relevant in the extant *C. pasteurianum* enzyme and its relatives [32]. However, rather than having specifically emerged to perform such physiological functions, they derive from a different kind of process, that is, branched electron transfer as in the Hnd-type enzymes, and have been attributed novel functions when the flavin module got lost. The low bootstrap support at the base of this clade cannot strictly rule out alternative scenarios which would have in common that they consider an [FeFe]-hydrogenase devoid of the flavin module as the ancestor of the Hnd family. However, such scenarios would entail the necessity for extensive lateral transfer of both the flavin and the Y-junction module across all families rendering them significantly less parsimonious than the above proposed one.

### 3.5.3. NAD(P)-dependent [NiFe]-hydrogenases

As mentioned above, this family features a deep branching into two clades (Group I and Group II in Fig. 5a and 5b). The first group shows a further, more recent divergence into two subclusters of sequences. The structurally and biochemically known enzyme from *H. thermoluteolus* (Fig. 1G-I [6, 85]) and the biochemically best characterised complex from *Cupriavidus necator* (formerly called *Alcaligenes eutrophus*) belong to Group I and represent relatively typical members of their particular subclade [86, 21, 22]. They contain only three iron-sulphur clusters in the Y-junction module (missing the C-terminal cluster as indicated in Fig. 1G and 1H) and furthermore lack the [2Fe-2S]-centre from the pair of clusters flanking the flavin (Fig. 3). This implies that (a) the Y-junction module cannot mediate branching into two distinct electron transfer pathways and (b) the reduction of the flavin cannot occur in a confurcating 2-electron process but the electrons must reach the flavin one-by-one on their way from the H<sub>2</sub>-oxidising catalytic site of the [NiFe]-hydrogenase or *vice versa*. Electron transfer in the enzymes from this clade therefore necessarily is linear as indicated in Fig. 1G. Consequently, we predict that in these enzymes the flavin features a relatively high stability constant for its semi-reduced form to facilitate one-by-one reduction [62].

By contrast, the second clade forms a single radiation (Group II in Fig. 5a and 5b). All members of Group II not only seem to contain the [2Fe-2S]-cluster next to the flavin but their Y-junction modules contain the full set of all four iron-sulphur centres and their Y-junction/flavin-module dyad therefore strongly resembles that of ferredoxin/flavodoxin-dependent and NAD(P)-dependent [FeFe]-hydrogenases and formate dehydrogenase. It hence is tempting to assume corresponding functional properties. Experimental evidences gathered on the cyanobacterium *Synechocystis* sp. *PCC 6803* give credence to this assumption. *In vitro* tests in cell-free extracts have shown a significant increase in hydrogen production upon addition of ferredoxin or flavodoxin, indicating a potential second donor/acceptor [80]. These experimental results may indicate a less-than-strict coupling of NAD(P) and ferredoxin redox reactions as also observed in the Hnd-enzyme from *D. fructosovorans* [8]. While we would suggest a bifurcating ability in these enzymes, to our

knowledge a coupling of the reductions of ferredoxin and of  $\text{NAD}^+$  by electrons derived from hydrogen has not been unambiguously demonstrated so far (however, see Artz *et al.* (2020) [87]).

We were unable to detect any archaeal representatives of this enzyme family and the detailed structure of its subtree shows only little resemblance to species trees. At the present state of available genome data, it therefore seems likely that this enzyme emerged in Bacteria and was subsequently distributed largely *via* horizontal gene transfer over the bacterial domain. As a caveat to these conclusions, it is worth mentioning that heterotetrameric, soluble NADP-dependent [NiFe]-hydrogenases can be found in the euryarchaeal family of Thermococcaceae with characterised proteins from *Thermococcus litoralis* [88], *Thermococcus kodakarensis* [89] and *Pyrococcus furiosus* [90]. Besides a certain degree of similarity in the catalytic hydrogenase subunits, we were unable to identify a Y-junction module nor could we find similarities in the sequences of the respective flavin subunits (the flavin can occur either as FAD or as FMN in these cases). While they have been called “Hox-type” hydrogenases by some authors [91], we have not considered them here since they are extremely divergent with respect to those dealt with in our phylogenetic analysis. Whether archaeal enzymes of this type will ultimately modify the tentative conclusions on the ancestry of this clade remains to be seen.

#### 3.5.4. NAD(P)-dependent formate dehydrogenases

A few isolated cases of this type of enzyme have been dealt with in the literature [11, 41, 49, 50, 92, 93, 7, 51, 47] and phylogenetic investigations are scarce [94]. In the phylogenetic tree reconstructed from sequences of the Y-junction module (Fig. 5a), all cases containing a formate-dehydrogenase of the Mo/W-*bis*PGD-type cluster into five groups, well-separated from the NAD(P)-dependent versions of both [FeFe]- and [NiFe]-hydrogenases. Three of these groupings are also seen in the tree based on flavin module sequences (Fig. 5b).

The common presence of the flavin module in all formate dehydrogenases in the tree implies that all these enzymes are indeed NAD(P)-dependent and belong to a homogeneous group. Similar to what we discussed in section 3.5.2 for [FeFe]-Hydrogenases, several naming schemes are in use for this enzyme family (*e.g.* Fds [41, 95], Fdh [92, 96, 94] or Fno [97]). To avoid confusion, we will apply the “Fds”-scheme to all these enzymes and specify nomenclature correspondences with Table 1 and Supplementary Table 2.

In the depicted phylogeny of the Y-junction module, the subtree of NAD-dependent complexes I grows out of one of the clades of the NAD(P)-dependent formate dehydrogenases. This intriguing observation will be discussed at length in the following section (section 3.5.5) dealing with the complex I subtree.

In striking contrast to NAD-dependent [FeFe]- and [NiFe]-hydrogenases, a few clusters of NAD(P)-dependent formate dehydrogenase-type enzymes show deep splittings into archaeal and bacterial subgroups. Group A3 in Fig. 5a, for example, is characterised by a deep (Eury-)Archaea/Bacteria divergence but unfortunately this clade so far contains only two representatives for each prokaryotic domain. In enzymes from this clade the Y-junction module retains all four clusters and the flavin module is equipped with both clusters flanking the flavin. All representatives are trimeric, *i.e.* their small subunit is fused to the flavin-binding

FdsB subunit. Since the complement of redox cofactors is equivalent to that of Hnd- and Fds-type enzymes for which ferredoxin reduction *via* electron bifurcation (see sections 3.2 and 3.3) has been experimentally observed, it seems tempting to assume that the enzymes in Group A3 are able to fulfill similar functions. However, while the flavin modules of the two archaeal representatives indeed cluster within the “ferredoxin/flavodoxin-dependent” domain (see section 3.5.1) of the tree shown in Fig. 5b, the two bacterial sequences are part of the lower, ferredoxin-independent subtree and biochemical information on the respective enzymes will be required to settle this question.

Enzymes from a further archaeal cluster (denoted as Group A1 in Fig. 5a) differ significantly from the above mentioned euryarchaeal examples. This group encompasses both Cren- and Euryarchaea as well as two additional bacterial species. They all lack the histidine cluster in the Y-junction module and most of them don’t possess a flavin-flanking [2Fe-2S]-cluster. Only the enzymes from the *Aciduliprofundum* genus retain this cluster. Again to the exception of these two euryarchaeal cases, all members of the group are dimeric enzymes with one polypeptide encompassing the entire flavin module. The absence of two crucial [Fe-S]-clusters makes the possibility of a communication with soluble ferredoxins in these enzymes doubtful. Two proteobacterial representatives, *Methylobacterium extorquens* (Met ex) and *Cupriavidus oxalaticus* (Cup ox), are associated with this archaeal clade (see Supplementary Fig. 3a). Notably, these two bacterial sequences do not cluster with archaeal ones in the tree of the flavin module. Their Y-junction module lacks the histidine-cluster, a feature not found in other bacterial NAD(P)-dependent formate dehydrogenases but characteristic for the mentioned archaeal clade. The low bootstrap support for the nodes of these two bacterial sequences favours the assumption that pulling them into the archaeal clade is driven by the common absence of the residues ligating the histidine cluster and therefore a tree-building artefact. This effect likely also explains the positioning of the sequence of the dimeric enzyme from the Archaeon *Methanothermobacter methylutens* (Met me) basal to the bacterial cluster Group A2 next to the archaeal Group A1. This sequence contains the histidine-ligating residues (just as all bacterial members of this clade) which likely drags its branch over from the archaeal clade. In the tree of the flavin module its branch is indeed part of the corresponding archaeal clade. We therefore consider it highly likely that the groups A1 and A2 in Fig. 5a in fact correspond to well-separated archaeal and bacterial clades, respectively. Potential evolutionary consequences of such a clustering will be discussed below.

Members of Group A2 contain both tri- and tetrameric manifestations of the enzyme in a strongly mingled manner. However, the 3D structure model of the A2-type enzyme from *Rhodobacter capsulatus* (Rho ca) [7] indicates that, in stark contrast to NAD(P)-dependent [FeFe]-hydrogenases, the fourth subunit (denoted FdsD, see Fig. 4) in the tetrameric versions (a) does not resemble thioredoxin-like [2Fe-2S]-proteins (see section 3.4.1) and (b) is located on the surface of the catalytic molybdopterin-subunit. Like the enzymes from Group A3, the members of this group feature the full set of cofactors required to bifurcate electrons and to connect to the ferredoxin pool. In fact, for the tetrameric enzyme from *Methylobacterium trichosporium* (Met tr) the absence of an observable flavin semiquinone species [49], indicative for inverted redox potentials, supports this scenario. Further biochemically char-

acterised members of Group A2 are the Fds enzymes from *Rhodobacter capsulatus* (Rho ca) [50] and *Cupriavidus (C.) necator* (Cup ne) [41] for which the structural models of the entire protein [7] or the flavin module [47] have been recently obtained. In addition, the latter study regarding the enzyme of *C. necator* observed a neutral semiquinone species by EPR, thereby suggesting a ferredoxin-independence of the reaction. Such a ferredoxin-independent reduction of CO<sub>2</sub> by NADH could also be shown for its relative *C. oxalaticus* (Cup ox) [98].

The cluster denoted as Group B again contains tri- and tetrameric enzymes exclusively from the bacterial kingdom. In contrast to the above mentioned Group A2, the fourth subunit in the tetrameric versions in this group is a homolog of the HndB protein of [FeFe]-hydrogenases. Since the name FdsB is already taken by the flavin/[4Fe-4S]-subunit of the flavin module, we have chosen to denote the gene coding for this subunit as *fdsX* in Fig. 4.

We have organised Group B into a larger group B2 encompassing a second enzymes cluster denoted as group B1, which homogeneously display the full complement of redox centres involved in electron bifurcation and in branched electron transfer. Members of the enclosing Group B2, by contrast, show more heterogeneous layouts of subunits and cofactors (specific information for each individual representative can be found in the Supplementary Table 3). Group B2 contains the only Fds-enzyme for which electron bifurcation involving reduction of ferredoxins has been demonstrated experimentally, *i.e.* the Fds from *Gottschalkia acidurici* [11]. Unfortunately, the gene cluster organisation of this very case deviates somewhat from that of the bulk of cases situated in this group. Its gene cluster features a Y-junction module (with the FdsA specific A1 cluster, see Tab. 1) as separate ORF together with two distinct catalytic formate dehydrogenase subunits, each of which contains its own Y-junction module as an N-terminal extension. The two latter Y-junction module sequences cluster together (not shown since difficult to discern on the scale of the tree), indicating relatively recent gene duplication events. However, the Y-junction module encoded by a separate gene located in between the duplicated catalytic subunits and the flavin module, is of yet indetermined origin (and not included in the trees). A traditional flavin module is also present. While the global conformation of all other enzymes can be reasonably well predicted based on the 3D-structure models shown in Fig. 1, imagining the spatial layout of this specific enzyme appears more challenging to us. For the more abundant cases, however, a similar cofactor arrangement as for Hnd-type enzymes is likely as already proposed by Hille *et al.* (2014) [95].

The representatives making up Group B1 in Fig. 5a are found to also cluster together in the tree of the flavin module (Fig. 5b) and more precisely in the subtree suggested to represent ferredoxin/ flavodoxin-dependent cases [24]. By contrast, members of Group B2 as defined by the tree based on the Y-junction module do not cluster in the flavin module tree but interestingly are interspersed either within clades representing Hnd-type enzymes or in between Hnds, [NiFe]-hydrogenases and “glutamate-synthase-insertion” (see below, section 3.5.6) clades. Nevertheless, all these latter disperse branches (to the sole exception of *Desulfobacca acetoxidans*) are still part of the “upper” subtree, possibly indicating ferredoxin/ flavodoxin-dependent enzymes in line with the biochemical characterisation of the *G. acidurici* enzyme [11]. This indicates that their flavin modules do not derive via

vertical inheritance from their counterparts in the ancestral enzymes but rather from genes transferred from diverse sources into the respective operons to replace the original flavin modules of these ancestors.

These enzyme therefore all represent “reassembled” forms merging genuine “Y-junction-module/formate-dehydrogenase” subunits with flavin modules from other enzyme systems as already mentioned in section 3.5.1. All enzymes but one from both Group B1 and B2 share the trait of belonging to the putatively ferredoxin/flavodoxin-dependent [24] subtree of Fig. 5b.

A small number of cases do not follow the gene cluster scheme for NAD(P)-dependent formate dehydrogenases as outlined in Fig. 4. These cases are further discussed in the Supplementary section 3.

### 3.5.5. NAD-dependent complex I

We again emphasise that we here deal only with a subgroup of what is frequently referred to as complex I, *i.e.* those complexes I which contain the NuoEFG module and thus certainly redox-convert NAD during their catalytic cycle (as discussed in Suppl. section 1). More specifically, the evolutionary conclusions we will draw below only pertain to this EFG-module and do not inform on the evolutionary history of the remaining modules of the enzyme, that is, the quinone-module and the actual ion-pumping subunits.

Rather than appearing as a separate family, the phylogenetic tree shown in Fig. 5a features complex I’s Y-junction module as emanating from within the B1-Group of NAD(P)-dependent formate dehydrogenases. While the bootstrap values supporting this topology are respectable (48%), they are certainly far from suggesting infallible reliability. However, a range of structural features which can be deduced from the amino acid sequences of the relevant protein subunits both of true NAD(P)-dependent formate dehydrogenase- and complex I-gene clusters corroborate a gradual transitioning from an NAD(P)-dependent formate dehydrogenase enzyme towards the NuoEFG module of complex I.

Members of Group B of the Fds therefore show the following features: (1) The presence of the Mo/W-*bis*Pterin centre which catalyses the CO<sub>2</sub>/ formate redox conversion. The distinguishing amino acid sequence motif for the presence of this cofactor is the conservation of a specific cysteine/selenocysteine residue in addition to a conserved histidine and an arginine which together form a catalytic triad in these enzymes [19]. (2) A Y-junction module which contains all four iron-sulphur clusters. (3) A flavin module which possesses the FMN binding motif as well as conserved cysteines for the two iron-sulphur centres flanking the flavin and which is located within the subtree of Fig. 5b which putatively corresponds to ferredoxin/flavodoxin-dependent enzymes. By contrast, none of the further subunits characterising complex I are found in their respective gene clusters to the exception of *Pelobacter propionicus*, which is discussed in the Supplementary section 3. The set of enzymes forming early branching clusters of genuine complex I still retain all iron-sulphur centres and the flavin but they additionally feature the complex I-specific membrane subunits as well as the NuoB/D genes in their gene clusters.

Intriguingly, the lowest-branching complex I cluster (made up from two *Geobacter* sequences in Fig. 5a) furthermore maintains a conserved cysteine at the sequence position

where the 5th Mo/W-ligand is found in formate dehydrogenases (all later-branching complex I enzymes in our tree lack this crucial cysteine/selenocysteine residue). However, the sequence stretch surrounding these cysteines diverges very strongly from those observed in all NAD(P)-dependent formate dehydrogenases.

The flavin-modules of these two enzymes cluster in the ferredoxin/flavodoxin-dependent subtree neighboring those of the B-type formate dehydrogenases. Sequence information thus suggests the tantalising possibility that these two complex I-type enzymes may still contain the Mo/W-*bis*Pterin cofactor and possibly even CO<sub>2</sub>/formate redox activity as well as involve ferredoxin/flavodoxin redox reactions in their catalytic processes. Biochemical characterisations of these enzymes will be required to assess the validity of the respective sequence-based functional predictions. The cluster formed by the two *Geobacter* sequences in Fig. 5b furthermore encompasses a complex I-type enzyme from the Planctomycetes *Kuenenia stuttgartensis*. The bootstrap value of the node for this enzyme's Y-junction module (Fig. 5a) is too low to allow conclusions on positioning within the evolutionary history of complex I.

All complex I representatives branching higher than the two *Geobacter* sequences in Fig. 5a feature flavin modules which are located in the lower subtree of Fig. 5b, possibly indicating ferredoxin/flavodoxin-independence, as would indeed be expected based on functional characterisations of canonical complex I enzymes. This may imply that the ancestor of all these higher-branching enzymes has replaced its originally ferredoxin-dependent flavin-module by a ferredoxin-independent one, a process which likely has occurred in many NAD(P)-dependent [FeFe]-hydrogenases and formate dehydrogenases (see section 3.5.4).

When moving up the tree towards more and more recent branchings, a progressive loss of the iron-sulphur centres not involved in the electron transfer chain connecting the flavin module to the quinone binding subunit is observed. Over the evolutionary journey from its inception out of an NAD(P)-dependent formate dehydrogenase up until its form in the mitochondrial complex I, the unit formed by the flavin module and the Y-junction module, obviously has successively lost all the dispensable iron-sulphur centres and individual transition states are still with us as living fossils in the species indicated on the tree of Fig. 5a [33]. In the light of this evolutionary offloading, it is intriguing that the [2Fe-2S] centre flanking the flavin, *alias* the N1a cluster, is still present even in the mitochondrial enzyme. The example of the NAD-reducing [NiFe]-hydrogenase of *H. thermoluteolus* (see section 3.4.1) demonstrates that redox conversions of NAD can be mediated by a flavin module lacking this cluster [85]. A scenario assuming that cluster N1a may be on its way out and disappear in the evolutionary future similar to the other out-of-line clusters, however, seems highly questionable to us. Site-directed mutations eliminating this cluster were seen to entail significant loss of NADH-oxidase activity in *E. coli* [59]. In humans, absence of the subunit harbouring cluster N1a was observed to be associated with debilitating conditions such as Alzheimer and Parkinson disease [99]. As outlined above, we suspect that kinetic optimization of the 2-electron oxidation of the flavin in complex I requires the presence of both flanking clusters.

## Repercussions on the debate of complex I's evolutionary origins

A plethora of scenarios attempting to retrace the evolutionary whereabouts of complex I have been proposed [100, 101, 35, 102, 5, 103, 104, 105, 106]. While they all agree that complex I was assembled from a range of functional units with differing enzyme ancestors, the specific evolutionary roots of particularly the membrane-extrinsic parts, that is, the NAD- and the quinone-redox-converting entities, are controversial. While some models proposed [FeFe]-hydrogenases or formate dehydrogenases as precursors of the NAD-redox converting module [106], others involve NAD-dependent [NiFe]-hydrogenases and formate hydrogenlyase (FHL) as ancestors of both the NAD- and the quinone-binding modules [5, 107]. FHL is indeed an intriguing enzyme since it features the presence of a Mo/W-*bis*Pterin subunit together with the two [NiFe]-hydrogenase subunits known to share substantial sequence similarity with NuoB and NuoD, the subunits involved in redox conversion of the quinone in complex I. To top up these similarities, one or more of the proton-pumping membrane subunits are also found to be part of FHL. However, the formate dehydrogenase subunit in FHL is a standard Mo/W-*bis*Pterin protein containing only the canonical adjacent [4Fe-4S] cluster but lacking the N-terminal Y-junction module domain. The small iron-sulphur protein (HycB) of FHLs proposed as homolog of the Y-junction module clearly has a different architecture since no significant conservation of cluster-ligating residues is observed [5]. The flavin module is also absent in FHL. Despite the intriguing overall structural similarities of FHL and complex I, the devil is in the above mentioned details which differ substantially between these two enzymes. Claims of direct evolutionary relationships between these two [108] may therefore have been premature.

The second oft-discussed candidate [101, 102, 104], the NAD(P)-dependent [NiFe]-hydrogenase, by contrast, contains the Y-junction module, the flavin module and the NuoB and NuoD homologs. It misses the Mo/W-subunit and the membrane subunits together with a few further small subunits. However, the tree of Fig. 5a suggests that the NAD(P)-dependent [NiFe]-hydrogenase is not the phylogenetically closest relative of the respective proteins in NAD-dependent complex I. Instead, the so far only relatively scantily studied NAD(P)-dependent formate dehydrogenase is! This enzyme family featured as the poor cousin among all the evolutionary precursor candidates of NAD-dependent complex I until the characterisation of the *G. acidurici* enzyme by the Thauer group [11] managed to appreciate the catalytic feats performed by this enzyme. The tree of Fig. 5a thus indicates that the NuoEFG module in NAD-dependent complex I directly derives from an NAD(P)-dependent formate dehydrogenase. The ultimate ancestor to complex I's EFG-module may have been the putative ancestral ferredoxin/ flavodoxin-dependent version (section 3.5.4) with a subsequent exchange of the type of flavin-module leading to canonical reaction scheme of complex I. During the evolutionary conversion from a Group B formate dehydrogenase to the NAD-redox converting unit of complex I, the histidine cluster of the Y-junction module saw its function of either connecting the two halves of a dimeric assembly or of mediating electron transfer to or from a ferredoxin in putative monomeric forms of NAD(P)-dependent formate dehydrogenases morphed into being part of the chain of redox cofactors leading “down” towards the quinone binding site of the ancestral, NAD-independent form of complex I. The CO<sub>2</sub>/formate converting function of the Mo/W-*bis*Pterin subunit was shut down, initially *via* elimination of the protein ligands to the molybdenum/tungsten and then step-by-step *via* the taking out of

further iron-sulphur centres connecting the central cluster of the Y-junction module to the Mo/W-*bis*Pterin subunit. While our data have no bearing on the question of how EFG-less forms of complex I may have arisen, they very strongly indicate that it was a previously overlooked subgroup of NAD(P)-dependent formate dehydrogenases which gave birth to the membrane-extrinsic part of the NADH-oxidising version found in many respiratory chains and eventually in mitochondria.

Astonishingly, the three rhizobial representatives show molecular architectures that deviate from that of other Alphaproteobacteria since they retain both all four clusters in the Y-junction module and the FS0 cluster of the formate dehydrogenase-derived domain similar to what is found in the much lower branching enzymes from the Aquificales. They furthermore branch substantially earlier than Alphaproteobacteria in the complex I clade of both the Y-junction module and the flavin module trees. It therefore seems likely that these Rhizobia acquired their complex I *via* lateral gene transfer.

### 3.5.6. The “glutamate-synthase-insertion”-clade

In 2008, Toth *et al.* [20] reported the characterisation of an enzyme from the Euryarchaeon *Thermococcus litoralis* which features puzzling similarities to Hnd-type enzymes but which stands out by the presence of a long insertion showing substantial sequence similarities to the small subunit of the enzyme glutamate synthase, within the Y-junction module sequence. Toth *et al.* determined that this enzyme might represent an NAD(P)-dependent polysulphide reductase. Sequences related to this enzyme popped up during our genome surveys and are denoted in the trees of Fig. 5a and 5b by a beige area. They cluster together and apart from the rest of the tree. Intriguingly, these sequences show a deep Archaea/Bacteria split and therefore potentially diverged from the bulk of other families already in the last universal common ancestor (LUCA). Together with the NAD(P)-dependent formate dehydrogenase enzymes, this clade therefore supports the notion that the functional dyad constituted from the Y-junction- and the flavin-module is evolutionarily ancient and represents a pre-LUCA entity. Enzymes belonging to the glutamate-synthase-insertion-clade certainly are tantalising and to our mind merit further scrutiny. Most recently, a NAD/NADP-transhydrogenase was characterised from the acetogenic Clostridium *Sporomusa ovata* which features a Y-junction module containing this glutamate-synthase-related insertion sequence [79].

## 4. Conclusions

### 4.1. *The functional unit formed by the Y-junction-/flavin-module dyad likely pre-dates the last universal common ancestor*

As mentioned earlier, firm inferences on the evolutionary emergence of the Y-junction-/flavin-module dyad and its earliest diversifications are hampered by low bootstrap support for the deepest nodes of the trees in Fig. 5a and 5b. However, a few leads can be extracted from these phylogenies. At the time of writing, neither Hnd-type nor NAD(P)-dependent [NiFe]-hydrogenases appear to contain archaeal representatives and therefore do not qualify as potentially present in the last universal common ancestor (LUCA). By contrast, deep Archaea/Bacteria cleavages are observed for NAD(P)-dependent formate dehydrogenase as well as for the “insertion-type” enzymes. The Y-junction/flavin-unit may therefore have emerged already in the LUCA as a functional part of one of these two enzymes. Phylogenies of further representatives and/or of further subunits in their parent enzymes may help to progress in our understanding of the earliest events during spawning of the superfamily of enzymes shown in Fig. 5a and 5b. For example, the NAD(P)-dependent formate dehydrogenases are a subgroup of a much larger superfamily of formate dehydrogenases which do not contain the dyad, such as the enzymes denoted as Fdh-H, Fdh-N, Fwd etc. The composite phylogeny of their molybdopterin-binding catalytic subunits will provide a means of rooting of the NAD(P)-dependent subtree and thus will help to decide whether the NAD(P)-dependent enzymes indeed may have been present in LUCA.

### 4.2. *Lessons to be drawn from the comparative analysis of the enzymes containing the dyad*

#### 4.2.1. *The Y-junction module is a versatile 3-way redox hub with a poorly understood function*

The structural peculiarity of the Y-junction module to provide electron transfer connectivity between 3 distinct electron entry/exit sites on its surface is exploited by diverse enzymes for building a range of dissimilar architectures. Some of these enzymes appear to have shut down one of these pathways by eliminating either the histidine- or the C-terminal iron-sulphur cluster. Others, by contrast, seem to take advantage of the crossroad-functionality in still poorly understood ways. For example, an inter-dimer redox-connection *via* the two respective histidine clusters appears inevitable in the light of the oligomerization arrangement seen in the 3D-structure of an NAD(P)-dependent formate dehydrogenase. The mechanistic implications of such inter-dimer connections have not been explored so far. However, the scarcity of data with respect to oligomerization properties in the vast majority of enzymes considered in this work leaves room for speculations that in some enzymes the 3rd redox pathway may be used to connect soluble carriers to the main electron transfer chain from the main redox substrate to NAD(P). The precise functional role of the peculiar electron transfer crossroad in the Y-junction module thus clearly deserves further exploration.

#### 4.2.2. *The electron bifurcating site is located within the Y-junction-/flavin-module dyad*

Electron bifurcation-based reduction of low potential ferredoxins or flavodoxins was experimentally observed in NAD(P)-dependent [FeFe]-hydrogenases and NAD(P)-dependent

formate dehydrogenases [10, 13, 12, 15, 45, 8, 11]. Complex I was shown to bifurcate electrons in its corresponding functional unit (the EFG-module) albeit without relaying to low potential soluble acceptors [58]. As we have discussed in section 3.5.3, selected NAD(P)-dependent [NiFe]-hydrogenases also couple ferredoxin redox reactions to their canonical catalytic activities, although the strength of coupling has not been explored so far. The only element common to these 4 distinct enzyme families is the Y-junction-/flavin-module dyad. We therefore find it ineluctable that it is this structural unit which confers the ability to perform electron bifurcation. In none of the remaining, extremely diverse protein subunits has electron bifurcation ever been observed. As argued in section 3.4.2, the Y-junction module lacks the electrochemical properties conducive to electron bifurcation and we therefore favour the flavin module as the actual electron bifurcating entity.

#### *4.2.3. The strong conservation of the dyad between all families renders the presence of a second flavin in Hnd-type [FeFe]-hydrogenases unlikely*

The absence of a 3D-structure for Hnd-type enzymes together with conceptual difficulties to rationalise the observed reaction patterns using only a single flavin led to the proposal that Hnd-type enzymes actually contain two distinct flavins [36, 8], one dedicated to the bifurcation of electrons and the second one performing the 2x1-electron-to-2-electron gating function for NAD(P)H oxidoreduction (as for example observed for the flavin moiety located on the small subunit of Nfn) [36, 8]. For NAD-dependent complex I, NAD(P)-dependent formate dehydrogenase and NAD(P)-dependent [NiFe]-hydrogenase, 3D-structures are now available [33, 109, 7, 47, 6, 110]. In all these structures only one flavin is found in a strictly conserved location. The high similarity of the amino acid sequences in the subunits constituting the Y-junction-/flavin-module dyad between the diverse enzymes and in particular between formate dehydrogenase and Hnd-type [FeFe]-hydrogenase families to our minds leaves no leeway for the presence of a second flavin in the Hnd-type enzymes. The observed mode of functioning thus must almost certainly be explained on the basis of only a single flavin moiety. The difficulties to come up with a straightforward rationalisation using only one flavin precisely was the reason for postulating the two-flavin models and the comparative analysis of the diverse enzymes dealt with here unfortunately takes us back to the initial mechanistic riddle. Novel ideas for solving this enigmatic process are thus required.

#### *4.2.4. The electron bifurcating site occurs in a range of electrochemical layouts*

As discussed in this article, various enzymes from the described superfamilies behave in substantially different ways with respect to the electron bifurcating reaction. Some of them strongly couple the reduction of low potential ferredoxins or flavodoxins to the electron bifurcating processes. Others, such as for example complex I, perform the electron bifurcating reaction without involving these low potential soluble electron shuttle proteins. When characterised, the electrochemical properties of the flavin in these latter enzymes do not indicate strongly cooperative redox transitions as considered crucial for truly energy conserving electron bifurcation [72, 62, 55, 56, 73, 71]. Still others even manage to interface the 2-electron redox conversion of NAD(P) to the intramolecular 1-electron transfer chain without bifurcating electrons as evidenced by the absence of the [2Fe-2S]-cluster next to

the flavin. In these cases, the flavin must feature reasonably stable semireduced states as discussed previously [62]. The ensemble of these observations suggest that the flavin moiety in the flavin module present in the enzymes we dealt with in this article comes in electrochemically quite diverse forms. As suggested by simulations recently reported by Yuly *et al.* (2020) [73], such variability of cooperativity will result in variable strength of coupling between the exergonic and the endergonic 1-electron transfer steps. It will be interesting to use the above described cases to experimentally test the predictions that the ability to involve low potential electron donors/acceptors is correlated to the degree of redox cooperativity of the bifurcating 2-electron centre.

#### *4.3. Perspective: towards an understanding of the structural determinants which tune the flavin cofactor's redox cooperativity*

As discussed previously [62, 73], the precise type of redox reactions occurring within flavin-based electron bifurcation is likely correlated to the electrochemical properties of the flavin moiety and in particular to the strength of its redox cooperativity. These properties are necessarily pre-set by the flavin's immediate protein environment [111] but the type of influence the flavin's surroundings exerts on its electrochemical properties are only badly understood so far. As transpires from the previous sections, the reaction schemes performed by the flavin module vary substantially over the ensemble of enzymes discussed. So far, only a handful of high resolution 3D structure models are available. The set of cases for which comprehensive measurements of redox parameters of the involved cofactors and in particular of the flavin have been reported is even smaller. However, the possibility of confronting the close surroundings of flavins from enzymes cases with dissimilar coupling strengths in the electron bifurcating process would go a long way towards unraveling the molecular underpinnings of the electron bifurcating reaction. We therefore argue that the integration of the diverse enzymes into a common scheme, as attempted by our article, may allow valuable cross-fertilisations between so far rather distinct research communities and, in particular, may be key to understanding of what allows a 2-electron compound to act as an electron bifurcating agent. The development of a unified naming scheme for enzymes and subunits thereof, possibly arrived at *via* a community-wide concertation as for example achieved for arsenics-dependent molybdoenzymes [112] would certainly facilitate such cross-fertilisations.

## **Acknowledgements**

We thank Barbara Schoepp-Cothenet (BIP/CNRS Marseille/France) for help with the formate dehydrogenase family of enzymes and Wolfgang Buckel (MPI Marburg/Germany) for many stimulating discussions and critical reading of the manuscript. Thanks are also due to Kirsten Gutekunst (Christian-Albrechts-Universität zu Kiel /Germany) for providing the pdf-file of an article in a monograph. We are furthermore grateful to Jacques Meyer (Grenoble/France) for his willingness to extensively discuss and benefit from his expertise concerning ligation patterns of thioredoxin-like [2Fe-2S]-ferredoxins. The first author would also like to acknowledge the Studienstiftung des deutschen Volkes for its funding.

## References

- [1] F. Baymann, E. Lebrun, M. Brugna, B. Schoepp-Cothenet, M.-T. Giudici-Orticoni et al. (2003); ‘The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes’. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, vol. 358(1429):pp. 267–274.
- [2] Y. Higuchi, H. Ogata, K. Miki, N. Yasuoka & T. Yagi (1999); ‘Removal of the bridging ligand atom at the Ni–Fe active site of [NiFe] hydrogenase upon reduction with H<sub>2</sub>, as revealed by X-ray structure analysis at 1.4 Å resolution’. *Structure*, vol. 7(5):pp. 549–556.
- [3] J. W. Peters, W. N. Lanzilotta, B. J. Lemon & L. C. Seefeldt (1998); ‘X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom resolution’. *Science*, vol. 282(5395):pp. 1853–1858.
- [4] J. H. Artz, O. A. Zadvornyy, D. W. Mulder, S. M. Keable, A. E. Cohen et al. (2019); ‘Tuning catalytic bias of hydrogen gas producing hydrogenases’. *Journal of the American Chemical Society*, vol. 142(3):pp. 1227–1235.
- [5] R. G. Efremov & L. A. Sazanov (2012); ‘The coupling mechanism of respiratory complex I – a structural and evolutionary perspective’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1817(10):pp. 1785–1795.
- [6] Y. Shomura, M. Taketa, H. Nakashima, H. Tai, H. Nakagawa et al. (2017); ‘Structural basis of the redox switches in the NAD<sup>+</sup>-reducing soluble [NiFe]-hydrogenase’. *Science*, vol. 357(6354):pp. 928–932.
- [7] C. Radon, G. Mittelstädt, B. R. Duffus, J. Bürger, T. Hartmann et al. (2020); ‘Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in *Rhodobacter capsulatus* formate dehydrogenase’. *Nature communications*, vol. 11(1):pp. 1–9.
- [8] A. Kpebe, M. Benvenuti, C. Guendon, A. Rebai, V. Fernandez et al. (2018); ‘A new mechanistic model for an O<sub>2</sub>-protected electron-bifurcating hydrogenase, Hnd from *Desulfovibrio fructosovorans*’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1859(12):pp. 1302–1312.
- [9] S. Wang, H. Huang, J. Moll & R. K. Thauer (2010); ‘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*’. *Journal of Bacteriology*, vol. 192(19):pp. 5115–5123.
- [10] K. Schuchmann & V. Müller (2012); ‘A bacterial electron-bifurcating hydrogenase’. *Journal of Biological Chemistry*, vol. 287(37):pp. 31165–31171.
- [11] S. Wang, H. Huang, J. Kahnt & R. K. Thauer (2013); ‘*Clostridium acidurici* electron-bifurcating formate dehydrogenase’. *Appl. Environ. Microbiol.*, vol. 79(19):pp. 6176–6179.
- [12] Y. Zheng, J. Kahnt, I. H. Kwon, R. I. Mackie & R. K. Thauer (2014); ‘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’. *Journal of bacteriology*, vol. 196(22):pp. 3840–3852.
- [13] S. Wang, H. Huang, J. Kahnt & R. K. Thauer (2013); ‘A reversible electron-bifurcating ferredoxin-and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*’. *Journal of bacteriology*, vol. 195(6):pp. 1267–1275.
- [14] S. Wang, H. Huang, J. Kahnt, A. P. Mueller, M. Kopke et al. (2013); ‘NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO’. *Journal of Bacteriology*, vol. 195(19):pp. 4373–4386.
- [15] G. J. Schut & M. W. W. Adams (2009); ‘The Iron-Hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production’. *Journal of Bacteriology*, vol. 191(13):pp. 4451–4457.
- [16] S. Poudel, M. Tokmina-Lukaszewska, D. R. Colman, M. Refai, G. J. Schut et al. (2016); ‘Unification of [FeFe]-hydrogenases into three structural and functional groups’. *Biochimica et Biophysica Acta (BBA)-General Subjects*, vol. 1860(9):pp. 1910–1921.
- [17] S. Thiemermann, J. Dornedde, M. Bernhard, W. Schroeder, C. Massanz et al. (1996); ‘Carboxyl-terminal processing of the cytoplasmic NAD-reducing hydrogenase of *Alcaligenes eutrophus* requires the hoxW gene product.’ *Journal of bacteriology*, vol. 178(8):pp. 2368–2374.

- [18] M. Guiral, L. Prunetti, S. Lignon, R. Lebrun, D. Moinier et al. (2009); ‘New insights into the respiratory chains of the chemolithoautotrophic and hyperthermophilic bacterium *Aquifex aeolicus*’. *Journal of proteome research*, vol. 8(4):pp. 1717–1730.
- [19] T. Hartmann, P. Schrapers, T. Utesch, M. Nimtz, Y. Rippers et al. (2016); ‘The molybdenum active site of formate dehydrogenase is capable of catalyzing C–H bond cleavage and oxygen atom transfer reactions’. *Biochemistry*, vol. 55(16):pp. 2381–2389.
- [20] A. Tóth, M. Takács, G. Groma, G. Rákhely & K. L. Kovács (2008); ‘A novel NADPH-dependent oxidoreductase with a unique domain structure in the hyperthermophilic archaeon, *Thermococcus litoralis*’. *FEMS microbiology letters*, vol. 282(1):pp. 8–14.
- [21] L. Lauterbach, Z. Idris, K. A. Vincent & O. Lenz (2011); ‘Catalytic properties of the isolated diaphorase fragment of the NAD<sup>+</sup>-reducing [NiFe]-hydrogenase from *Ralstonia eutropha*’. *PloS one*, vol. 6(10).
- [22] L. Lauterbach, J. Liu, M. Horch, P. Hummel, A. Schwarze et al. (2011); ‘The hydrogenase subcomplex of the NAD<sup>+</sup>-reducing [NiFe] hydrogenase from *Ralstonia eutropha* – insights into catalysis and redox interconversions’. *European Journal of Inorganic Chemistry*, vol. 2011(7):pp. 1067–1079.
- [23] G. Peng, B. Meyer, L. Sokolova, W. Liu, S. Bornemann et al. (2018); ‘Identification and characterization two isoforms of NADH: ubiquinone oxidoreductase from the hyperthermophilic eubacterium *Aquifex aeolicus*’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1859(5):pp. 366–373.
- [24] N. A. Losey, S. Poudel, E. S. Boyd & M. J. McInerney (2020); ‘The beta subunit of non-bifurcating NADH-dependent [FeFe]-hydrogenases differs from those of multimeric electron-bifurcating [FeFe]-hydrogenases’. *Frontiers in microbiology*, vol. 11:p. 1109.
- [25] N. Guex & M. C. Peitsch (1997); ‘SWISS-MODEL and the Swiss-PDB Viewer: an environment for comparative protein modeling’. *electrophoresis*, vol. 18(15):pp. 2714–2723.
- [26] J. Rozewicki, S. Li, K. M. Amada, D. M. Standley & K. Katoh (2019); ‘MAFFT-DASH: integrated protein sequence and structural alignment’. *Nucleic acids research*, vol. 47(W1):pp. W5–W10.
- [27] I. Letunic & P. Bork (2019); ‘Interactive Tree Of Life (iTOL) v4: recent updates and new developments’. *Nucleic acids research*, vol. 47(W1):pp. W256–W259.
- [28] D. W. Mulder, E. S. Boyd, R. Sarma, R. K. Lange, J. A. Endrizzi et al. (2010); ‘Stepwise [FeFe]-hydrogenase H-cluster assembly revealed in the structure of HydA  $\Delta$ EFG’. *Nature*, vol. 465(7295):pp. 248–251.
- [29] Y. Nicolet, C. Piras, P. Legrand, C. E. Hatchikian & J. C. Fontecilla-Camps (1999); ‘*Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center’. *Structure*, vol. 7(1):pp. 13–23.
- [30] J. Alric, J. Lavergne, F. Rappaport, A. Verméglio, K. Matsuura et al. (2006); ‘Kinetic performance and energy profile in a roller coaster electron transfer chain: a study of modified tetraheme-reaction center constructs’. *Journal of the American Chemical Society*, vol. 128(12):pp. 4136–4145.
- [31] M. G. Bertero, R. A. Rothery, M. Palak, C. Hou, D. Lim et al. (2003); ‘Insights into the respiratory electron transfer pathway from the structure of nitrate reductase A’. *Nature Structural & Molecular Biology*, vol. 10(9):pp. 681–687.
- [32] C. Gauquelin, C. Baffert, P. Richaud, E. Kamionka, E. Etienne et al. (2018); ‘Roles of the F-domain in [FeFe] hydrogenase’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1859(2):pp. 69–77.
- [33] L. A. Sazanov & P. Hinchliffe (2006); ‘Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*’. *science*, vol. 311(5766):pp. 1430–1436.
- [34] M. Finel (1998); ‘Organization and evolution of structural elements within complex I’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1364(2):pp. 112–121.
- [35] T. Friedrich & B. Böttcher (2004); ‘The gross structure of the respiratory complex I: a Lego System’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1608(1):pp. 1–9.
- [36] W. Buckel & R. K. Thauer (2013); ‘Energy conservation via electron bifurcating ferredoxin reduction and proton/Na<sup>+</sup> translocating ferredoxin oxidation’. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1827(2):pp. 94–113.

- [37] C. C. Page, C. C. Moser, X. Chen & P. L. Dutton (1999); ‘Natural engineering principles of electron tunnelling in biological oxidation–reduction’. *Nature*, vol. 402(6757):pp. 47–52.
- [38] M. Świerczek, E. Cieluch, M. Sarewicz, A. Borek, C. C. Moser et al. (2010); ‘An electronic bus bar lies in the core of cytochrome bc1’. *Science*, vol. 329(5990):pp. 451–454.
- [39] S. Hong, D. Victoria & A. R. Crofts (2012); ‘Inter-monomer electron transfer is too slow to compete with monomeric turnover in bc1 complex’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1817(7):pp. 1053–1062.
- [40] T. Wagner, J. Koch, U. Ermler & S. Shima (2017); ‘Methanogenic heterodisulfide reductase (HdrABC-MvhAGD) uses two noncubane [4Fe-4S] clusters for reduction’. *Science*, vol. 357(6352):pp. 699–703.
- [41] J.-I. Oh & B. Bowien (1998); ‘Structural analysis of the Fds operon encoding the NAD<sup>+</sup>-linked formate dehydrogenase of *Ralstonia eutropha*’. *Journal of Biological Chemistry*, vol. 273(41):pp. 26349–26360.
- [42] B. Schoepp-Cothenet, R. Van Lis, A. Atteia, F. Baymann, L. Capowiez et al. (2013); ‘On the universal core of bioenergetics’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1827(2):pp. 79–93.
- [43] F. Grein, A. R. Ramos, S. S. Venceslau & I. A. Pereira (2013); ‘Unifying concepts in anaerobic respiration: insights from dissimilatory sulfur metabolism’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1827(2):pp. 145–160.
- [44] P. M. Vignais & B. Billoud (2007); ‘Occurrence, classification, and biological function of hydrogenases: an overview’. *Chemical reviews*, vol. 107(10):pp. 4206–4272.
- [45] N. Chongdar, K. Pawlak, O. Rüdiger, E. J. Reijerse, P. Rodríguez-Maciá et al. (2019); ‘Spectroscopic and biochemical insight into an electron-bifurcating [FeFe] hydrogenase’. *JBIC Journal of Biological Inorganic Chemistry*, pp. 1–15.
- [46] J. Meyer, S. L. Andrade & O. Einsle (2011); ‘Thioredoxin-like [2Fe-2S] ferredoxin’. *Encyclopedia of inorganic and bioinorganic chemistry* (2011).
- [47] T. Young, D. Niks, S. Hakopian, T. K. Tam, X. Yu et al. (2020); ‘Crystallographic and kinetic analyses of the FdsBG subcomplex of the cytosolic formate dehydrogenase FdsABG from *Cupriavidus necator*’. *Journal of Biological Chemistry*, vol. 295(19):pp. 6570–6585.
- [48] K. Schuchmann, N. P. Chowdhury & V. Müller (2018); ‘Complex multimeric [FeFe] hydrogenases: biochemistry, physiology and new opportunities for the hydrogen economy’. *Frontiers in microbiology*, vol. 9:p. 2911.
- [49] D. R. Jollie & J. D. Lipscomb (1991); ‘Formate dehydrogenase from *Methylosinus trichosporium OB3b*. purification and spectroscopic characterization of the cofactors.’ *Journal of Biological Chemistry*, vol. 266(32):pp. 21853–21863.
- [50] T. Hartmann & S. Leimkühler (2013); ‘The oxygen-tolerant and NAD<sup>+</sup>-dependent formate dehydrogenase from *Rhodobacter capsulatus* is able to catalyze the reduction of CO<sub>2</sub> to formate’. *The FEBS journal*, vol. 280(23):pp. 6083–6096.
- [51] R. Hille, T. Young, D. Niks, S. Hakopian, T. K. Tam et al. (2020); ‘Structure: Function studies of the cytosolic, Mo-and NAD<sup>+</sup>-dependent formate dehydrogenase from *Cupriavidus necator*’. *Inorganics*, vol. 8(7):p. 41.
- [52] J. J. Kearny & R. D. Sagers (1972); ‘Formate dehydrogenase from *Clostridium acidurici*’. *Journal of bacteriology*, vol. 109(1):pp. 152–161.
- [53] P. Mitchell (1975); ‘Protonmotive redox mechanism of the cytochrome bc complex in the respiratory chain: Protonmotive ubiquinone cycle’. *FEBS Letters*, vol. 56(1):pp. 1–6.
- [54] P. Mitchell (1975); ‘The protonmotive Q cycle: A general formulation’. *FEBS Letters*, vol. 59(2):pp. 137–139.
- [55] W. Buckel & R. K. Thauer (2018); ‘Flavin-based electron bifurcation, a new mechanism of biological energy coupling’. *Chemical Reviews*, vol. 118(7):pp. 3862–3886.
- [56] W. Buckel & R. K. Thauer (2018); ‘Flavin-based electron bifurcation, ferredoxin, flavodoxin and anaerobic respiration with protons (Ech) or NAD<sup>+</sup> (Rnf) as electron acceptors: A historical review’. *Frontiers in Microbiology*, vol. 9.
- [57] N. A. Losey, F. Mus, J. W. Peters, H. M. Le & M. J. McInerney (2017); ‘*Syntrophomonas wolfei* uses an NADH-dependent, ferredoxin-independent [FeFe]-hydrogenase to reoxidize NADH’. *Applied and*

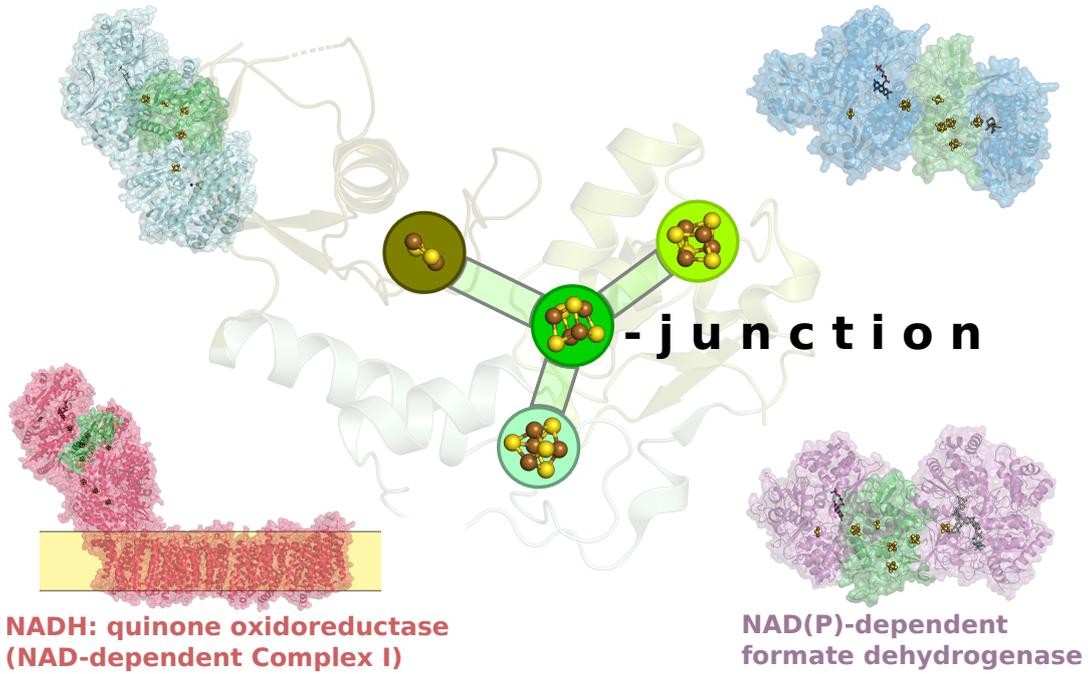
- environmental microbiology, vol. 83(20).
- [58] E. Gnannt, J. Schimpf, C. Harter, J. Hoerer & T. Friedrich (2017); ‘Reduction of the off-pathway iron-sulphur cluster N1a of *Escherichia coli* respiratory complex I restrains NAD<sup>+</sup> dissociation’. *Scientific reports*, vol. 7(1):pp. 1–10.
- [59] K. Dörner, M. Vranas, J. Schimpf, I. R. Straub, J. Hoerer et al. (2017); ‘Significance of [2Fe-2S] cluster N1a for electron transfer and assembly of *Escherichia coli* respiratory complex I’. *Biochemistry*, vol. 56(22):pp. 2770–2778.
- [60] E. Aubert-Jousset, M. Cano, G. Guedeney, P. Richaud & L. Cournac (2011); ‘Role of HoxE subunit in *Synechocystis sp. PCC 6803* hydrogenase’. *The FEBS journal*, vol. 278(21):pp. 4035–4043.
- [61] L. A. Sazanov (2007); ‘Respiratory complex I: mechanistic and structural insights provided by the crystal structure of the hydrophilic domain’. *Biochemistry*, vol. 46(9):pp. 2275–2288.
- [62] F. Baymann, B. Schoepp-Cothenet, S. Duval, M. Guiral, M. Brugna et al. (2018); ‘On the Natural History of Flavin-Based Electron Bifurcation’. *Frontiers in Microbiology*, vol. 9.
- [63] S. Malki, I. Saimmaime, G. De Luca, M. Rousset, Z. Dermoun et al. (1995); ‘Characterization of an operon encoding an NADP-reducing hydrogenase in *Desulfovibrio fructosovorans*’. *Journal of bacteriology*, vol. 177(10):pp. 2628–2636.
- [64] M.-P. Golinelli, C. Chatelet, E. C. Duin, M. K. Johnson & J. Meyer (1998); ‘Extensive ligand rearrangements around the [2Fe-2S] cluster of *Clostridium pasteurianum* ferredoxin’. *Biochemistry*, vol. 37(29):pp. 10429–10437.
- [65] M. F. Verhagen, T. O’Rourke & M. W. Adams (1999); ‘The hyperthermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1412(3):pp. 212–229.
- [66] M. K. Wikström & J. A. Berden (1972); ‘Oxidoreduction of cytochrome b in the presence of antimycin’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 283(3):pp. 403–420.
- [67] G. Herrmann, E. Jayamani, G. Mai & W. Buckel (2007); ‘Energy conservation via electron-transferring flavoprotein in anaerobic bacteria’. *Journal of Bacteriology*, vol. 190(3):pp. 784–791.
- [68] F. Li, J. Hinderberger, H. Seedorf, J. Zhang, W. Buckel et al. (2008); ‘Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/Etf complex from *Clostridium kluyveri*’. *Journal of bacteriology*, vol. 190(3):pp. 843–850.
- [69] L. Michaelis (1932); ‘Theory of the reversible two-step oxidation’. *Journal of Biological Chemistry*, vol. 96(3):pp. 703–715.
- [70] W. M. Clark et al. (1960); ‘Oxidation-reduction potentials of organic systems’. .
- [71] W. Nitschke & M. J. Russell (2011); ‘Redox bifurcations: Mechanisms and importance to life now, and at its origin’. *BioEssays*, vol. 34(2):pp. 106–109.
- [72] A. R. Crofts, S. Hong, C. Wilson, R. Burton, D. Victoria et al. (2013); ‘The mechanism of ubiquinol oxidation at the Q<sub>o</sub>-site of the cytochrome bc<sub>1</sub> complex’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1827(11-12):pp. 1362–1377.
- [73] J. L. Yuly, P. Zhang, C. E. Lubner, J. W. Peters & D. N. Beratan (2020); ‘Universal free energy landscape produces efficient and reversible electron bifurcation’. *arXiv preprint arXiv:2005.14052*.
- [74] L. Bergdoll, F. ten Brink, W. Nitschke, D. Picot & F. Baymann (2016); ‘From low- to high-potential bioenergetic chains: Thermodynamic constraints of Q-cycle function’. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1857(9):pp. 1569–1579.
- [75] J. W. Peters, D. N. Beratan, G. J. Schut & M. W. W. Adams (2018); ‘On the nature of organic and inorganic centers that bifurcate electrons, coupling exergonic and endergonic oxidation–reduction reactions’. *Chemical Communications*, vol. 54(33):pp. 4091–4099.
- [76] S. Duval, J. M. Santini, D. Lemaire, F. Chaspoul, M. J. Russell et al. (2016); ‘The H-bond network surrounding the pyranopterins modulates redox cooperativity in the molybdenum-*bis*PGD cofactor in arsenite oxidase’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1857(9):pp. 1353–1362.
- [77] S. Grimaldi, B. Schoepp-Cothenet, P. Ceccaldi, B. Guigliarelli & A. Magalon (2013); ‘The prokaryotic Mo/W-*bis*PGD enzymes family: a catalytic workhorse in bioenergetic’. *Biochimica et Biophysica*

- Acta (BBA)-Bioenergetics, vol. 1827(8-9):pp. 1048–1085.
- [78] M. Schulte, K. Frick, E. Gnanndt, S. Jurkovic, S. Burschel et al. (2019); ‘A mechanism to prevent production of reactive oxygen species by *Escherichia coli* respiratory complex I’. Nature communications, vol. 10(1):pp. 1–9.
- [79] F. Kremp, J. Roth & V. Müller (2020); ‘The Sporomusa type Nfn is a novel type of electron-bifurcating transhydrogenase that links the redox pools in acetogenic bacteria’. Scientific reports, vol. 10(1):pp. 1–14.
- [80] K. Gutekunst, X. Chen, K. Schreiber, U. Kaspar, S. Makam et al. (2014); ‘The bidirectional NiFe-hydrogenase in *Synechocystis* sp. PCC 6803 is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions’. Journal of Biological Chemistry, vol. 289(4):pp. 1930–1937.
- [81] K. Gutekunst & R. Schulz (2018); ‘The physiology of the bidirectional NiFe-hydrogenase in cyanobacteria and the role of hydrogen throughout the evolution of life’. In Microalgal Hydrogen Production, pp. 107–138 (European Society for Photobiology).
- [82] J. Meyer (2007); ‘[fefe] hydrogenases and their evolution: a genomic perspective’. Cellular and molecular life sciences, vol. 64(9):p. 1063.
- [83] B. Boxma, R. M. de Graaf, G. W. van der Staay, T. A. van Alen, G. Ricard et al. (2005); ‘An anaerobic mitochondrion that produces hydrogen’. Nature, vol. 434(7029):pp. 74–79.
- [84] B. Boxma, G. Ricard, A. H. van Hoek, E. Severing, S.-Y. Moon-van der Staay et al. (2007); ‘The [FeFe] hydrogenase of *Nyctotherus ovalis* has a chimeric origin’. BMC evolutionary biology, vol. 7(1):pp. 1–12.
- [85] J. Preissler, S. Wahlefeld, C. Lorent, C. Teutloff, M. Horch et al. (2018); ‘Enzymatic and spectroscopic properties of a thermostable [NiFe]-hydrogenase performing H<sub>2</sub>-driven NAD<sup>+</sup>-reduction in the presence of O<sub>2</sub>’. Biochimica et Biophysica Acta (BBA)-Bioenergetics, vol. 1859(1):pp. 8–18.
- [86] A. Tran-Betcke, U. Warnecke, C. Böcker, C. Zaborosch & B. Friedrich (1990); ‘Cloning and nucleotide sequences of the genes for the subunits of NAD-reducing hydrogenase of *Alcaligenes eutrophus* H16’. Journal of bacteriology, vol. 172(6):pp. 2920–2929.
- [87] J. H. Artz, M. Tokmina-Lukaszewska, D. W. Mulder, C. E. Lubner, K. Gutekunst et al. (2020); ‘The structure and reactivity of the HoxEFU complex from the cyanobacterium *Synechocystis* sp. PCC 6803’. Journal of Biological Chemistry.
- [88] G. Rákhely, Z. H. Zhou, M. W. Adams & K. L. Kovács (1999); ‘Biochemical and molecular characterization of the [NiFe] hydrogenase from the hyperthermophilic archaeon, *Thermococcus litoralis*’. European journal of biochemistry, vol. 266(3):pp. 1158–1165.
- [89] T. Kanai, S. Ito & T. Imanaka (2003); ‘Characterization of a cytosolic NiFe-hydrogenase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1’. Journal of bacteriology, vol. 185(5):pp. 1705–1711.
- [90] K. Ma, R. Weiss & M. W. Adams (2000); ‘Characterization of hydrogenase ii from the hyperthermophilic archaeon *Pyrococcus furiosus* and assessment of its role in sulfur reduction’. Journal of bacteriology, vol. 182(7):pp. 1864–1871.
- [91] J. Maróti, A. Farkas, I. K. Nagy, G. Maróti, É. Kondorosi et al. (2010); ‘A second soluble hox-type NiFe enzyme completes the hydrogenase set in *Thiocapsa roseopersicina* BBS’. Appl. Environ. Microbiol., vol. 76(15):pp. 5113–5123.
- [92] M. Laukel, L. Chistoserdova, M. E. Lidstrom & J. A. Vorholt (2003); ‘The tungsten-containing formate dehydrogenase from *Methylobacterium extorquens* AM1: Purification and properties’. European journal of biochemistry, vol. 270(2):pp. 325–333.
- [93] A. Graentzdoerffer, D. Rauh, A. Pich & J. R. Andreessen (2003); ‘Molecular and biochemical characterization of two tungsten- and selenium-containing formate dehydrogenases from *Eubacterium acidaminophilum* that are associated with components of an iron-only hydrogenase’. Archives of microbiology, vol. 179(2):pp. 116–130.
- [94] C. F. Nielsen, L. Lange & A. S. Meyer (2019); ‘Classification and enzyme kinetics of formate dehydrogenases for biomanufacturing via CO<sub>2</sub> utilization’. Biotechnology advances.
- [95] R. Hille, J. Hall & P. Basu (2014); ‘The mononuclear molybdenum enzymes’. Chemical reviews, vol.

- 114(7):pp. 3963–4038.
- [96] D. Niks & R. Hille (2019); ‘Molybdenum-and tungsten-containing formate dehydrogenases and formyl-methanofuran dehydrogenases: Structure, mechanism, and cofactor insertion’. *Protein Science*, vol. 28(1):pp. 111–122.
- [97] L. G. Ljungdahl & J. R. Andreesen (1978); ‘Formate dehydrogenase, a selenium-tungsten enzyme from *Clostridium thermoaceticum*’. In *Methods in enzymology*, vol. 53, pp. 360–372 (Elsevier).
- [98] U. Ruschig, U. Müller, P. Willow & T. Höpner (1976); ‘CO<sub>2</sub> reduction to formate by NADH catalysed by formate dehydrogenase from *Pseudomonas oxalaticus*’. *European journal of biochemistry*, vol. 70(2):pp. 325–330.
- [99] H.-Y. Liu, P.-C. Liao, K.-T. Chuang & M.-C. Kao (2011); ‘Mitochondrial targeting of human NADH dehydrogenase (ubiquinone) flavoprotein 2 (NDUFB2) and its association with early-onset hypertrophic cardiomyopathy and encephalopathy’. *Journal of biomedical science*, vol. 18(1):pp. 1–17.
- [100] U. Weidner, S. Geier, A. Ptock, T. Friedrich, H. Leif et al. (1993); ‘The gene locus of the proton-translocating NADH: ubiquinone oxidoreductase in *Escherichia coli*: organization of the 14 genes and relationship between the derived proteins and subunits of mitochondrial complex I’. *Journal of molecular biology*, vol. 233(1):pp. 109–122.
- [101] T. Friedrich & D. Scheide (2000); ‘The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases’. *FEBS letters*, vol. 479(1-2):pp. 1–5.
- [102] R. Hedderich (2004); ‘Energy-converting [NiFe] hydrogenases from archaea and extremophiles: ancestors of complex I’. *Journal of bioenergetics and biomembranes*, vol. 36(1):pp. 65–75.
- [103] P. J. Castro, A. F. Silva, B. C. Marreiros, A. P. Batista & M. M. Pereira (2016); ‘Respiratory complex I: a dual relation with H<sup>+</sup> and Na<sup>+</sup>?’ *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1857(7):pp. 928–937.
- [104] H. Yu, C.-H. Wu, G. J. Schut, D. K. Haja, G. Zhao et al. (2018); ‘Structure of an ancient respiratory system’. *Cell*, vol. 173(7):pp. 1636–1649.
- [105] K. Parey, C. Wirth, J. Vonck & V. Zickermann (2020); ‘Respiratory complex I—structure, mechanism and evolution’. *Current Opinion in Structural Biology*, vol. 63:pp. 1–9.
- [106] W. Chow, I. Ragan & B. H. Robinson (1991); ‘Determination of the cDNA sequence for the human mitochondrial 75-kDa Fe–S protein of NADH-coenzyme Q reductase’. *European journal of biochemistry*, vol. 201(3):pp. 547–550.
- [107] F. Sargent (2016); ‘The model [NiFe]-hydrogenases of *Escherichia coli*’. In *Advances in microbial physiology*, vol. 68, pp. 433–507 (Elsevier).
- [108] W. Nitschke (2018); ‘Carbon Fixation: “Let Things Flow Naturally Forward in Whatever Way They Like”’. *Current Biology*, vol. 28(3):pp. R110–R112.
- [109] R. Baradaran, J. M. Berrisford, G. S. Minhas & L. A. Sazanov (2013); ‘Crystal structure of the entire respiratory complex I’. *Nature*, vol. 494(7438):pp. 443–448.
- [110] J. A. Letts, K. Fiedorczuk, G. Degliesposti, M. Skehel & L. A. Sazanov (2019); ‘Structures of respiratory supercomplex I+III<sub>2</sub> reveal functional and conformational crosstalk’. *Molecular cell*, vol. 75(6):pp. 1131–1146.
- [111] K. Kayastha, S. Vitt, W. Buckel & U. Ermler (2021); ‘Flavins in the electron bifurcation process’. *Archives of Biochemistry and Biophysics*, vol. in press.
- [112] M.-C. Lett, D. Muller, D. Lièvreumont, S. Silver & J. Santini (2012); ‘Unified nomenclature for genes involved in prokaryotic aerobic arsenite oxidation’. *Journal of bacteriology*, vol. 194(2):pp. 207–208.

**NAD(P)-dependent  
[NiFe]-Hydrogenase**

**NAD(P)-dependent  
[FeFe]-Hydrogenase**



Graphical abstract

**Table 1: Subunit and cofactor nomenclature**

Subunit and cofactor composition of the characterised protein complexes containing the Y-junction module. Coloring scheme for each protein complex and symbols used to denote the presence of specific cofactor binding sites same throughout the paper. Subunits in the same rows (from left to right) are homologous to each other. Cofactor binding sites in the same row (from right to left) are homologous to each other. Bold names are used throughout the publication to refer back to specific subunits. A more extensive nomenclature table specifically for [FeFe]-Hydrogenases and Formate Dehydrogenase is given in Supplementary Table 2.

n.a. - not available, *i.e.* cofactor not present

n.n. - no naming convention

SU - subunit

<sup>a</sup> - Nomenclature according to Peters *et al.* (1998) for [FeFe]-hydrogenase (in blue) [3]

<sup>b</sup> - Nomenclature according to Sazanov and Hincliffe (2006) for complex I (in red) [33]

<sup>c</sup> - Nomenclature according to Shomura *et al.* (2017) for bidirectional [NiFe]-hydrogenase (in cyan) [6]

<sup>d</sup> - Nomenclature according to Radon *et al.* (2020) for formate dehydrogenase (in purple) [7]

Subunit nomenclature used in this article	Multimeric [FeFe]-hydrogenase <sup>a</sup>	Complex I <sup>b</sup>	Bidirectional [NiFe]-hydrogenase <sup>c</sup>	Formate dehydrogenase <sup>d</sup>	Symbol	Cofactor			
YJM-subunit	HndD (HydA)	NuoG (Nqo3, 75 kDa SU)	HoxU	FdsA (alpha SU)	A5	[2Fe-2S]			
					FS2	N1b	U1		
					FS4C	N5	U3		[4Fe-4S]
					FS4B	N4	U2		[4Fe-4S]
					FS4A	n.n.	n.n.		[4Fe-4S]
					n.a.	N7	n.a.		[4Fe-4S]
					n.a.	n.a.	n.a.	Mo/W-pterin	
	H-cluster	n.a.	n.a.	n.a.					
flavin/[4Fe-4S] subunit	HndC (HydB)	NuoF (Nqo1, 51 kDa SU)	C-terminal HoxF	FdsB (beta SU)		[2Fe-2S]			
						flavin			
					n.n.	N3	F1		[4Fe-4S]
[2Fe-2S]-binding subunit	HndA (HydC)	NuoE (Nqo2, 24 kDa SU)	N1a	HoxE or N-terminal HoxF	n.n.	FdsG (gamma SU)	G7		[2Fe-2S]
small subunit	HndB (HydD)	n.a.		n.a.		FdsX			
		n.a.		n.a.		FdsD (delta SU)			

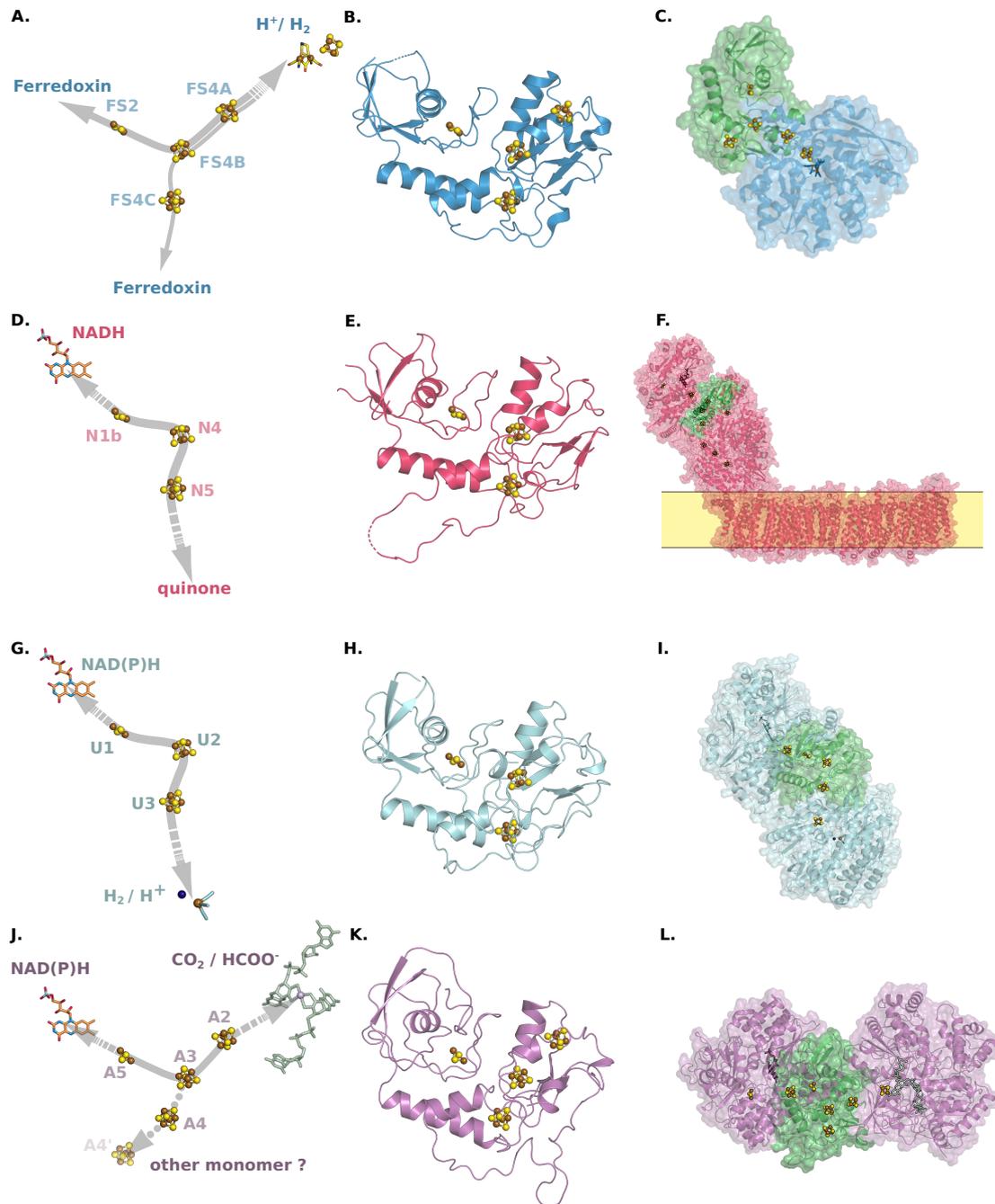


Figure 1: **Electron-transfer pathways through the Y-junction modules from different enzymes.** Comparison of the electron-transfer pathways in the first column (A., D., G. and J. with donor/acceptor location not to scale) enabled by the structurally conserved Y-junction module in the second column (B., E., H. and K., color of Y-junction module according to the general coloring scheme for each of these four protein complexes will be used throughout this article) from 4 bioenergetic proteins. Location of the Y-junction module (in green) in the structural representatives of the prototypical [FeFe]-hydrogenase (C., in blue, PDB: 6N59, Y-junction module in residue 1-209), of membrane-bound complex I (F., in red, PDB: 4HEA, membrane in beige, Y-junction module in residue 1-239 of NuoG), bidirectional [NiFe]-hydrogenase (I., in cyan, PDB: 5XF9, Y-junction module is hoxU) and in formate dehydrogenase (L., in purple, PDB: 6TG9, Y-junction module in residue 1-248 of FdhA) in the last column (relative protein sizes not to scale). The denomination of the iron-sulphur clusters in the first column (as in Table 1) follows Peters *et al.* (1998) [3] for [FeFe]-hydrogenase in A., Sazanov and Hinchliffe (2006) [33] for complex I in D., Shomura *et al.* (2017) [6] for [NiFe]-hydrogenase in G. and Radon *et al.* (2020) [7] for formate dehydrogenase in J. The same color coding for each of these four protein complexes will be used throughout this article.

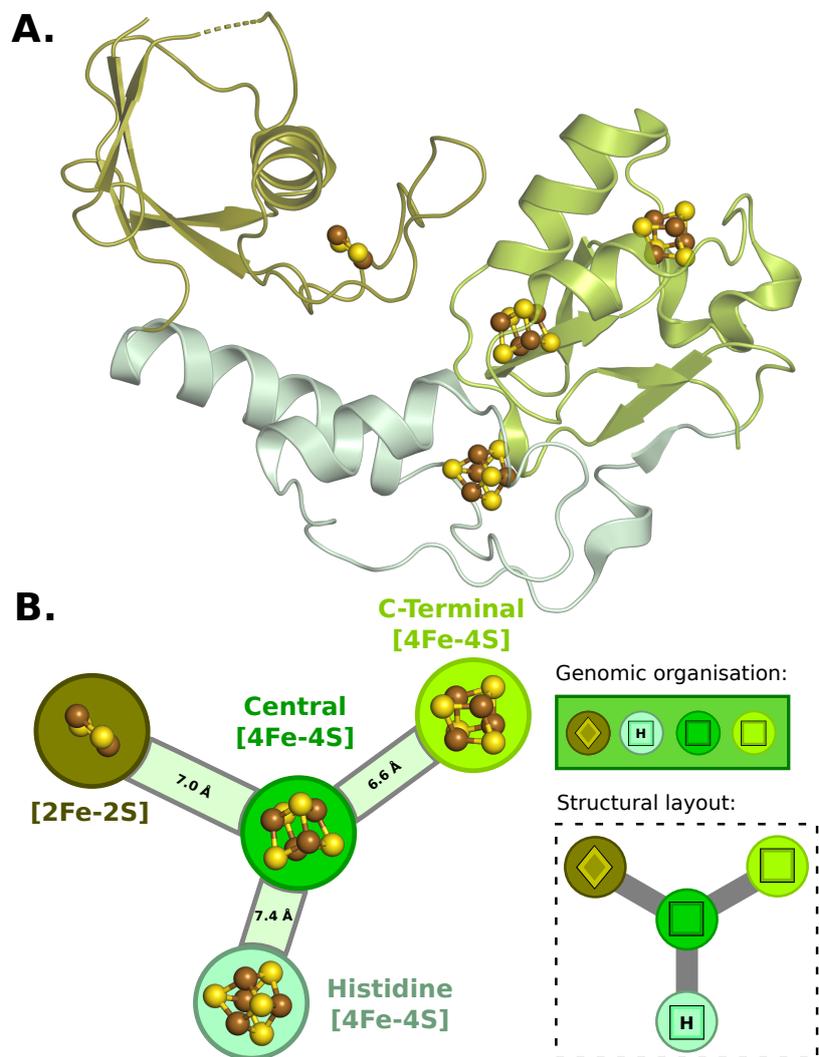
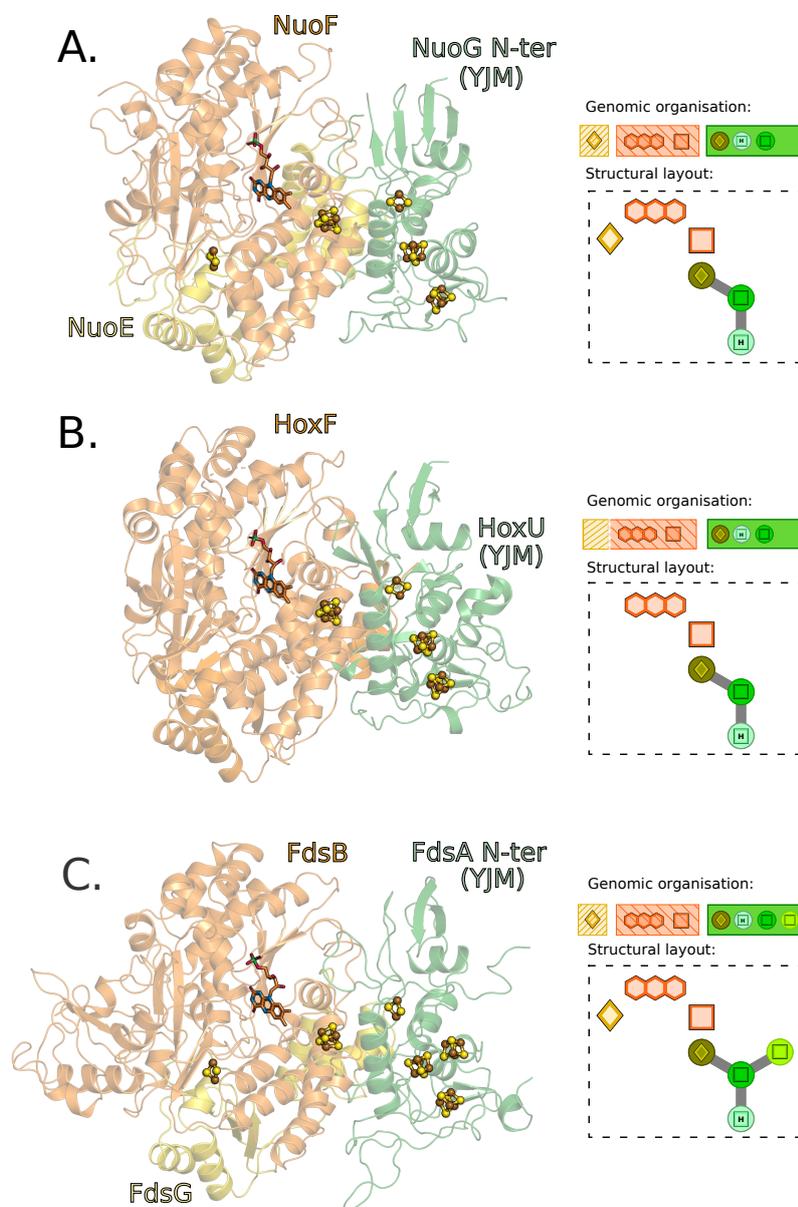


Figure 2: **Domain-architecture of the Y-junction module.**

A. The Y-junction module is constructed of a plant-type-[2Fe-2S]-ferredoxin domain (in olive), a histidine-[4Fe-4S]-binding domain (in mint) and a regular 2x[4Fe-4S]-ferredoxin domain (in lime) [41].

B. Nomenclature of the 4 electron-conducting iron-sulphur clusters in the Y-junction module and subsequent representation of the genomic organisation and structural layout of this protein module as symbols and pictograms. Numbers on the edges between the clusters show averaged edge-to-edge distances (in Å) that include the ligating residues. PDB-structure used for this module, [FeFe]-hydrogenase from *C. pasteurianum* (CpI): 6N59.



**Figure 3: Structural comparison of the “flavin module” from different complexes.**

Structural juxtaposition of the “flavin module” of the respiratory complex I from *Thermus thermophilus* (A., PDB: 4HEA), the NAD-dependent [NiFe]-hydrogenase from *Hydrogenophilus thermoluteolus* (B., PDB: 5XF9) and the NAD-dependent formate dehydrogenase from *Rhodobacter capsulatus* (C., PDB: 6TG9). Subunits are colored according to homology with the flavin-binding subunit in orange, the [2Fe-2S]-cluster-containing subunit in yellow and the Y-junction module in green. Iron-sulphur clusters are brown and yellow for iron and sulphur, respectively. Carbons in the flavin are orange. General genomic organisation and structural layout in accordance with Fig. 4, and Fig. 5a and 5b, where the pictograms for the operon depictions and structural arrangements are given.

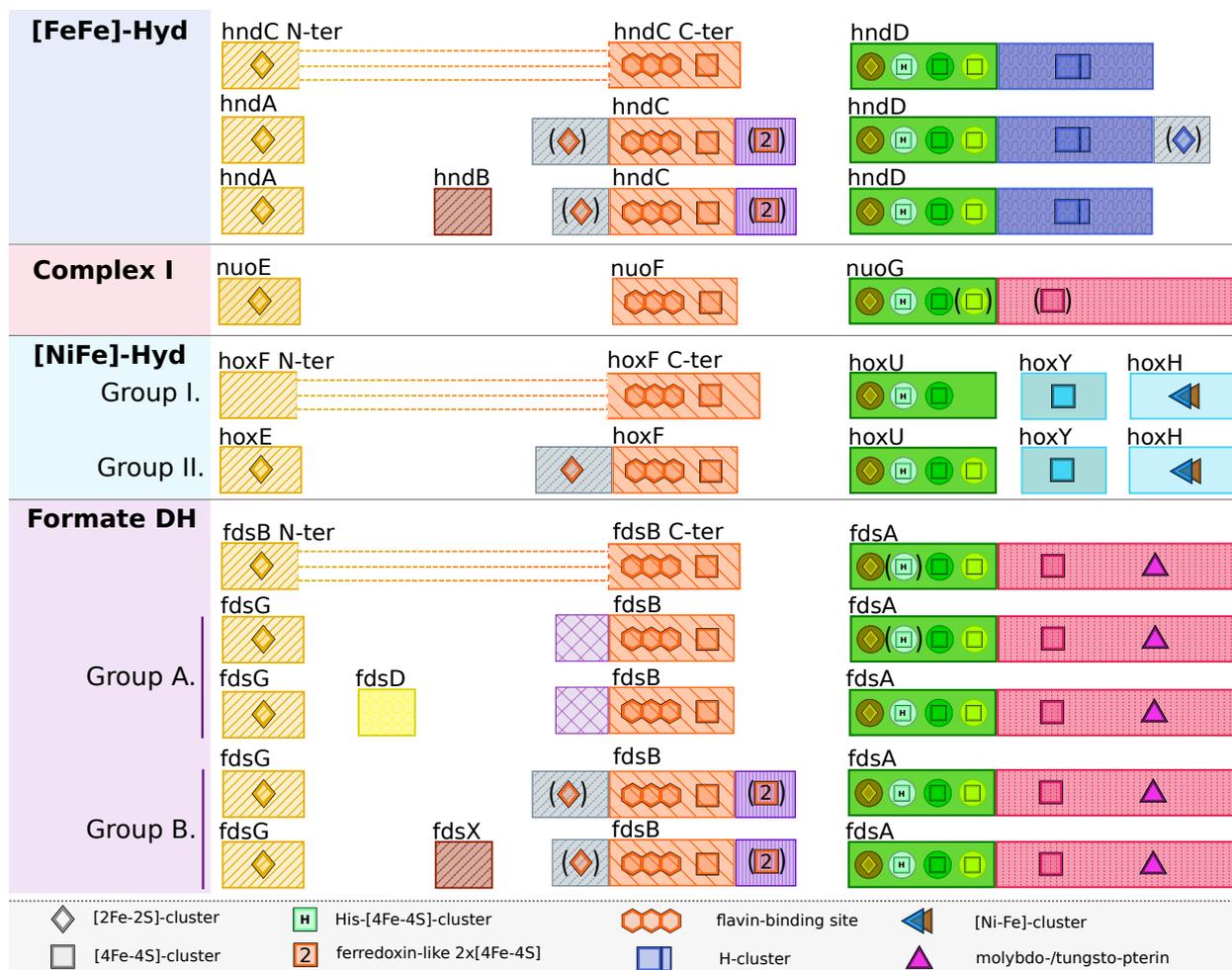


Figure 4: **Operon organization for the bioenergetic proteins containing the Y-junction module.** Gene cluster organization in [FeFe]-hydrogenase (in blue), complex I (in red), bidirectional [NiFe]-hydrogenase (in cyan) and the formate dehydrogenase (in purple). Sequence motifs for the coordination of cofactors are depicted by colored symbols in accordance with the pictograms used in Fig. 2, Fig. 3, Fig. 5a and 5b and summarised in Tab. 1. Symbols enclosed by parenthesis indicate the presence of the cofactor and/or binding domain of that cofactor in a majority of considered sequences. Identical colors and underlying patterns of the sequence stretches indicate homologous regions. Y-junction module is shown in green. Group definitions follow compositional and functional differences of proteins belonging the same protein complex. Coloring of the iron-sulphur clusters corresponds to its location in the operon analogously as in the Fig. 2, Fig. 3, Fig. 5a and 5b. Sequence lengths not to scale.

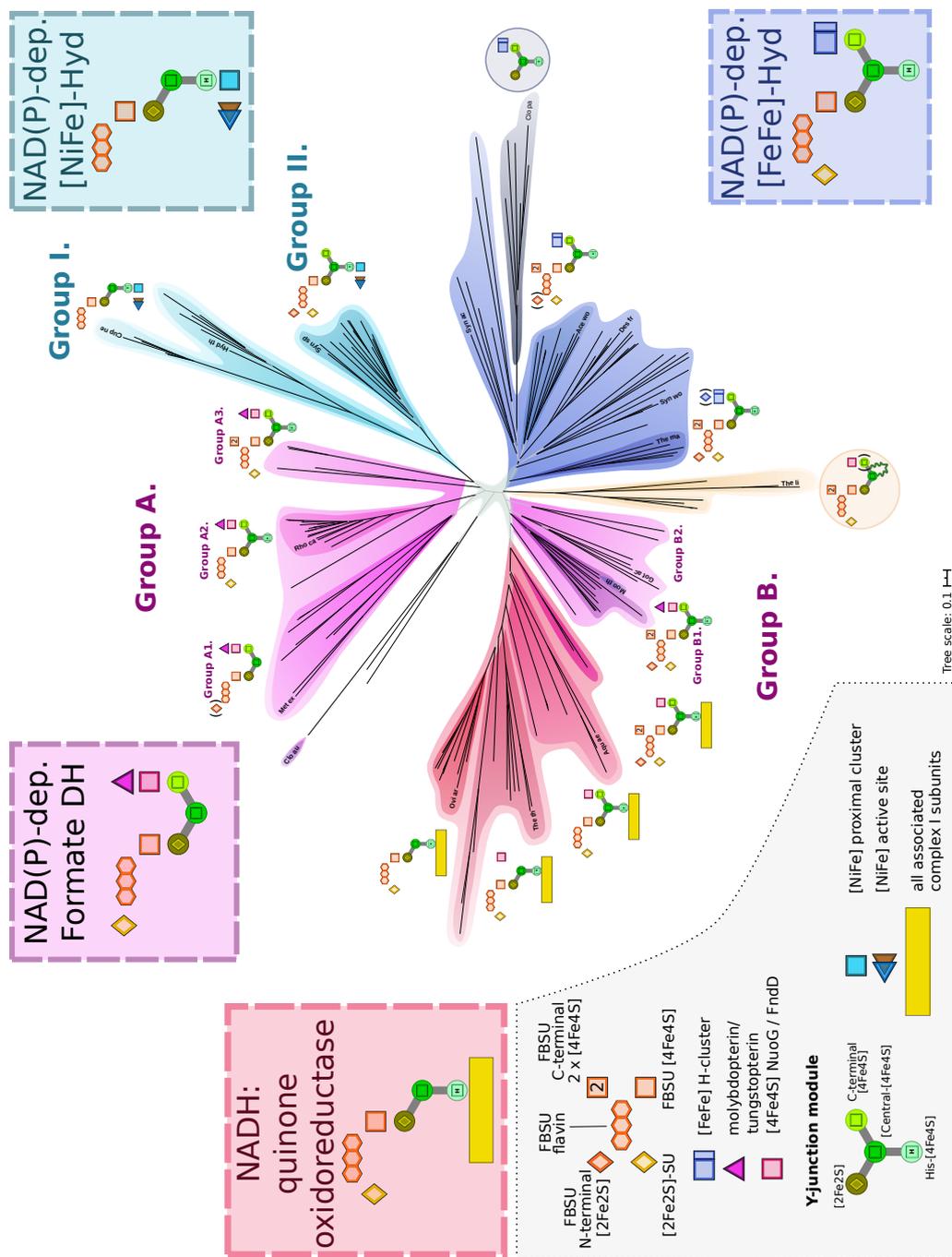


Figure 5a: **Neighbor-Joining Tree of the Y-junction module in representative sequences.**

Similarity clustering of the sequences belonging to the [FeFe]-hydrogenase (in blue), complex I (in red), [NiFe]-hydrogenase (in cyan) and formate dehydrogenase (in purple). Minimal subunit and cofactor composition of the proteins including the [2Fe-2S]-containing subunit ([2Fe-2S]-SU), flavin/[4Fe4S]-binding subunit (FBSU), as well as specific subunits for all members of each protein family are shown as pictograms in the colored square. Additional specific cofactors characterising a specific group inside each protein family can be found as small pictograms at the representative leaves of the tree. Symbols enclosed by parentheses indicate the presence of the cofactor and/or the binding domain of that cofactor in the majority of sequences of the respective group. The symbols used are the same throughout all figures and tables. Grey clouds and grey-colored branches indicate bootstrap values below 50%. Species name abbreviations are shown for selected characterised enzyme complexes. Fully annotated tree can be found in Supplementary Figure 4a.

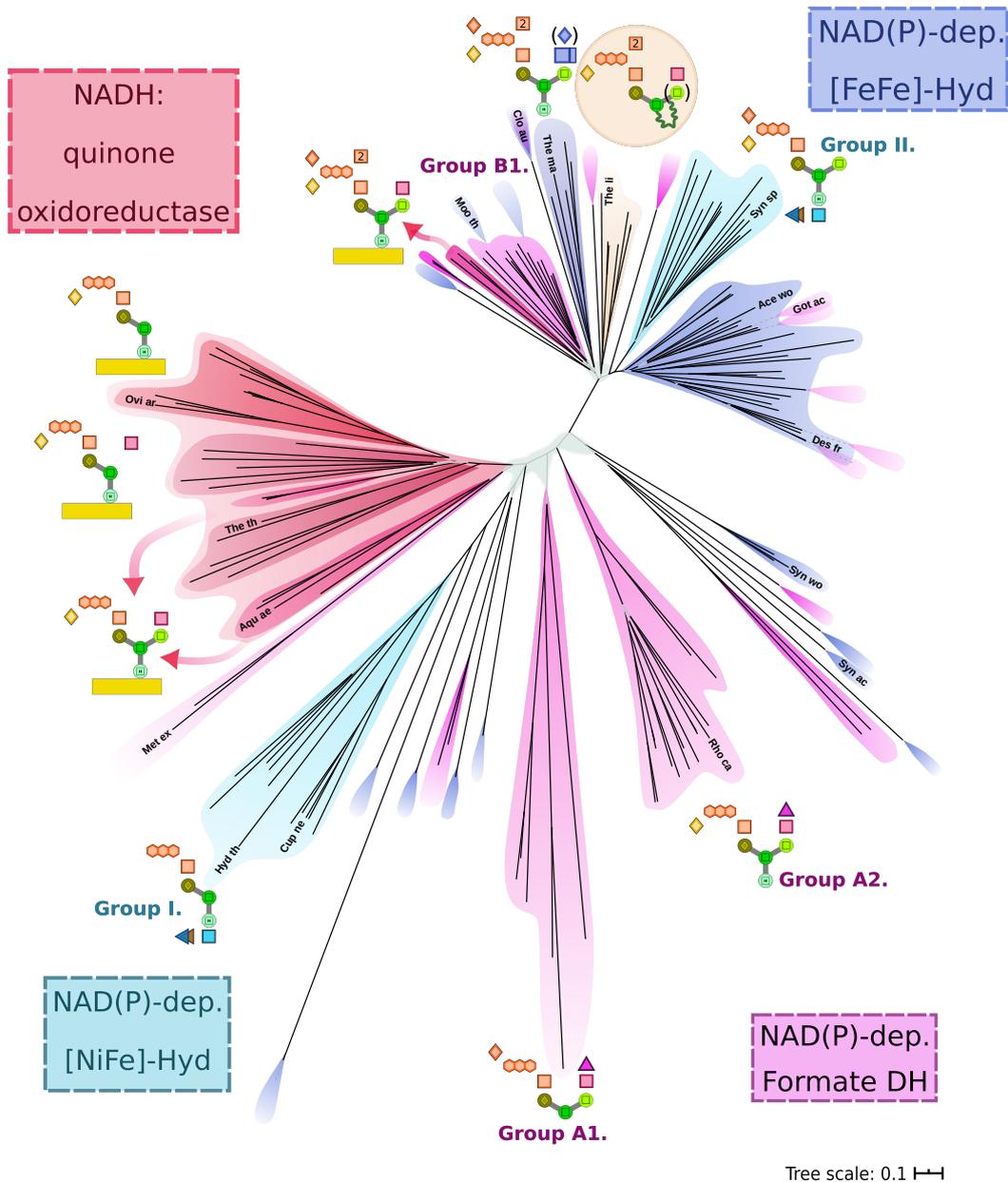


Figure 5b: **Neighbor-Joining Tree of the flavin module in representative sequences.** Similarity clustering of the sequences belonging to the [FeFe]-hydrogenase (in blue), complex I (in red), [NiFe]-hydrogenase (in cyan) and formate dehydrogenase (in purple). Cofactor composition of the enzymes are shown as pictograms as in Figure 5a. Additional specific cofactors characterising a specific group inside each protein family can be found as small pictograms at the representative leaves of the tree. Symbols enclosed by parentheses indicate the presence of the cofactor and/or the binding domain of that cofactor in the majority of sequences of the respective group. The symbols used are the same throughout all figures and tables. Grey clouds and grey-colored branches indicate bootstrap values below 50%. Species name abbreviations are shown for selected characterised enzyme complexes. Fully annotated tree can be found in Supplementary Figure 4b.

# Supplementary - The dyad of the Y-junction- and a flavin module unites diverse redox enzymes

Kilian Zuchan<sup>a</sup>, Frauke Baymann<sup>a</sup>, Carole Baffert<sup>a</sup>, Myriam Brugna<sup>a,\*</sup>, Wolfgang Nitschke<sup>a</sup>

<sup>a</sup>*Aix Marseille Univ, CNRS, BIP, 31 Chemin Joseph Aiguier, 13402, Marseille Cedex 09, France*

---

Supplementary Table 3 containing all sequences considered in this study are given as a separate file: SupplementaryTable3.ods

## Section 1. Complex I, an enzyme with a somewhat ambiguous naming scheme

The name complex I historically derives from the fact that it is the first in a chain of membrane-integral enzymes through which electrons pass in mitochondrial respiration on their way to the terminal electron acceptor, O<sub>2</sub>. This enzyme couples the redox conversions of NAD<sup>+</sup>/NADH and quinone/quinol to the translocation of protons across energy converting membranes and is frequently also referred to as “type 1 NADH dehydrogenase” or “NDH-1” (NDH-2’s are structurally much simpler enzymes which perform the exactly same redox conversions but without pumping protons [10]).

We contend that the term “type 1 NAD:quinone oxidoreductase” is more appropriate since (a) the enzyme is known to operate in both directions and (b) the second redox substrate always is a quinone [11]. By contrast, the NAD<sup>+</sup>/NADH couple can be substituted by other substrates. Several enzymes have in fact been characterised which lack the subunits involved in the redox conversion of NAD but which retain the full remaining complement of subunits, e.g. in cyanobacteria [12, 13]. These enzymes are sometimes also referred to as complex I but they aren’t “NAD-dependent” but transfer electrons between other redox compounds and the quinol/quinone couple. The terms complex I and NDH-1 therefore cannot be used interchangeably.

Part of the evolutionary analyses detailed in the main text pertains to the evolutionary origins of precisely that subgroup of complexes I which truly can operate with NAD<sup>+</sup>/NADH and, *a fortiori*, to the evolutionary history of the NuoEFG-module.

We therefore refer throughout this article to this subgroup as “NAD-dependent” complex I to distinguish it from the wider family of proton-translocating complexes I, which may use a range of other redox compounds instead of NAD, such as for example ferredoxins or coenzyme F<sub>420</sub> [14, 15]. To address the different subunits of NAD-dependent complex I we will furthermore use the older “Nuo”-nomenclature originally referring to NAD:ubiquinone oxidoreductase. However, since complex I also functions in menaquinone species, the naming

---

\*Corresponding author

Email address: mbrugna@imm.cnrs.fr (Myriam Brugna)

of subunits introduced from the enzyme from *Thermus thermophilus* (Nqo) would rigorously be the more appropriate one [16]. For readers used to these alternative nomenclatures we compiled currently used abbreviations for concerned subunits of this and the other enzymes discussed in this article in Table 1.

## Section 2. The Histidine-[4Fe-4S] cluster - a case for structure informed alignments

Multiple sequence alignments can be guided by the available 3D structures depicted in Fig. 1 and improvements in reliability of trees (as judged by their bootstrap values) are observed upon correcting algorithm-based alignments by structure-based ones [17, 18]. An intriguing example of deviation between these two methods is provided by the position of the histidine residue ligating the "histidine-[4Fe-4S]" cluster in Suppl. Fig. 3. The full conservation of a histidine as ligand biases algorithm-based alignment methods into aligning all these histidines by generating indels shortly before and after the histidine residue to compensate for the fact that in NAD-dependent [NiFe]-hydrogenases the histidine doesn't align straightforwardly with those of all the other enzymes. The structural alignment of the known protein models (Suppl. Fig. 3A), however, shows that the polypeptide chains overlap perfectly without any indels and that the histidine in the Y-junction module of the NAD-reducing [NiFe]-hydrogenase is indeed at a sequence and structural position that differs from that of the other cases (Fig. 3B). To still be able to serve as a ligand to the cluster, coordination of the iron-atom is achieved by the  $N_{\delta}$ -atom of the imidazole moiety in the NAD-dependent [NiFe]-hydrogenases rather than by the  $\epsilon$ -nitrogen in all other cases (Fig. 3A). This particularity of the NAD-dependent [NiFe]-hydrogenase seen in the *H. thermoluteolus* structure was found to be conserved in all other members of this family detected in genomes as described in this article. However, we also detected a few cases of enzymes with histidines simultaneously present in both positions (e.g. the case of the complex I enzyme in *Escherichia coli* shown in Fig. 3B) suggesting a likely pathway for the evolutionary transition from one position to the other. Whether the strict conservation of the deviant histidine position in NAD-dependent [NiFe]-hydrogenases is a mere phylogenetic trait or whether it plays a functional role (e.g. by tuning the redox midpoint potential of the cluster) remains to be elucidated.

## Section 3. Cases of NAD(P)-dependent formate dehydrogenases which appear as structural and possibly functional outliers

A specific member of Group B2 in the NAD(P)-dependent formate dehydrogenases displays an intriguing architecture. In addition to having exchanged its proper flavin module against one from an NAD(P)-dependent [FeFe]-hydrogenase, the gene cluster encoding a putative NAD(P)-dependent formate dehydrogenase in *Pelobacter propionicus* DSM 2379 (107-Pel pr, Suppl. Fig. 4a and Fig. 4b) contains all subunits required for the assembly of a complex I. Albeit this complex harbours a NuoD-like subunit featuring the binding motif (CxxC-CxxC) for the [NiFe]-catalytic centre and a NuoG subunit predicted to harbour all

four [Fe-S]-clusters in the Y-junction module, the FS0-cluster of the formate dehydrogenase catalytic domain and a cysteine residue in the canonical position of the protein-ligand to the molybdenum. The first few features were already noticed by Marreiros *et al.*(2013), whereas the additional occurrence of a CO<sub>2</sub>-reducing site would allow the Y-junction module to function as a three-way electron conduit between these different reaction sites [19]. The assembled genome of this organism, however, features several strictly identical (with respect to sequences of ORFs) gene clusters coding for *bona fide* complexes I. Such a situation being extremely unlikely, a further verification of the reliability of this genome sequence might be required before firm conclusions can be drawn.

Three formate dehydrogenase sequences (to the left of Group A1 in Suppl. Fig. 4a) retrieved in our genomic survey do not cluster with any of the above delineated Fds-groups as mentioned in the main text (nor with any other group in Suppl. Fig. 4a and Fig. 4b). They furthermore strongly diverge among each other with respect to their structural layout. Future expansion of genomic databases may yield the emergence of further classes encompassing these cases or else they may represent phylogenetic artifacts resulting from long-branch attraction, a common pitfall in phylogenetic analyses [20]. Excessively long branches may be due to rapid evolution entailed by the emergence of uncommon functional roles. The enzymatically characterised enzyme from *Clostridium autoethanogenum* [6, 21] which is part of this group, might favour the latter explanation. This enzyme actually represents a chimera between an NAD(P)-dependent formate dehydrogenase and an NAD(P)-dependent [FeFe]-hydrogenase and probably serves as a “ferredoxin- and hydrogen-dependent CO<sub>2</sub>-reductase” as discussed by Schuchmann *et al.* (2018). Its gene cluster as well as its biochemically verified subunit composition suggests the presence of the flavin module, a Y-junction module (again represented by a separate protein and lacking not only the histidine-cluster but also the entire histidine-cluster subdomain) as well as formate dehydrogenase (lacking an N-terminal Y-junction module) and [FeFe]-hydrogenase catalytic subunits (with two cubane iron-sulphur clusters N-terminal domain to the H-domain of the hydrogenase), supplemented by two additional tetracubane [16Fe-16S]-ferredoxin subunits [22]. Neither the formate dehydrogenase subunit nor the [FeFe]-hydrogenase one contain a complete Y-junction module as N-terminal domain. A straightforward deduction of the architecture of this enzyme based on known 3D-structures is thus not possible.

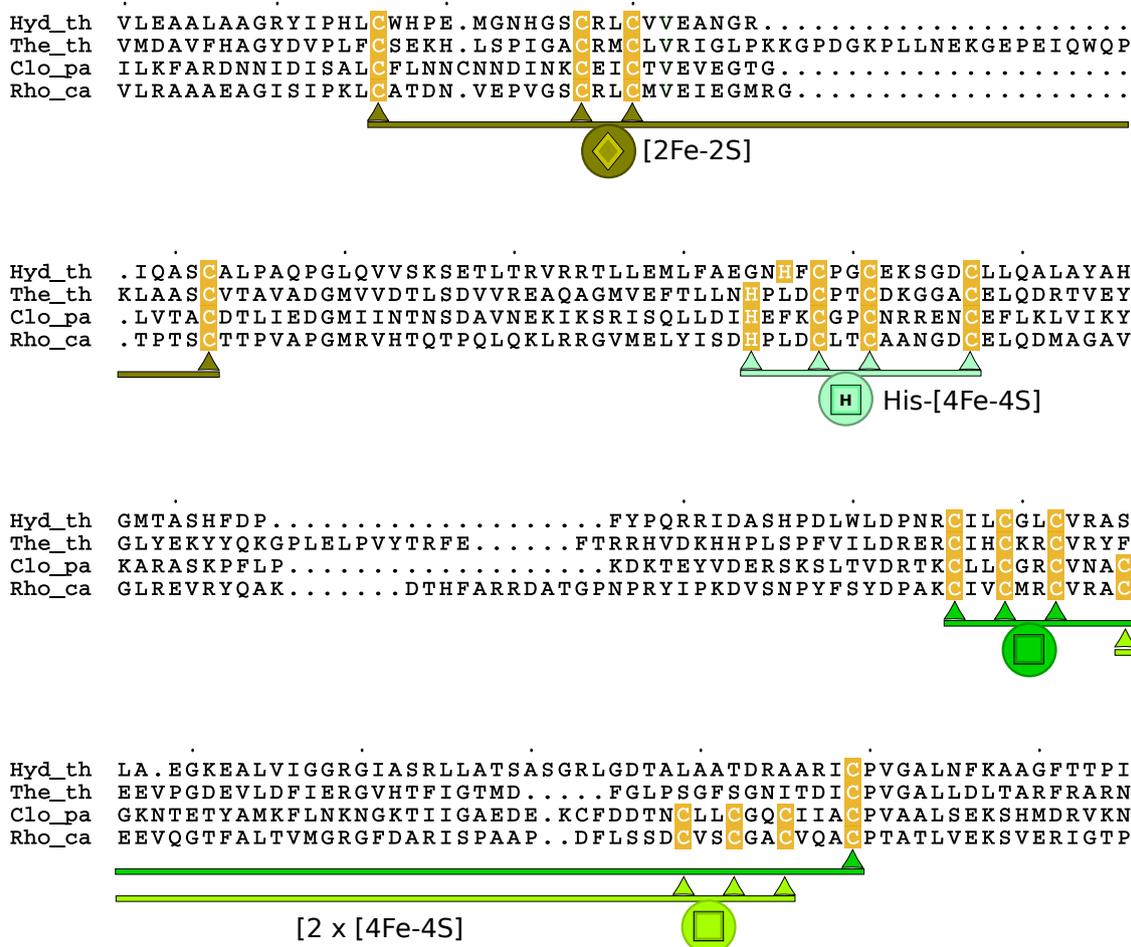
Supplementary Table 1: Sequence motifs used for the identification of cofactors in the respective proteins.

Symbol	Respective cofactor	Motif
	[2Fe-2S]-cluster in [2Fe-2S] binding SU	<b>CxxxxC-...-CxxxC</b>
	[2Fe-2S]-cluster in flavin/[4Fe-4S]-SU	
	YJM [2Fe-2S]-cluster	<b>Cx{10}CxxCx{10-13}C</b>
	YJM His-[4Fe-4S]-cluster	<b>see Supplementary Section 2 and Fig. 1 &amp; 3</b>
	YJM Central and C-Terminal [4Fe-4S]-cluster	<b>CxxCxxCxxxC-...-CxxCxxCxxxC</b>
	2x[4Fe-4S] in flavin/[4Fe-4S]-SU	
	[4Fe-4S] in flavin/[4Fe-4S]-SU	<b>CxxCxxC-...-C</b>
	flavin-binding site in flavin/[4Fe-4S]-SU	<b>[GSA]xxG[ASR][FY]CG[ED]</b>
	NAD(P)-binding site in flavin/[4Fe-4S]-SU	<b>GxGxx[GA]xxx[GA]x{6}[G]</b>
	[4Fe-4S]-proximal cluster in [NiFe] Hydrogenases	<b>CxxC-...-C-...-C</b>
	[NiFe] catalytic cluster in [NiFe] Hydrogenases	<b>CxxC-...-CxxC</b>
	H-cluster in [FeFe] Hydrogenases	<b>C-...-C-...-CxxxC</b>
	N1/A7 [4Fe-4S]-cluster in complex I/Fds	<b>CxxCxxxC-...-C</b>
	molybdopterin/tungstopterin in Fds	<b>[CU]-...-H-...-R</b>

Supplementary Table 2: **Overview subunit nomenclature of [FeFe]-Hydrogenase and Formate dehydrogenase.**

Recently used subunit nomenclature for homologous subunits in [FeFe]-Hydrogenase and formate dehydrogenase. Additional non-homologous subunits characterising specific protein complexes not shown.

Organism	Oligomerization state	Name	[2Fe-2S] subunit	Flavin/[4Fe-4S] subunit	YJM containing subunit	H-cluster containing subunit	Mo-/W-containing subunit	
<i>Desulfovibrio fructosovorans</i>	Tetrameric	<i>Hnd</i>	HndA	HndC	HndD	HndD	-	Kpebe et al. (2018)
<i>Desulfovibrio fructosovorans</i>	Trimeric	<i>Hnt</i>	HntA	HntC	HntD	HntD	-	Baffert et al. (2019)
<i>Thermotoga maritima</i>	Trimeric	<i>Hyd</i>	HydC	HydB	HydA	HydA	-	Chongdar et al. (2019)
<i>Acetobacterium woodii</i>	Tetrameric	<i>Hyd</i>	HydC	HydB	HydA	HydA	-	Schuchmann & Müller (2012)
<i>Eubacterium acidaminophilum</i>	Tetrameric	<i>Hym</i>	HymA	HymB	HymC	HymC	-	Graentzdoerffer et al. (2002)
<i>Clostridium autoethanogenum</i>	Heptameric	<i>Hyt</i>	HytC	HytB	HytD	HytA	FdhA	Wang et al. (2013)
<i>Clostridium acidurici</i>	Tetrameric	<i>Hyl</i>	HylC	HylB	HylA, FdhF2, FdhF1	-	FdhF2, FdhF1	Wang et al. (2013)
<i>Cupriavidus necator</i>	Trimeric	<i>Fds</i>	FdsG	FdsB	FdsA	-	FdsA	Hille et al. (2020)
<i>Rhodobacter capsulatus</i>	Tetrameric	<i>Fds</i>	FdsG	FdsB	FdsA	-	FdsA	Radon et al. (2020)
<i>Methylobacterium extorquens</i>	Dimeric	<i>Fdh</i>	Fdh1B	Fdh1B	Fdh1A	-	Fdh1A	Laukel et al. (2003)



Supplementary Figure 1: **Structural alignment of the Y-junction module.**

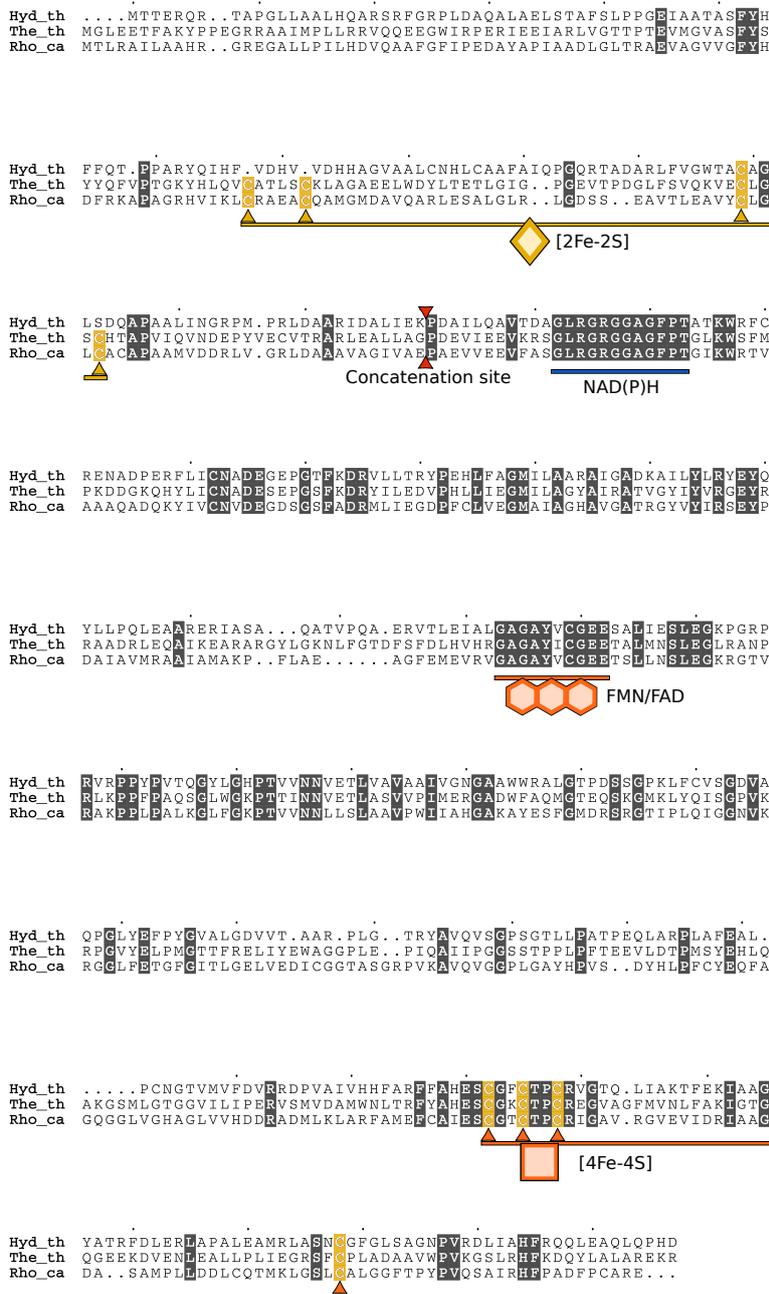
Structural alignment of the four structurally known template sequences of the Y-junction module were constructed using MAFFT-DASH by employing the iterative refinement algorithm as described in the methods section [18]. Multiple sequence alignment initially visualized with the ESPrict 3.0 Server [23]. Cofactor annotation according to specified motifs in Supplementary Table 1. Functional cluster annotation congruent with the pictograms used throughout the article.

Hyd th - HoxU subunit, [NiFe]-Hydrogenase of *Hydrogenophilus thermoluteolus TH-1*

Clo pa - N-terminal [FeFe]-Hydrogenase of *Clostridium pasteurianum DSM 525*

The th - N-terminal NuoG subunit, complex I of *Thermus thermophilus HB8*

Rho ca - N-terminal FdsA subunit, Formate dehydrogenase of *Rhodobacter capsulatus YW2*



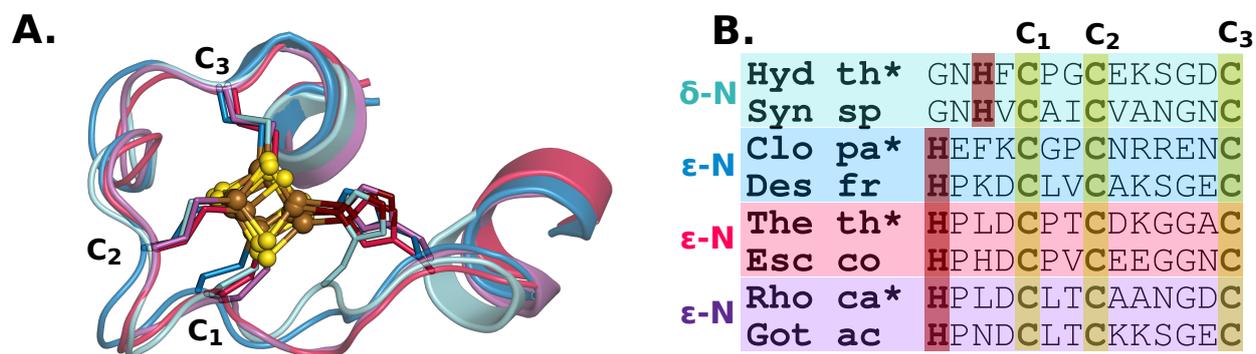
Supplementary Figure 2: **Structural alignment of the flavin module.**

Structural alignment of the three structurally characterised template sequences of the "flavin module" as shown in Figure 3 of the main text were constructed using MAFFT-DASH by employing the iterative refinement algorithm as described in the methods section [18]. Multiple sequence alignment initially visualized with the ESPript 3.0 Server [23]. Cofactor annotation according to specified motifs in Supplementary Table 1. Functional cluster annotation congruent with the pictograms used throughout the article.

Hyd th - HoxF subunit, [NiFe]-Hydrogenase - *Hydrogenophilus thermoluteolus TH-1*

The th - concatenated NuoE and NuoF subunit, complex I - *Thermus thermophilus HB8*

Rho ca - concatenated FdsG and FdsB subunit, Formate dehydrogenase - *Rhodobacter capsulatus YW2*

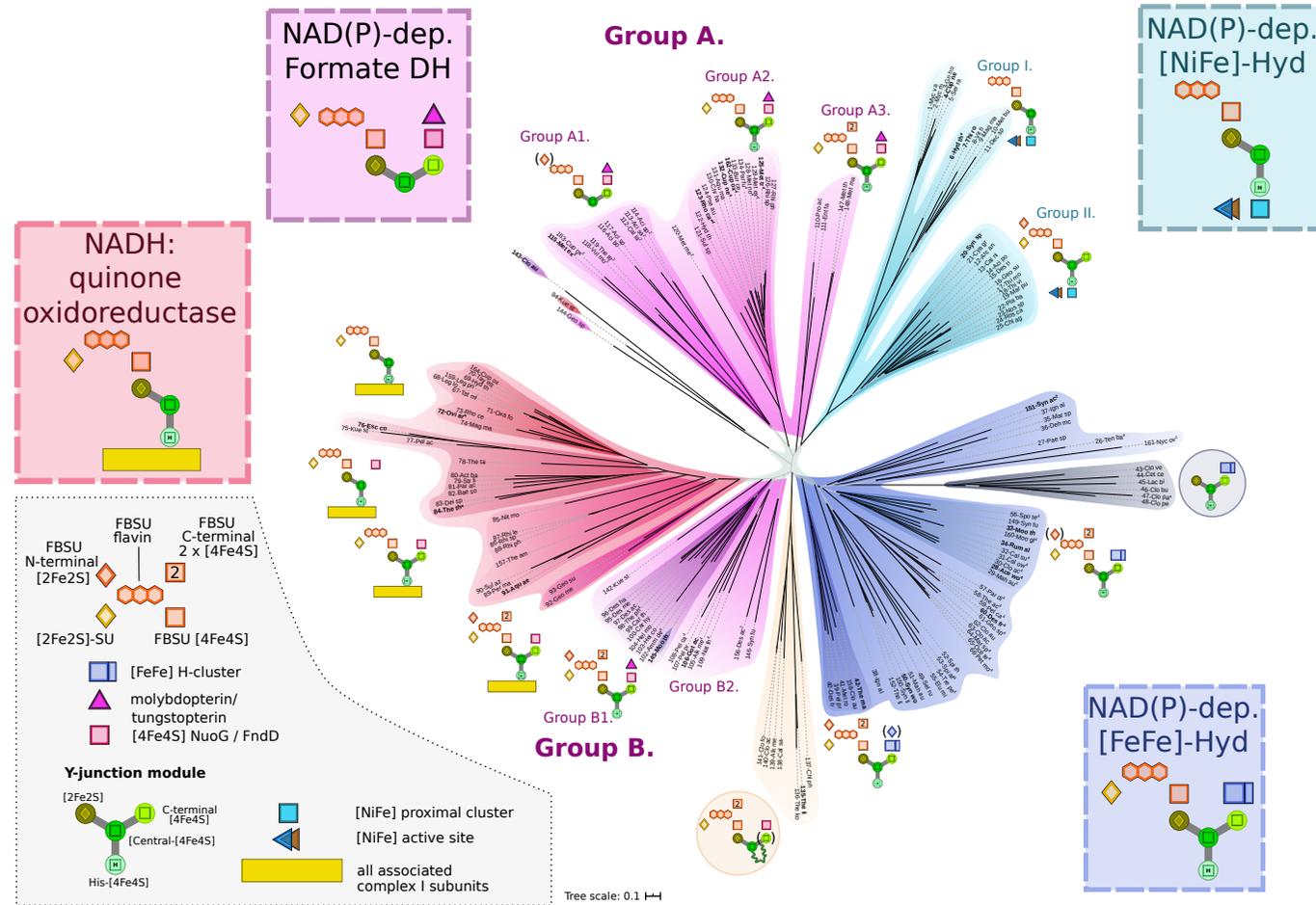


Supplementary Figure 3: **Coordination of the Histidine-[4Fe4S]-cluster of the examined bioenergetic proteins.**

Structural (A.) and sequence (B.) comparison of the coordination of the histidine-[4Fe-4S]-cluster (iron in brown, sulphur in yellow) from the bidirectional [NiFe]-hydrogenase (in cyan, PDB: 5XF9), [FeFe]-hydrogenase (in blue, PDB: 6N59), complex I (in red, PDB: 4HEA) and formate dehydrogenase (in purple, PDB: 6TG9) indicate a specific coordination through the  $\delta$ -nitrogen (in maroon) for the [NiFe]-hydrogenases. All other examined bioenergetic proteins exhibit coordination through the  $\epsilon$ -nitrogen.

Asterisk (\*) denotes the sequences of the structural representative:

Hyd th - *Hydrogenophilus thermoluteolus TH-1*  
 Syn sp - *Synechocystis sp. PCC 6803*  
 Clo pa - *Clostridium pasteurianum DSM 525*  
 Des fr - *Desulfovibrio fructosovorans DSM 3604*  
 The th - *Thermus thermophilus HB8*  
 Esc co - *Escherichia coli strain M19*  
 Rho ca - *Rhodobacter capsulatus YW2*  
 Got ac - *Gottschalkia acidurici 9a*



Supplementary Figure 4a: **Fully annotated Neighbor-Joining Tree of the Y-junction module in representative sequences.**

Similarity clustering of the sequences belonging to the [FeFe]-hydrogenase (in blue), complex I (in red), [NiFe]-hydrogenase (in cyan) and formate dehydrogenase (in purple) as in Figure 5a. Minimal subunit and cofactor composition of the proteins including the [2Fe-2S]-containing subunit ([2Fe-2S]-SU), flavin/[4Fe4S]-binding subunit (FBSU), as well as specific subunits for all members of each protein family are shown as pictograms in the colored square. Additional specific cofactors characterising a specific group inside each protein family can be found as small pictograms at the representative leaves of the tree. Symbols enclosed by parentheses indicate the presence of the cofactor and/or the binding domain of that cofactor in the majority of sequences of the respective group. The symbols used are the same throughout figures and tables. Numbers in front of each species name refer to Supplementary Table 3, where RefSeq accessions and additional information to each sequence are given. Grey clouds and grey-colored branches indicate bootstrap values below 50%. Additional information include:

\* - representative sequences that were used for the structural alignment

4 - tetrameric Hnd/Fds

2 - dimeric Hnd/Fds

Experimentally studied protein complexes as indicated in bold in Supplementary Fig. 4a and 4b:

### **[NiFe]-hydrogenase (in cyan)**

*Cupriavidus necator* H16 (**4-Cup ne**) [24, 25, 26]  
*Hydrogenophilus thermoluteolus* TH-1 (**6-Hyd th\***) [27, 28]  
*Thiocapsa roseopersicina* strain DSM 217 (**7-Thi ro**) [29]  
*Synechocystis* sp. PCC 6803 (**20-Syn sp\***) [30, 31, 32]

### **[FeFe]-hydrogenase(in blue)**

*Acetobacterium woodii* DSM 1030 (**28-Ace wo<sup>4</sup>**) [4]  
*Moorella thermoacetica* ATCC 39073 (**33-Mor th**) [33]  
*Ruminococcus albus* DSM 20455 (**34-Rum al**) [34]  
*Thermotoga maritima* DSM 3109 (**42-The ma**) [35, 3]  
*Syntrophomonas wolfei* subsp. *wolfei* str. Goettingen G311 (**50-Syn wo**) [36]  
*Desulfovibrio fructosovorans* DSM3604 (**60-Des fr<sup>4</sup>**) [2]  
*Syntrophus aciditrophicus* SB (**151-Syn ac<sup>2</sup>**) [37]

### **Complex I (in red)**

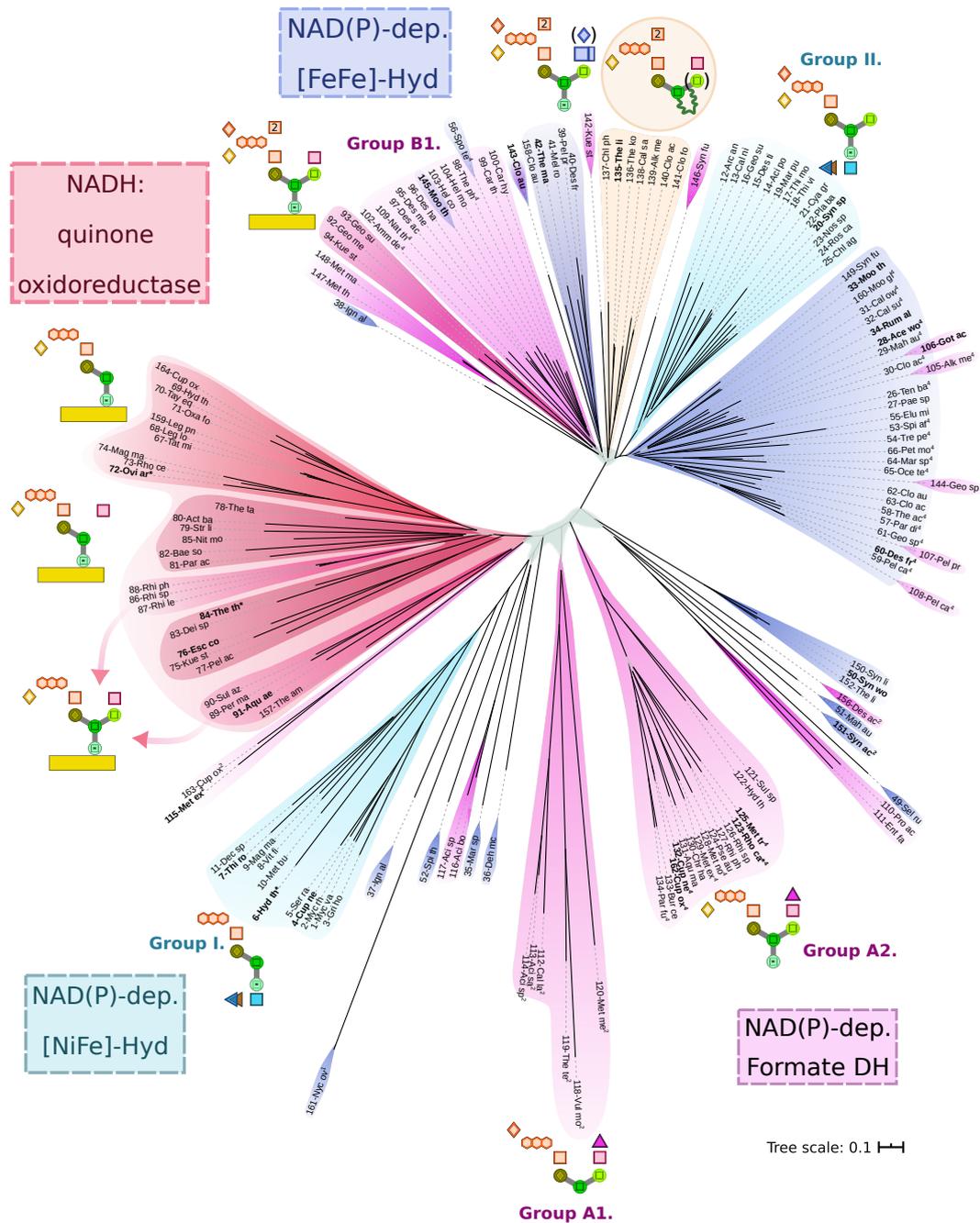
*Ovis aries* (**72-Ovi ar\***) [38, 39]  
*Escherichia coli* strain M9 (**76-Esc co**) [40]  
*Thermus thermophilus* DSM 579 (**84-The th\***) [41, 42, 16, 43, 44]  
*Aquifex aeolicus* VF5 (**91-Aqu ae**) [45, 46]

### **Formate Dehydrogenase (in purple)**

*Gottschalkia acidurici* 9a (**106-Got ac**) [7]  
*Methylobacterium extorquens* AM1 (**115-Met ex<sup>2</sup>**) [47]  
*Rhodobacter capsulatus* YW2 (**123-Rho ca\*<sup>4</sup>**) [48, 9]  
*Methylosinus trichosporium* OB3b (**125-Met tr<sup>4</sup>**) [49]  
*Cupriavidus necator* N-1 (**132-Cup ne<sup>4</sup>**) [50, 51, 8]  
*Cupriavidus oxalaticus* strain T2 (**162-Cup ox<sup>4</sup>**) [52]

### **Other proteins**

*Thermococcus litoralis* DSM 5473 (**135-The li**) [53]  
*Clostridium autoethanogenum* DSM 10061 (**143-Clo au**) [6]  
*Moorella thermoacetica* ATCC 39073 (**145-Moo th**) [54]



Supplementary Figure 4b: **Fully annotated Neighbor-Joining Tree of the flavin module in representative sequences.**

Similarity clustering of the sequences belonging to the [FeFe]-hydrogenase (in blue), complex I (in red), [NiFe]-hydrogenase (in cyan) and formate dehydrogenase (in purple) as in Figure 5b. Cofactor composition of the enzymes are shown as pictograms as in Figure 4a. Additional specific cofactors characterising a specific group inside each protein family can be found as small pictograms at the representative leaves of the tree. Symbols enclosed by parentheses indicate the presence of the cofactor and/or the binding domain of that cofactor in the majority of sequences of the respective group. The symbols used are the same throughout figures and tables. Numbers in front of each species name refer to Supplementary Table 3, where RefSeq accessions and additional information to each sequence are given. Grey clouds and grey-colored branches indicate bootstrap values below 50%. Additional information include:

\* - representative sequences that were used for the structural alignment

4 - tetrameric Hnd/Fds

2 - dimeric Hnd/Fds

## References

- [1] C. Baffert, A. Kpebe, L. Avilan & M. Brugna (2019); ‘Hydrogenases and H<sub>2</sub> metabolism in sulfate-reducing bacteria of the *Desulfovibrio* genus’. In *Advances in microbial physiology*, vol. 74, pp. 143–189 (Elsevier).
- [2] A. Kpebe, M. Benvenuti, C. Guendon, A. Rebai, V. Fernandez et al. (2018); ‘A new mechanistic model for an O<sub>2</sub>-protected electron-bifurcating hydrogenase, Hnd from *Desulfovibrio fructosovorans*’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1859(12):pp. 1302–1312.
- [3] N. Chongdar, K. Pawlak, O. Rüdiger, E. J. Reijerse, P. Rodríguez-Maciá et al. (2019); ‘Spectroscopic and biochemical insight into an electron-bifurcating [FeFe] hydrogenase’. *JBIC Journal of Biological Inorganic Chemistry*, pp. 1–15.
- [4] K. Schuchmann & V. Müller (2012); ‘A bacterial electron-bifurcating hydrogenase’. *Journal of Biological Chemistry*, vol. 287(37):pp. 31165–31171.
- [5] A. Graentzdoerffer, D. Rauh, A. Pich & J. R. Andreesen (2003); ‘Molecular and biochemical characterization of two tungsten- and selenium-containing formate dehydrogenases from *Eubacterium acidaminophilum* that are associated with components of an iron-only hydrogenase’. *Archives of microbiology*, vol. 179(2):pp. 116–130.
- [6] S. Wang, H. Huang, J. Kahnt, A. P. Mueller, M. Kopke et al. (2013); ‘NADP-specific electron-bifurcating [FeFe]-hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO’. *Journal of Bacteriology*, vol. 195(19):pp. 4373–4386.
- [7] S. Wang, H. Huang, J. Kahnt & R. K. Thauer (2013); ‘*Clostridium acidurici* electron-bifurcating formate dehydrogenase’. *Appl. Environ. Microbiol.*, vol. 79(19):pp. 6176–6179.
- [8] R. Hille, T. Young, D. Nicks, S. Hakopian, T. K. Tam et al. (2020); ‘Structure: Function studies of the cytosolic, Mo- and NAD<sup>+</sup>-dependent formate dehydrogenase from *Cupriavidus necator*’. *Inorganics*, vol. 8(7):p. 41.
- [9] C. Radon, G. Mittelstädt, B. R. Duffus, J. Bürger, T. Hartmann et al. (2020); ‘Cryo-EM structures reveal intricate Fe-S cluster arrangement and charging in *Rhodobacter capsulatus* formate dehydrogenase’. *Nature communications*, vol. 11(1):pp. 1–9.
- [10] K. Björklöf, V. Zickermann & M. Finel (2000); ‘Purification of the 45 kDa, membrane bound NADH dehydrogenase of *Escherichia coli* (NDH-2) and analysis of its interaction with ubiquinone analogues’. *FEBS Letters*, vol. 467(1):pp. 105–110.
- [11] T. Friedrich & B. Böttcher (2004); ‘The gross structure of the respiratory complex I: a Lego System’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1608(1):pp. 1–9.
- [12] T. Friedrich & D. Scheide (2000); ‘The respiratory complex I of bacteria, archaea and eukarya and its module common with membrane-bound multisubunit hydrogenases’. *FEBS letters*, vol. 479(1-2):pp. 1–5.
- [13] T. G. Laughlin, D. F. Savage & K. M. Davies (2018); ‘Revealing the subunit architecture of NAD(P)H dehydrogenase type-1 from cyanobacteria through Cryo-EM’. *Biophysical Journal*, vol. 114(3):p. 424a.
- [14] J. M. Schuller, J. A. Birrell, H. Tanaka, T. Konuma, H. Wulfhorst et al. (2019); ‘Structural adaptations of photosynthetic complex I enable ferredoxin-dependent electron transfer’. *Science*, vol. 363(6424):pp. 257–260.
- [15] H. Brüggemann, F. Falinski & U. Deppenmeier (2000); ‘Structure of the F<sub>420</sub>H<sub>2</sub>: quinone oxidoreductase of *Archaeoglobus fulgidus*: Identification and overproduction of the F<sub>420</sub>H<sub>2</sub>-oxidizing subunit’. *European Journal of Biochemistry*, vol. 267(18):pp. 5810–5814.
- [16] R. Baradaran, J. M. Berrisford, G. S. Minhas & L. A. Sazanov (2013); ‘Crystal structure of the entire respiratory complex I’. *Nature*, vol. 494(7438):pp. 443–448.
- [17] B. Schoepp-Cothenet, R. Van Lis, P. Philippot, A. Magalon, M. J. Russell et al. (2012); ‘The ineluctable requirement for the trans-iron elements molybdenum and/or tungsten in the origin of life’. *Scientific reports*, vol. 2:p. 263.
- [18] J. Rozewicki, S. Li, K. M. Amada, D. M. Standley & K. Katoh (2019); ‘MAFFT-DASH: integrated protein sequence and structural alignment’. *Nucleic acids research*, vol. 47(W1):pp. W5–W10.
- [19] B. C. Marreiros, A. P. Batista, A. M. Duarte & M. M. Pereira (2013); ‘A missing link between com-

- plex I and group 4 membrane-bound [NiFe] hydrogenases'. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1827(2):pp. 198–209.
- [20] H. Brinkmann & H. Philippe (1999); 'Archaea sister group of Bacteria? Indications from tree reconstruction artifacts in ancient phylogenies.' *Molecular biology and evolution*, vol. 16(6):pp. 817–825.
- [21] J. Mock, Y. Zheng, A. P. Mueller, S. Ly, L. Tran et al. (2015); 'Energy conservation associated with ethanol formation from h<sub>2</sub> and co<sub>2</sub> in clostridium autoethanogenum involving electron bifurcation'. *Journal of bacteriology*, vol. 197(18):pp. 2965–2980.
- [22] K. Schuchmann, N. P. Chowdhury & V. Müller (2018); 'Complex multimeric [FeFe] hydrogenases: biochemistry, physiology and new opportunities for the hydrogen economy'. *Frontiers in microbiology*, vol. 9:p. 2911.
- [23] X. Robert & P. Gouet (2014); 'Deciphering key features in protein structures with the new endscript server'. *Nucleic acids research*, vol. 42(W1):pp. W320–W324.
- [24] A. Tran-Betcke, U. Warnecke, C. Böcker, C. Zaborosch & B. Friedrich (1990); 'Cloning and nucleotide sequences of the genes for the subunits of NAD-reducing hydrogenase of *Alcaligenes eutrophus H16*'. *Journal of bacteriology*, vol. 172(6):pp. 2920–2929.
- [25] L. Lauterbach, Z. Idris, K. A. Vincent & O. Lenz (2011); 'Catalytic properties of the isolated diaphorase fragment of the NAD<sup>+</sup>-reducing [NiFe]-hydrogenase from *Ralstonia eutropha*'. *PloS one*, vol. 6(10).
- [26] L. Lauterbach, J. Liu, M. Horch, P. Hummel, A. Schwarze et al. (2011); 'The hydrogenase subcomplex of the NAD<sup>+</sup>-reducing [NiFe] hydrogenase from *Ralstonia eutropha* – insights into catalysis and redox interconversions'. *European Journal of Inorganic Chemistry*, vol. 2011(7):pp. 1067–1079.
- [27] Y. Shomura, M. Taketa, H. Nakashima, H. Tai, H. Nakagawa et al. (2017); 'Structural basis of the redox switches in the NAD<sup>+</sup>-reducing soluble [NiFe]-hydrogenase'. *Science*, vol. 357(6354):pp. 928–932.
- [28] J. Preissler, S. Wahlefeld, C. Lorent, C. Teutloff, M. Horch et al. (2018); 'Enzymatic and spectroscopic properties of a thermostable [NiFe]-hydrogenase performing H<sub>2</sub>-driven NAD<sup>+</sup>-reduction in the presence of O<sub>2</sub>'. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1859(1):pp. 8–18.
- [29] G. Rákhely, A. T. Kovács, G. Maróti, B. D. Fodor, G. Csanádi et al. (2004); 'Cyanobacterial-type, heteropentameric, NAD<sup>+</sup>-reducing NiFe hydrogenase in the purple sulfur photosynthetic bacterium *Thiocapsa roseopersicina*'. *Appl. Environ. Microbiol.*, vol. 70(2):pp. 722–728.
- [30] O. Schmitz, G. Boison, H. Salzmann, H. Bothe, K. Schütz et al. (2002); 'HoxE—a subunit specific for the pentameric bidirectional hydrogenase complex (HoxEFUYH) of cyanobacteria'. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1554(1-2):pp. 66–74.
- [31] F. Germer, I. Zebger, M. Saggu, F. Lenzian, R. Schulz et al. (2009); 'Overexpression, isolation, and spectroscopic characterization of the bidirectional [NiFe] hydrogenase from *Synechocystis sp. PCC 6803*'. *Journal of Biological Chemistry*, vol. 284(52):pp. 36462–36472.
- [32] K. Gutekunst, X. Chen, K. Schreiber, U. Kaspar, S. Makam et al. (2014); 'The bidirectional NiFe-hydrogenase in *Synechocystis sp. PCC 6803* is reduced by flavodoxin and ferredoxin and is essential under mixotrophic, nitrate-limiting conditions'. *Journal of Biological Chemistry*, vol. 289(4):pp. 1930–1937.
- [33] S. Wang, H. Huang, J. Kahnt & R. K. Thauer (2013); 'A reversible electron-bifurcating ferredoxin-and NAD-dependent [FeFe]-hydrogenase (HydABC) in *Moorella thermoacetica*'. *Journal of bacteriology*, vol. 195(6):pp. 1267–1275.
- [34] Y. Zheng, J. Kahnt, I. H. Kwon, R. I. Mackie & R. K. Thauer (2014); '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'. *Journal of bacteriology*, vol. 196(22):pp. 3840–3852.
- [35] G. J. Schut & M. W. W. Adams (2009); 'The Iron-Hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production'. *Journal of Bacteriology*, vol. 191(13):pp. 4451–4457.
- [36] N. A. Losey, F. Mus, J. W. Peters, H. M. Le & M. J. McInerney (2017); '*Syntrophomonas wolfei* uses an NADH-dependent, ferredoxin-independent [FeFe]-hydrogenase to reoxidize NADH'. *Applied and environmental microbiology*, vol. 83(20).

- [37] N. A. Losey, S. Poudel, E. S. Boyd & M. J. McInerney (2020); ‘The beta subunit of non-bifurcating NADH-dependent [FeFe]-hydrogenases differs from those of multimeric electron-bifurcating [FeFe]-hydrogenases’. *Frontiers in microbiology*, vol. 11:p. 1109.
- [38] J. A. Letts, G. Degliesposti, K. Fiedorczuk, M. Skehel & L. A. Sazanov (2016); ‘Purification of ovine respiratory complex I results in a highly active and stable preparation’. *Journal of Biological Chemistry*, vol. 291(47):pp. 24657–24675.
- [39] J. A. Letts, K. Fiedorczuk, G. Degliesposti, M. Skehel & L. A. Sazanov (2019); ‘Structures of respiratory supercomplex I+III<sub>2</sub> reveal functional and conformational crosstalk’. *Molecular cell*, vol. 75(6):pp. 1131–1146.
- [40] T. Friedrich (1998); ‘The NADH: ubiquinone oxidoreductase (complex I) from *Escherichia coli*’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1364(2):pp. 134–146.
- [41] T. Yagi, K. Hon-Nami & T. Ohnishi (1988); ‘Purification and characterization of two types of NADH-quinone reductase from *Thermus thermophilus* HB-8’. *Biochemistry*, vol. 27(6):pp. 2008–2013.
- [42] L. A. Sazanov & P. Hinchliffe (2006); ‘Structure of the hydrophilic domain of respiratory complex I from *Thermus thermophilus*’. *science*, vol. 311(5766):pp. 1430–1436.
- [43] R. G. Efremov, R. Baradaran & L. A. Sazanov (2010); ‘The architecture of respiratory complex I’. *Nature*, vol. 465(7297):pp. 441–445.
- [44] R. G. Efremov & L. A. Sazanov (2012); ‘The coupling mechanism of respiratory complex I – a structural and evolutionary perspective’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1817(10):pp. 1785–1795.
- [45] M. Guiral, L. Prunetti, S. Lignon, R. Lebrun, D. Moinier et al. (2009); ‘New insights into the respiratory chains of the chemolithoautotrophic and hyperthermophilic bacterium *Aquifex aeolicus*’. *Journal of proteome research*, vol. 8(4):pp. 1717–1730.
- [46] G. Peng, B. Meyer, L. Sokolova, W. Liu, S. Bornemann et al. (2018); ‘Identification and characterization two isoforms of NADH: ubiquinone oxidoreductase from the hyperthermophilic eubacterium *Aquifex aeolicus*’. *Biochimica et Biophysica Acta (BBA)-Bioenergetics*, vol. 1859(5):pp. 366–373.
- [47] M. Laukel, L. Chistoserdova, M. E. Lidstrom & J. A. Vorholt (2003); ‘The tungsten-containing formate dehydrogenase from *Methylobacterium extorquens* AM1: Purification and properties’. *European journal of biochemistry*, vol. 270(2):pp. 325–333.
- [48] T. Hartmann & S. Leimkühler (2013); ‘The oxygen-tolerant and NAD<sup>+</sup>-dependent formate dehydrogenase from *Rhodobacter capsulatus* is able to catalyze the reduction of CO<sub>2</sub> to formate’. *The FEBS journal*, vol. 280(23):pp. 6083–6096.
- [49] D. R. Jollie & J. D. Lipscomb (1991); ‘Formate dehydrogenase from *Methylosinus trichosporium* OB3b. purification and spectroscopic characterization of the cofactors.’ *Journal of Biological Chemistry*, vol. 266(32):pp. 21853–21863.
- [50] J.-I. Oh & B. Bowien (1998); ‘Structural analysis of the Fds operon encoding the NAD<sup>+</sup>-linked formate dehydrogenase of *Ralstonia eutropha*’. *Journal of Biological Chemistry*, vol. 273(41):pp. 26349–26360.
- [51] T. Young, D. Nix, S. Hakopian, T. K. Tam, X. Yu et al. (2020); ‘Crystallographic and kinetic analyses of the FdsBG subcomplex of the cytosolic formate dehydrogenase FdsABG from *Cupriavidus necator*’. *Journal of Biological Chemistry*, vol. 295(19):pp. 6570–6585.
- [52] U. Ruschig, U. Müller, P. Willow & T. Höpner (1976); ‘CO<sub>2</sub> reduction to formate by NADH catalysed by formate dehydrogenase from *Pseudomonas oxalaticus*’. *European journal of biochemistry*, vol. 70(2):pp. 325–330.
- [53] A. Tóth, M. Takács, G. Groma, G. Rákhely & K. L. Kovács (2008); ‘A novel NADPH-dependent oxidoreductase with a unique domain structure in the hyperthermophilic archaeon, *Thermococcus litoralis*’. *FEMS microbiology letters*, vol. 282(1):pp. 8–14.
- [54] J. R. Andreesen & L. G. Ljungdahl (1974); ‘Nicotinamide adenine dinucleotide phosphate-dependent formate dehydrogenase from *Clostridium thermoaceticum*: purification and properties’. *Journal of bacteriology*, vol. 120(1):pp. 6–14.