Supporting Information

Secondary structure changes as the potential H_2 sensing mechanism of Group D [FeFe]-hydrogenases

Ivan Voloshyn[a], Conrad Schumann[b], Princess R. Cabotaje[b], Afridi Zamader[b,†], Henrik Land[b] and Moritz Senger[a,b]*

a) Department of Chemistry – BMC, Biochemistry, Uppsala University, 75120 Uppsala, Sweden

b) Department of Chemistry – Ångström Laboratory, Molecular Biomimetics, Uppsala University, 75120 Uppsala, Sweden

†) Current address: Laboratoire d'Electrochimie Moléculaire (LEM), Université Paris Cité, CNRS, F-75006, Paris, France.

*To whom correspondence should be addressed

moritz.senger@kemi.uu.se

Contents

Materials and Methods

In silico genomic analysis of region

The genomic region containing *Tam*HydS and a gene encoding Group A [FeFe]-hydrogenase was analysed for operon structure, promoters, terminators, as well as ribosome binding sites (RBS). To determine bacterial operons, an **Operon-Mapper** prediction tool was used.[1] For the prediction of putative promoters and terminators, tools **BPROM** and **FindTerm** were used.[2] For ribosome binding site prediction, the **RBS Calculator** from DeNovoDNA was used (https://salislab.net/software/predict_rbs_calculator).
The results of each prediction were manually aligned with the genome sequence to choose the best suitable options among predicted.

Attenuated Total Reflectance Fourier-Transform Infrared (ATR-FTIR) spectroscopy

A solution of 2 µL enzyme (0.375−1 mM *Tam*HydS) in 10 mM Tris buffer (pH 8.0), 3 µl Eosin Y (6mM) and 0.5µl 100 mM Ethylenediaminetetraacetic acid (EDTA) was deposited on the ATR crystal in the anaerobic atmosphere of a Braun Glove box. The ATR unit (BioRadII from Harrick) was sealed with a custom-built PEEK cell that allowed for gas exchange and illumination (similar to Stripp 2021[3] and Senger et al., 2016[4]) mounted in a FTIR spectrometer (Vertex V70v, Bruker). Illumination for photo-reduction experiments was facilitated via a Schott KL2500 lamp optically coupled to the ATR crystal surface with fibre optics. The sample was dried under 100% nitrogen gas and rehydrated with a humidified aerosol (100 mM Tris-HCl, pH 8) as described before.[5] For H_2 exposure experiments, a mixed buffer (pH 8, 50mM Tris-HCl, 50 mM Citric acid) was used. Spectra were recorded with 2 cm^{-1} resolution, a scanner velocity of 80 Hz and averaged of varying number of scans (mostly 1000 Scans). All measurements were performed at ambient conditions (room temperature and pressure, hydrated enzyme films). Gases (N_2, H_2) were applied at a flow rate of 0.5-1.5 L/min. The data was analysed and plotted to our protocols described previously.[6, 7]

Enzyme purification

Preparation of *Tam*HydS^{ADT}

The preparation of *Tam*HydS^{ADT} was performed as previously reported with minor changes to the procedure.[8] For the expression of the apo-form, sequence-confirmed plasmids were transformed in chemically competent *E. coli* BL21(DE3) cells. The protein expression was induced at $O.D._{600} \approx 0.5$ with 1 mM IPTG with concomitant supplementation of the culture with 100 µM $FeSO_4$ in 1% HCl solution. The harvested cell pellets were lysed by 3 cycles of freezing/thawing in liquid N_2 in 100 mM Tris-HCl, 150 mM NaCl pH 8.0 supplemented with 1 g/L lysozyme, 0.05 g/L DNase, 0.05 g/L RNase, 2 g/L $MgCl_2$ * 6 H_2O and cOmplete™ Protease Inhibitor Cocktail (Roche). The cell lysis, protein purification, the reconstitution of the [4Fe-4S] clusters, as well as the activation of the enzyme were carried out in an MBRAUN glovebox under argon atmosphere (app. 1 ppm O_2). The protein was purified using StrepTrap affinity chromatography (StrepTrap XT (GE Healthcare)) following the manufacturer's instructions and applying an additional washing step with 1 M urea in 100 mM Tris-HCl, 150 mM NaCl pH 8.0. After purification, the protein yield was 2.8 mg L^{-1} of cell culture with an iron/protein content of 11.8 Fe/protein. By incubating the enzyme (50 µM) in a reaction with 500 µM dithiothreitol, 500 nM cysteine desulferase (*E. coli* IscS), 525 µM L-cysteine and 525 µM $(NH_4)_2Fe(SO_4)_2(H_2O)_6$ in 100 mM Tris-HCl, 150 mM NaCl pH 8.0 all [4Fe-4S] clusters were fully reconstituted (17.0 ± 1.4 Fe/protein). Subsequently, the yielded apo-form of *Tam*HydS (200 µM) was semi-artificially activated by incubating it for 30 min with 4 mM sodium dithionite and 2.4 mM of of [2Fe]^{ADT} in 100 mM Tris-HCl, 150 mM NaCl pH 8.0. Eventually the activation reaction was desalted using 10 mM Tris-HCl pH 8.0 containing 5 mM NaDT. The generated holo-form of [2Fe]^{ADT}-activated *Tam*HydS aliquots were concentrated to 2.0 mM.

Aliquots were prepared in air-tight vials, flash-frozen in liquid N_2 and stored at −80 °C. The successful activation and the correct cofactor integration were verified by ATR-FTIR analyses.

The expression and purification procedures were further optimized for improved yields and protein purity. For the cell lysis, the above described lysis buffer was supplemented with 5 g/L sodium deoxycholate and 50 g/L Succrose. Additionally, the freeze-thaw-cycles were replaced by sonication. During the purification an additional washing step with 5 mM ATP and 10 mM $MgCl_2$ in 100 mM Tris-HCl, 150 mM NaCl pH 8.0 was applied to remove contaminant Hsp70 molecular chaperones which replaced the 1M urea washing step.[9, 10] The generated holo-form of $[2Fe]^{ADT}$-activated *Tam*HydS was concentrated to 375 µM. As before, aliquots were prepared in air-tight vials, flash-frozen in liquid N_2 and stored at −80 °C. The successful activation and the correct cofactor integration were verified by ATR-FTIR spectroscopic analysis.

Synthetic chemistry

The synthesis of $(Et_4N)_2[Fe_2(\mu\text{-ADT})(CO)_4(CN)_2]$[11] was performed using previously reported protocols, and purity verified using a combination of FTIR, NMR and UV/Vis spectroscopy.

Table S1-3: Promoters, terminators and ribosome binding sites.

The output from prediction tools was manually aligned with the genomic sequence to annotate the region surrounding *Tam*HydS. The promoter and terminator results with the highest scores are provided in the Table S1 and S2, respectively. Ribosome binding sites are listed in Table S3.

Table S1 Promoters:

Name	Position	-35 sequence	-10 sequence	LDF score[1]
Promoter 1	900768 - 900789	TTTAAA	TAGTATAAT	7.34
Promoter 2	904809 - 904840	TTGTTA	TGATATAAT	10.46

[1]Linear discriminant function (LDF) score provides a measurement of promoter strength.

Table S2 Terminators:

Name	Chain	Length (nt)	Sequence	Score
Terminator 1	+	35	uccuugcccaaccccuucacccuuucaaaGUAACCUCA caucaaacucucuuuuuaaaau	-14.5
Terminator 2	+	50	Uucuauucgaacaaagucuaccgcuucaggauuaucccuaucu gcaauaguuucuccaaucucaauaucgcuuAUGCCAGAAac ugcuacaauaucuccuaaacuugcuucuucuacuug	-21.9

Table S3 Ribosome binding site:

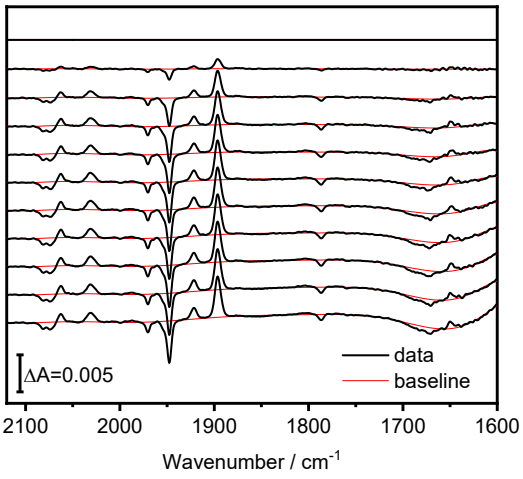
Ribosome binding site number	**Position**	**Sequence**
859	900797	UAAAGG
860	901604	GUUAGG
862	902382	AGUAGG
863	903695	AGGAGU
865	904903	AGGAGG
868	906326	AGGAGG
869	908141	AGGAGU

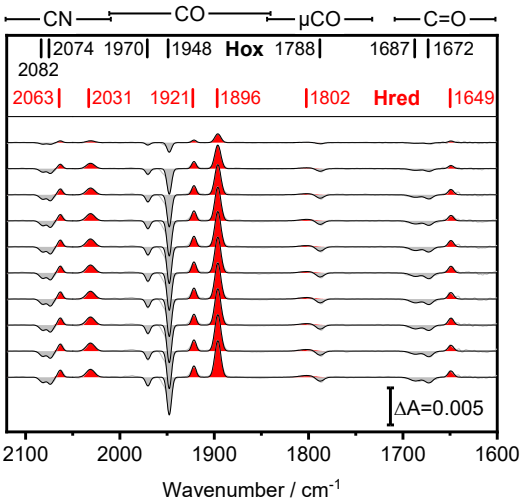
Table S4 Infrared band patterns of different redox states in *Tam*HydS (Group D) and *Cr*HydA1 (Group A) [FeFe]-hydrogenase

	*Tam*HydS (Group D) bands in cm^{-1}	*Cr*HydA1 (Group A) bands in cm^{-1}
Hox	2082, 2074, 1970, 1948, 1787[8]	2088, 2072, 1964, 1940, 1800[12]
Hred (H_2)	2062, 2030, 1961, 1922, 1895[8] 2063, 2032, 1922, 1996, 1801 (this study)	2072, 2033, 1961, 1915, 1891[13] At 90 K, n.d, n.d., 1919, 1891, 1817[14]
Hred (photo-reduction)	2063, 2032, 1922, 1895, 1803 (this study)	2072, 2033, 1961, 1915, 1891[15]

Bands are indicated in the order CN, CN, tCO, tCO, μ/tCO

Figure S1: Redox state specific band population over the time course of the photo-reduction experiment





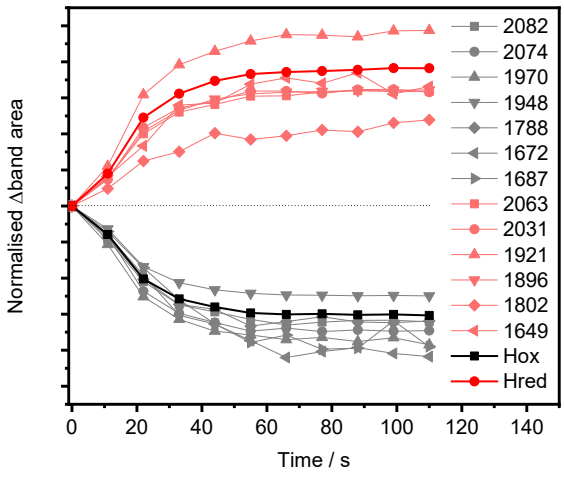


Figure S1: (top) Visualization of the baseline subtraction process. Data in black, Baseline in red. (middle) Baseline subtracted difference spectra with redox state specific bands highlighted (bottom) Plotting the band area of redox specific peaks over time highlights the co-population of bands associated with one redox state. Note that band areas are normalized to allow for better visual comparison.

Figure S2: ATR-FTIR Hred-Hox difference spectra of *Tam*HydS (Group D) and CrHydA1 (Group A) [FeFe]-hydrogenases

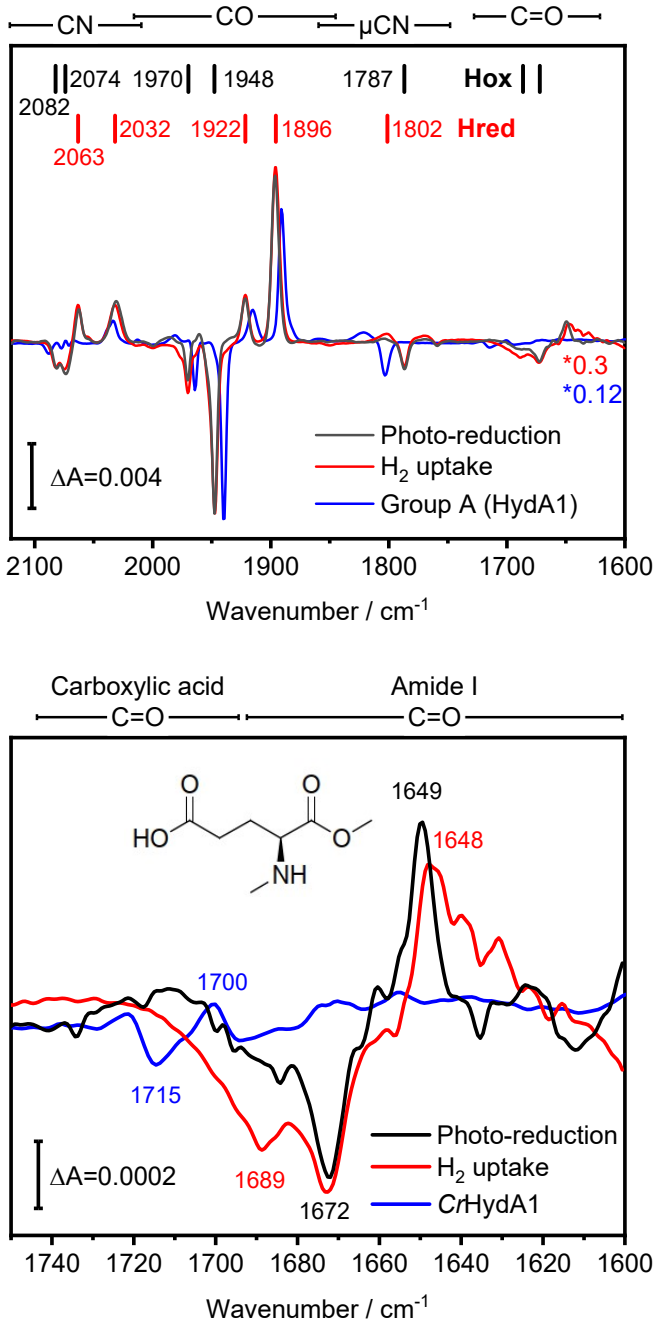


Figure S2: (top) Upon exposure of *Tam*HydS to either H_2 (red spectrum) or illumination of *Tam*HydS in the presence of Eosin Y and EDTA as sacrificial electron donor (black spectrum, photo-reduction) we observe the population of the reduced state Hred (Positive peaks: 2063, 2032, 1922, 1896, 1802 cm-1) and a de-population of Hox (negative peaks: 2082, 2074, 1970, 1948, 1787 cm-1). The same transition from Hox to Hred was observed in *Cr*HydA1 (blue spectrum) induced by the photoreduction protocol (peak positions are shifted due to the different [FeFe]-hydrogenase). The respective vibrational regions are indicated above the graph. Band positions are indicated by coloured bars.

(bottom) Zoom in on the C=O region. We observe a large difference feature in the Amide I C=O region with a positive peak at 1649 cm-1 and a negative peak at 1672 cm-1 in the Hred-Hox difference spectrum of TamHydS induced via photo-reduction (black) or induced by H_2 exposure (red) which we assign to a change in secondary structure. In the Group A [FeFe]-hydrogenase *Cr*HydA1 (blue spectrum) no changes in the Amide I region are detected in the difference spectrum of the Hox to Hred transition. Instead we detected a difference feature in the carboxylic acid C=O region (1715 and 1700 cm-1) which indicates a rearrangement of the hydrogen bonding network of the PTP absent in TamHydS. The respective vibrational regions are indicated above the graph.

The H_2 uptake spectrum of *Tam*HydS is scaled by 0.3 and the photoreduction spectrum of *Cr*HydA1 by a factor of 0.12 for better comparability. *Cr*HydA1 data modified from ref[15].

Figure S3: H/D effect on the ATR-FTIR Hred-Hox difference spectra of *Tam*HydS

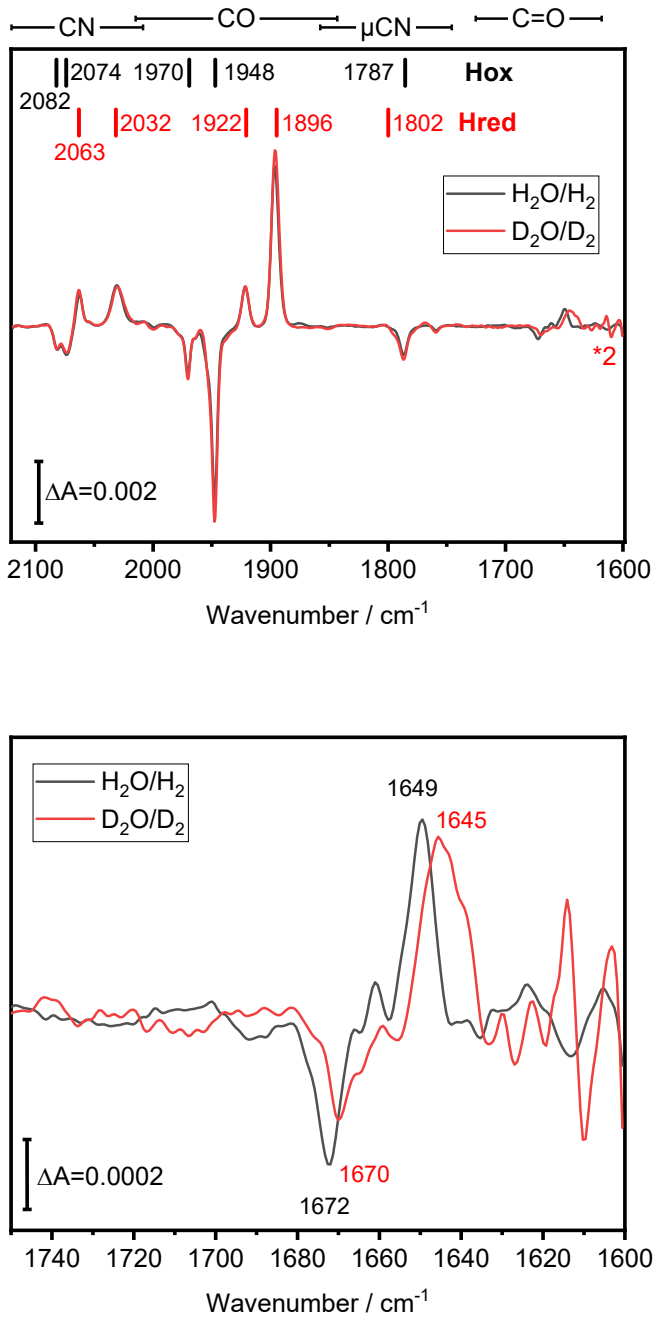


Figure S3: (top) The transition from Hox (negative bands) to Hred (positive bands) was induced either in H_2O via H_2 (black spectrum) or in D_2O via D_2 (red spectrum). The difference spectra look overall similar but exhibit slight differences in the C=O (Amide I) region. (Bottom) Zoom in on the C=O (Amide I) region for the transition shown in the top. We detected a down-shift of the negative band at 1672 to 1670 cm-1 and of the positive band at 1649 to 1645 cm-1 upon deuteration.
The D_2O/D_2 spectrum is scaled by a factor of two to allow for better visual comparison.

Fit parameters Fig.S1

Fit parameters used to analyse and plot the data as described before.[6, 7] In short band position and band width (Full Width Half Max (FWHM)) for each redox state are fitted to sets of spectra. In some cases, the relative amplitudes of the peaks related to one redox state are derived from the fit.

*Tam*HydS^{ADT} Fig.S1

Band position	**FWHM**	**Redox state**	
2081,73277048695	-	7,81297001312315	Hox
2073,57909496923	-	6,51261389374464	Hox
1970,30216362434	-	5,72919867410981	Hox
1947,68473038479	-	5,64983595248925	Hox
1787,96633434964	-	9,57488904726338	Hox
2063,28761581669	-	5,76870662850313	Hred
2031,19703921083	-	10,4127838002636	Hred
1921,48715276992	-	5,52050765083911	Hred
1896,23423787334	-	7,53581753427441	Hred
1801,6632547387	-	18	Hred
1672,42617114634	-	9,60741974551811	Hox
1649,09320466939	-	6,82534955653803	Hred
1686,55425474694	-	13,1130249689038	Hox

Annotated Thermoanaerobacter mathranii genome

***Thermoanaerobacter mathranii* genome at position 897,980 – 910,900**

NCBI database: Thermoanaerobacter mathranii subsp. mathranii str. A3, complete genome (https://www.ncbi.nlm.nih.gov/nuccore/CP002032.1?report=graph&from=902393&to=904300)

Gene sequences marked-up with the starting codon underlined

ACCTTATGAAATTTGTAAGAGAAGATGTAAGAAATATAGCCATTATTGCTCATGTTGATCATGGAAA
AACTACTTTGGTTGATGCTATGCTTAAACAAAGCGGTATTTTTAGGTCAAATGAAAAAGTTGAGGAA
AGGATTTTGGACTTTAACGATTTAGAAAGAGAGAGAGGAATAACGATTCTTGCAAAAAATACTGCTA
TACGATATAAAGATGTGAAAATAAATATAGTGGATACCCCAGGACATGCGGATTTTAGCGGTGAAGT
AGAGCGAGTATTAAAAATGGTAGATGGGGTACTTTTGCTGGTGGATTCTTTCGAAGGACCTATGCCA
CAGACCCGTTTTGTGTTAAGCAAAGCTTTGGAACTAGATTTAAAACCTATAGTTGTTATAAATAAAAT
TGATAGACCTGATGCAAGGCCAGAGGAGGTTATTGACGAAGTCTTAGATTTATTTATTGAATTAGGA
GCAAATGATGACCAAATCGATTTTCCTGTGGTTTACACTTCTGCAAAAGAAGGTATAGCAAAATTAA
GCTTAGATGAAGAATCTCATGACTTGAGACCTCTTTTTGATACGATTTTAGAGTATATTCCTGCTCCCT
CAGGAAACATTGAAGCCCCCCTGCAACTTATAGTGACTACTTTAGATTATGATGATTACATTGGGAGA
ATTGCCATTGGAAAAATAGTTAGAGGGAAGATAATTTCAGGAGAAGAAGCGGCCATATGTAAAAGA
GACGGGTCGATACAAAAAGTTCACATAAATAATTTATATCAGTTTGAAGGACTAAAAAGAGTGCAAG
TAGAAGAAGCAAGTTTAGGAGATATTGTAGCAGTTTCTGGCATAAGCGATATTGAGATTGGAGAAA
CTATTGCAGATAGGGATAATCCTGAAGCGGTAGACTTTGTTCGAATAGAAGAGCCTACAGTTACTAT
GACTTTTAGTGTAAATACCAGTCCTTTTGCAGGAACCGAAGGTAAATACGTTACTTCAAGGCATTTAA
GAGAAAGGCTTTTTAAGGAGTTAGAGACAAACGTAGCGTTGCGTGTTGAAGAAACTGATTCTCCTGA
TTCTTTTAAAGTATCTGGAAGAGGAGAGTTACATCTTTCTATTTTAATTGAGACTATGAGGCGGGAAG
GGTATGAATTGCAGGTTTCTAAGCCTACAGTAATTTTTAAAGAGGAAAATGGAGTAAAGATGGAACC
TATTGAACTTTTGACAATAGATATCCCAGAAGAATATATGGGAGTTGTAATTGAAAAATTAGGGCCG
AGAAAAGCTGAGTTGATGGATATGCACACATTGAAACCGGGGACTGTAAGGCTTAAGTTTAAAATA
CCTACAAGAGGCTTGATAGGGTATCGTTCTGAATTTTTAACAGATACCAAAGGAAATGGGATAATGA
CTTCGGTATTTTATGATTATGAACCTTATAAAGGAGACATTCCTTCTCGGGGAAGAGGCGCTCTTGTA
GCTTTTGAGACAGGCGTTGCGACAACGTATGGGCTTTACAACGCGCAAGAAAGAGGAACACTTTTTA
TAGAGCCAGGTACAAAGGTATATGAAGGGATGGTTGTGGGGATTAATGCAAGAAGTGGAGATATT
GATGTAAATGTATGTAAGAAAAAACATGTGACAAATTTAAGGTCTGCGACGGCAGATGAGGCTTTG
AGGCTATCACCTATAAAAAAGATGTCGTTAGAAGAAGCATTAGAATTTATAGACAATGATGAATTAG
TTGAAGTTACTCCTCAAAGCATTAGAATACGAAAAAAGATTTTGGATTCTCAACAGAGGTATAAAAG
TGCAAAATATAACAAATAAAGGAGCTTTATGCATCCTTCAGCTTGTTGACAAAGTAAAAATTTTTTAAA
ATAGGATATTTTGCATCGTCACTCTGATGTTCCACAGCGCAAAACTAAGCTCCACCTTCGGGTTCCGGCA
GGGTCACGGGCACATTCGACCTCCTGTCTCAGGTGCCCGCCTCCGCCATCCTTGGCTTCGGCCCTGCCTC
CGCCCTCGGTCTCGCTAAGTTTTGTTAACGCTTTGTCACAAGTTGTGCCTTATGCAAAATATCCTATTTCC
GAAAGTTTGTATACAGTCTAAAGGAGCTTGATGGCTCCTTCTATTTATATAATTCATTTAAAATTTTGCT
TATATAATTTATAGTTTCTTGGTAAGGGGGAATACCTTGATATTTTTCTACTGATTGAGGACCAGCATT
ATATGCAGCTAATGCTAATCTTATGTCATGGTAAGTGTCTAGCAAATTTTTTAGGTATCTTACTCCTCC
ATCTATATTTTGAGAAGGGTCAAAAGGATTGGATACATTCAATTCCTTTGCAGTCGAAGGCATAAGCT
GCATAAGTCCCATTGCTCCTGCAGGAGAAACTGCGAAAGGATTAAAATTAGACTCTGCTTTGATAACT
GCTCTAATAAGGTTTGCATCTATATTATATTTTTGACTGGCCCGTTGTATTAAATCTTCAATTTGTTGTA
AATTTGTATTTGTTGATGATGCATTGTTACTTACATTTCCTGAATTGTATTGTTGATTTTCTATAGCATT
CTTTAAAACTTCGGCAAAAGGTTTATCTGTATTAAATATTTTTACATTAGCAGGGAGTCTACTTTGAAT
TTCTCTAAATTTATTTTCAATTATTTGATTTACAATATTTATCAATTCTACCCCTCCCATCATATATAATTC

857: TypA GTP-binding brotein

858: Lytic trans-glyocosylase

(here reverse complement of CDS with starting codon TTG)

GACAAAATTCCTCTAAAACCTCTTTTAAAAGTGTAAATTAGTAAAAAAGTCCTATTAAAATTTTTAAAGG
ATTTTTTAAAGGAGTATAGTATAATATAAGTAAAGGTTTATGGTAAGGGTTGATGAAGCAATGAAAAA
CGAAAGAATACTGGAAGTGTTAATTGTTGTATTAATAATTCCTCTGGTGTCTCTTTTAGCAGGTTATTTT
GTTTCAAAAGATATTATAGTGCCTAACATTTCAAAAACAAATGACAGCAATAATTTTAAAGAGATGGTT
GTTAAAGGCGTAGACTTGTTTGAAATTATTTTAGGCAGCTATTCAGATTTTGAGACTGCAAAATATCAA
GAAGATTTAATGAGGATGAAAAGAGTTTACTCTTTCGTTAGTAAATATTCTGATAAGTATTTGTTGATAG
GTGGTATATTTTTAGATAAAGAACAGGCTGCTTTATTTTCTTCCTCTTTAAAGTCTCAAGGAATTTCCAGT
GAAGTATATTTAAAAAGGGGACCTTCTCTCAGAATTAAATATGATAAGAAATTAACATCAGATATGGAT
GATTTTACAAGTAAATTACAAGATTTTAAACAAATTCTTGATTATATAGCTACTCTTTCTTATAAGGCTTA
TGATGACAAGTTAGAACGACAAGAATTGGAGGAATTAAAAAGGAAAGTAGATAGTTTTAAAAATGGA
AATGGGGATTTTGGAAAAAAAGAAATGACTGCACTAGTACAAAAGACGACAGATATTGTCGATAAAAT
AACAACTGATATTGAAAAAATTGAAATATCAGCTGCTTTAGAAGATGGAAATACATTTAGTTTGTTGCA
AGAAAGTTTATGGCAATCTTGTGAAGAATACAACGCTTTTTTGAAAACAATTGCTGTGCAGAATCAATG
ATGGGGTGTTAGGGTATGAGATTAAAAGAGGTAAAAGAAATTTTAAATGCAGAAGTTATGATAGGAG
AAGAAAAATTGGAAGAAGAAGTGTTTACTGCTTGTGGAGCGGACTTGATGAGTGATGTTTTGGCATCA
CGTGACGAAAAGGCTGTGCTTTTGACAGGGCTTACTAATGTGCAGGTTATAAGGACTGCAGAAGTAGT
TGGAGATATTAAATGTATAGTATTTGTAAGAGGTAAGAATCCGGGAGAGGACATATTGGAACTTGCTA
AAAAGGCAGGGCTAGTGATAATGAAAACGAAATATCCTCTTTACATTGCTTGTGGCCTTTTGTACTCCA
ATGGGCTTACTTGTAAAAGTGGTGAAGGAAATGGTACTTTATAGTTTAGATTTTGATGTAAAAGCAA
GAGATTTTGATTCTGCGGGAGAGGTTTCCAGCAATTTAAGGTCAGTATTAAAACAGCTTTCCTTAAAT
CCAGATGTTGTAAGGAGAAGTTGCATAGCTTGCTATGAAGCAGAGATGAATATAATAATTCATTCTT
ATGGAGGTCACATGAAAATAGAGATTTTTGGCGATAAAATACGCATAACTGCAGAAGACACAGGCC
CTGGAATAGAAGACATTGAACTTGCCATGAAAGAAGGATATTCTACTGCTCCTGAAGAGATAAGGG
AAATGGGATTTGGAGCCGGCATGGGGCTTCCTAATATGAAGGAGAATTCGGATTTCATGGATATAAA
AAGCAGTAAAGAGGGAACAACTGTGGTGATGGAAATTTTGTTTAAGTAGGTGATAAAATGTTATATT
TTCACTCTGTAACATTAGATAAGGATAGATGTAGAGGGTGTACTAATTGCATAAAAAGATGTCCTAC
AGAGGCAATAAGGGTAAGAGATGGGAAAGCAAGGATTATAAATGAAAGGTGTATAGATTGTGGTG
AGTGTATACGTGTGTGTCCATATCATGCTAAATTGGCGGTAACTGATAGTTTGGATATGATGAAAGA
TTTCAAGTACAAGATTGCACTTCCTGCTCCTTCTTTATATGGACAATTTAGGGATTTAACTATTAATCA
AATATTGAGTGCTTTACTGGATGTGGGATTTGATGAGGTTTTTGAAGTAGCATATGCTGCTGAAATTG
TTTCTAAATTTACGCGGGAAGCATTAGCAAAAGGGAATTTAAAAAAACCAGTTATATCTTCTGCTTGT
CCTGCTGTTGTGCGACTCGTACAAATCAGATTTCCCTCTTTGATTGATAATTTATTGGATATATGTTCTC
CTATGGATACTGCGGCGATATTAGCAAAAAAGGAAGCAATTAAAAAGACGGGACTGAAGGAAGAG
GAAATAGGAGTTTTTTTCATTTCTCCCTGTGCTGCAAAAGTGACAAGTGTGAAAAATCCTATTGGTAT
AGAAAAATCCAAAATTGATGGGGTTTTTTCTATGAAAGAAATATATGGCCTAATAATTCAAAAAGCT
AAGACTACTGTTGTTAGAGATTTAAGCAAGGCTTCTATGATAGGTGTTGGCTGGGCAAACTCTGGGG
GAGAAGCTTTTGGTACTTTTACAGAAAATAGTATTTATGTAGACGGGATACACAATGTTGTGGATGTT
TTAGAAGAGATTGAATTAGGAAAATTGAACGATTTAGACTTTTTTGAAGGGCTTGCATGTATCGGTG
GGTGTATTGGTGGACCGTTGACTGTTGAAAATAATTTTGTGGCCAAAAATAGAATAAGAAAACTGAC
AGAAAAACTTCCTAAAAAAGAAGAGGCTTTGTTTGATGAGGAAGAGATTGATTTTGAAGAGGTGAA
ATGGAAAAAGAAAATAGAAAAAAGTGAAGTTATGAAGTTGGACAAAGACATTTCAAAAGCGCTTGA
AATGATGAAACAAATAGACACTCAATACAAAGCCTTACCAGGACTTGACTGTGGTTCTTGTGGTTCTC
CTACTTGCAGGGCATTGGCGGAAGATATAGTGAAAGGGTATGCTACGGAATATGATTGCATTTTTAT
ATTAAAAGACAAAATAAAAAATTTGTCTCAAGAATTAAATGACCTTGCGGGTAAAATTCCTCCAGTTT
TAAGTGACGAAAAGGAGTGAAATTTATGAGGATAAAAGTAGAAGATCTTGTCAATAGCGGATTTAA
ATTAATTGCAGGGGACAACGGAATCAGCAAAGAAATTGAGGGGGTGTATATATGCGACCTTTTAAG
TTGGGTTATGGCTCATGCAAAAGCAAAAAGCGCTTGGATAACAATTCAAACTCATATAAATATTGTA
GCAGTGGCTTTGCTGGCAGAAATAAGTTGCATAATAATTCCCGAAAATGCGAAATTAGATGAGGAA

861: Ser/Thr protein kinase

862: TamHydS

863: DRTGG protein

GCCAAGAGGAAAGCCGATGAGGAGGGCATCCCTATATTAAGTTTTTCAGGAACTTCTTATGAAGCAG
CAATAACTTTGTACGAGATGATGAAGTGATGCTTTACTATGATTTGCACATTCATACAGCGCTATCTCC
TTGTGCTTCTGATGATATGACTCCTAACAATATTGTCAATATGGCTTCTATAAAAGGACTTGATGTAAT
AGCTATAACAGATCATAACAGTGCAAAAAATGTAAAAGCTGTGTATAATCTTGGATTAAAAAAAGGC
TTAATAGTAGTTCCTGGTATAGAGGTACAGACAAGAGAAGAGGTTCACATACTTTGTTATTTTTATTC
AGTAGATGAATGTATAAAATTTAGTGAAATTATTAACAAAAATTTGATAAAAATTAAAAACAAAAAA
ACAATCTTTGGGAATCAATTTGTAATGAATGAAGAAGATAACATAGCAGAGGAAATAGATTACTCAT
TGTTAGTTTCTTCTAATTTAAGTGTTAATGAAATTTTCGACTACATGGAGGGCAGGGGGGTAGCGGTC
CCTGCCCATGTGGATCGATATTCCTATAGTATTATATCAAATCTTGGTTTTATTCCGAATATTAAAAAT
TTAATTACAATCGAGGTATCGAAAGCCATCAGAAAAGAAAATTTTTTACACTCATATCCCGAATATAA
AAAATATAAAATTATAAGGTCTTCAGATGCTCATTATTTAGGAGATATTTCTGAAAGAGAAGAATTCT
TACTGTGCGGGTCTGAATTAAAAAGTATTGTAGACTGGCTAAGAGGTTATTGATTTTTCCTATAATCCT
ACTAAATTAGTTGGACTTATTAGAAAATAAAAATTTTTTATAAAAAATTGTTAAAGTTTTAATAAAGTTA
TGATATAATATTGACAAAGAATGCACAAAAAATTTATATGAGATAAGCTATATGAAATTTTAATTCACAA
GGAGGGAAATAAATGGATGCTTTGTGCCAAAAATTCGGTGAAGAGAAAGTAGAGAGGTTTAAAAAG
GCTTTGGAGGAATTAAAAAATATCCCTGGTTCGTTGATAGCAATTATGAATGAGGCTCAGGAAATTTT
TGGATATCTTCCTATTGAAGTTCAACTTTATATTTCAAAAGAAATGAATGTGCCTTTAACAGAAATATT
TGGAATAGCTACTTTTTATTCAAGGTTTACTTTAAAGCCTTCGGGTAAATACAAGATCAATTTGTGTAT
GGGAACAGCTTGTTATGTAAGAGGAGCTGCAATGGTACTGGAAAAAATAAAAGAAAAATTGGGAAT
AGAAGTAGGTGAAGCAACTGAAGACGGGAAATTTTCTTTAGAACCTACTAGGTGTCTTGGAGCTTGC
GGATTGGCTCCTGTTATGATGATAAATGGTGAAGTTTTTGGAAGACTAACTCCTGATGATGTGGATG
AAATATTAAGTAAATTTGAGTAAGAGTTTGATTGTTATGAAAGAATTAGCACTTTATATATTAGATTT
GTCACAAAATAGCATAAGAGCGGGAGCTAAAAATATTTGTATTGAAATAAATATAGATACTTCAAAA
GATATGCTTAAAGTGTCTATAGAAGATGATGGATGTGGAATGAAGGAGGAATTATTAAAAAAAGTC
ACAGACCCTTTTGTAACTACAAGGAAGGAAAGAAAAGTAGGGCTTGGGATTCCGCTCTTTAAGGAGC
TTGTCCAGCAGTGTGAAGGGAATTTTGAGATTTTTTCTGAAGAAGGAAAAGGAACAAAAATAATGG
GTACTTTTAAACTTTCCAGTGTAGACCTTGTTCCACTAGGAGATATAGCGTCAACAATTGTTTCGGTGA
TATTATCTGCTCCAGAAGTGGATATTGTATACAAATACAACAAAGATAACCATGAATTTTTGTTTGAC
ACTAAAGAACTTAAAAAAATACTAAAGGGTGTCAATATAAATGACATCAAGGTGTTGAACTGGATA
AAAGAATATATAAATGAGAATATGAAAGGTGATATGGAGGTGGAATAATGAAATCTATAGAGGAAT
TGGAAAAAATCAGAAAAGAAACATTAGAAAGAATCAATTTGAGAAAGGATAGGTCAGGTGTAAGAAT
TGCTGTTGGTATGGCTACTTGCGGTATTGCAGCAGGAGCGAGACCTGTCATGATGGCGATATTGGATG
AACTTAGCAAGAGAAATGTCACGGATGTAATTGTTACAGAGACAGGGTGTATAGGTATGTGTAAATTA
GAACCAATTGTAGATGTTTATGTTCCCGGCCAAGAGAAAGTGACTTACGTAAAGGTTGATGAGAAAAA
GGCAAGGCAAATAGTTGCTGAACATGTTATAAATGGCCATCCAATCAGAGAATGGACTATTGAAAATT
ACGAATAGAGGAGGGGAAATAGTATGCTTTATAGGTCACACGTAATGGTATGTGGTGGTACCGGATG
TACATCTTCAGATTCTGATAAAGTAGCAGAGCGTTTTACAGAGGAAATAAAAAAGGCTGGTCTAGAT
AAAGAAGTATTAGTTGTAAGAACTGGATGCTTTGGCCTTTGCGAATTGGGACCAGTTGTTGTAGTGT
ATCCTGAAGGAGTTTTTTATAGCAGAGTTAAACCTGACTATGTTCCGGAAATTGTGGAAGAGCACCT
GCTAAAAGGAAGACCTGTTAAGAAATATCTCTATGGAGAAAGCGTGACAGAAAGAGAGATTAAGCC
ATTGGAAGAGACTCCTTTCTTTAGAAAGCAAAAGAGAATTGCTCTTAGAAATTGTGGTATTATCAACC
CTGAAGATATAAGAGAAGCAATTGCTTTTGATGGATATAGGGCTTTAGCAAAGGTTCTTACTCAAAT
GACACCAAAAGAGGTTATCGAAGAAGTCAAAAAATCTGGCTTAAGGGGTAGAGGCGGCGGTGGAT
TCCCGACAGGTGTGAAATGGGAATTTGCTTATAACCAAAAAGAAACTCCAAAGTATGTAGTTTGCAA
CGCTGATGAAGGGGACCCTGGTGCTTTTATGGATAGAAGCATATTAGAAGGAGACCCCCACAGTGTT
ATAGAGGCTATGGCAATAGCAGGATACGCAATTGGAGCAAACTATGGCTACATCTACGTAAGGGCA
GAGTATCCTTTGGCAGTAAAAAGGCTTAAGATTGCTATTGAACAGGCTAGAGAATATGGCCTTTTAG
GAAAAGACATATTTGGTACAGGTTTTGATTTTGACATTGAAATAAGGTTAGGAGCAGGCGCTTTTGT

Overlapping ORF: Stopp-codon A = start codon A

864: PHP protein domain

865: NADH-quinone oxidoreductase

866: Histidine kinase

868: NADH dehydrogenase

CTGTGGAGAAGAGACTGCGCTTTTAAATTCAGTAATGGGAAGAAGAGGCGAACCAAGGCCAAGACC
ACCATTTCCTGCTGTAAAGGGTGTATGGAGCAAACCTACAATTATTAATAACGTAGAGACATTCGCTA
ATATTCCGCCAATTATACTAAATGGTGCGGAGTGGTTTTCAAGCATAGGTACTGAAAAATCGAAAGG
AACAAAGGTTTTTGCCCTGACCGGTAAGGTAAACAATACAGGTCTTATTGAAGTTCCAATGGGGACC
ACTTTAAGAGAAATTATATACGAAATAGGTGGAGGAATACCTAACGGAAAGAAATTTAAAGCTGCTC
AAACTGGAGGACCTTCAGGTGGCTGCATTCCTGCAGAGCATTTAGATACACCAATAGATTACGATTCC
TTAATAAATGTAGGTTCTATGATGGGTTCAGGTGGACTTGTTGTAATGGATGAAGACACTTGTATGG
TCAATGTTGCAAAATTCTTCCTTGAGTTTACAGTTGATGAGTCCTGCGGTAAATGTGCGCCTTGTAGG
ATTGGCACAAAAAGAATGCTGGAGCTTTTAGAAAAGATAACTTCTGGTAAAGGTGAAGAAGGAGAC
ATTGAAAAGTTGGAGGAACTTGCGAAAACCATTAAAGCGACAGCTTTGTGCGGCCTTGGACAAACA
GCCCCAAATCCGGTTTTATCTACAATTAGATATTTCAGAGATGAATATGAGGCGCATATAAAAGAGA
AAAAATGCCCAGCTGGCGTATGCCAAGCCCTCTTAAGATTTTATATTGACCCCGATAAGTGTAAAGGC
TGTGGAATTTGTGCTAAAAATTGTCCTGTCAATGCTATTTCTGGAAAACCAAGGCAGCCTTATGTGAT
TGACCAAGACAAATGTATAAAGTGCGGCACATGCATTGAAAAATGTCCATTTGGCGCAATCTACAAA
AAATAATTGAAGGAAGGAGTGTAAAAGATGGACAAAGTTCGTGTTACGATAGATGGCATCACTGTG
GAAGTTCCCTCCCATTATACCGTATTGGAAGCAGCAAAAGAAGCAGGAATAGATATCCCCACACTGT
GTTACCTCAAGGAAATCAACCAAATTGGCGCTTGCCGTATATGTGTAGTTGAAATAGAGGGAGTTAG
AAATTTACAAACCTCCTGCACTTATCCGGTATTTGATGGCATGAAAGTGTATACAAATACAGCTAAAA
TAAGAGAAGCGCGGAAATTAAATCTTGAGCTTATACTTTCAAATCATGATAGAAGTTGTTTGACTTGT
ATTAGAAATACTAACTGTGAGCTTCAATCATTGTCTAAAAAATTAGGAGTAAATGAAATAAGGTTTG
AAGGGGAAAATATAAAGTATTCTATAGATAATGCTTCTCCTGCTATAGTTCGAGATCCAAATAAGTGT
GTACTTTGTAGAAGATGTGTAGCAGTATGTTCAGAAGTGCAAAACGTATTTGCCATTGGAATGGTAA
ATAGAGGATTTAAAACAATGGTAGCACCTTCTTTTGGGAGAAGTTTAAAAGATTCCCCTTGTATAAGT
TGTGGACAGTGTATTGAGGTATGTCCTGTTGGAGCTATATATGAAAAAGACCATACAAAAAGAGTTT
ATGAAGCTTTAGCTGATGAGAAAAAATATGTAGTAGCTCAAACAGCTCCAGCAGTAAGAGTAGCATT
AGGTGAAGAGTTTGGTATGCCTATAGGAAGTATCGTTACTGGAAAAATGGTAGCAGCTTTGAAGAG
GATGGGTTTTGACGCGGTATTTGATACGGATTTTGCGGCGGACTTGACTATAATGGAAGAAGGCTCG
GAACTTCTTGAAAGGCTTAAAAATGGTGGAAAACTTCCTATGATAACTTCCTGTAGTCCTGGTTGGAT
AGCTTTCTGTGAAAAATATTATCCAGAATTTATAGATAATCTTTCAACTTGTAAGTCTCCTCACATGAT
GATGGGGGCATTAGTAAAGAGTTATTATGCAGAAAAGGAAGGGCTTAATCCTGAGGATATATATGT
GGTATCAATTATGCCATGTACTGCTAAAAAACTAGAGATTGAAAGACCAGAAATGCAGCATAATGGA
ATAAAAGATGTAGATGCTGTCCTTACCACAAGAGAATTGGCGAGAATGATAAAAGAAATGGGCATC
GACTTTGTAAATCTTCCTGATGAAGAATATGATGAACCTCTTGGAATGTCCACTGGCGCTGGAGTGAT
ATTCGGAGCTACAGGTGGAGTTATGGAGGCAGCTTTAAGGACAGTTGCAGATATTGTAGAAGGAAG
GGAATTAGACAAGGTTGACTTTGAAGAAGTGAGAGGCTTAGAAGGGGTAAGAGAAGCTACCATAA
AAATAGATGGTATGGACATAAAAGTTGCCATAGCAAATGGTACAGGAAATGCTAAAAAACTTCTTGA
CAAAATAAAGGCTGGTGAAGTTGAGTACCACTTCATAGAGGTAATGGGTTGCCCAGGAGGATGCAT
AATGGGAGGCGGTCAGCCGATTCACAATCCAAATGAAATGGGAAAAGTAAAAGAATTAAGAGCAA
AAGCCATTTATGAAGCTGATAAAAGCTTACCCATAAGAAAATCTCACAAAAATCCTGCAATACAAAA
ACTTTATGAAGAATTTTTGGGAAGTCCTTTAAGTGAGAAGTCTCATCATTTACTTCACACTCACTATTC
CAAGAAGGAATTGTATCCTTTGGTGAGATAAGCTTAATCAAAAGTAAAGCTTCCAATATATTATTGGG
AGCTTTACCATTTTTTATTTTTAAAGTTAGTTATAGAAATGTATATAATTTATGATATAATAAAAAAGAAT
AGTAATTTATGATGCTATTAGAACTTAAGGAGTGAAAGAAATTGAGAAGAGGTAAGAGTCCATATACA
TTGGTGTTTTTACTCATTGTCGGCCTAATTTTAGGAGAATTCTTAGGAGATCTGCTTGCACATTTTTTTAA
AGAATTAGCTTATTATCAAATAATAGGCATGGACAATCCACTTACTTTGAACTTGAGTTTTATTAGGTTT
TCTTTTCAGTTATCTCTTAAAATAAATATAGGAACAGTTGTAGGATTAGTTCTTGCAATTTACGCTTATTA
CAAAATGTAGGTGATAAGATGAAAATTGTCCTTGCTTCTAAATCGCCAAGGAGAAGAGAGTTACTTTC
AAATTTAGGACTTGACTTTGAAGTGGTAGAAAGTAATGTAAAAGAGTTTTCAAGAGAAAAGCAGCCT

869: Group A [FeFe]-hydrogenase

871: maf protein

TCAAGATATGTCATGGATTTATCTTTTAATAAAGCTATGTCGGTTGCTAAAAAATTAAGTGAAGAAGC
TATAGTGATAGGTGCTGATACCGTTGTAGTAATTGGAGATAAAATTTTAGGAAAGCCAAAGGATAG
AAATGAAGCTTATGTTATGCTTAAAAAGTTACAGGGAAAAGCTCATACAGTATATACAGGAATTACT
ATCGTCAGAACTTACGATTTTAAATATGTAAGCGATTTTGAAAAGACAAAAGTTTGGATTAGAAAAT
TACAAGATGAGGAAATTTTTAATTATATCGATACTGGCGAAGTTTATGATAAAGCTGGGGCATATGC
TATACAAGGAGTGGGTGCTCTTATTGTTGAGAAAATTGAAGGAGATTATTTTAATGTGGTAGGTCTTC
CAATTTCAAAACTTTTTGATATTTTAAAAAGAGAGTTTGATGTGAGGTTACTTTGAAAGGGTGAAGGG
GTTGGGCAAG

Supporting References

1. B. Taboada, K. Estrada, R. Ciria and E. Merino, *Bioinformatics*, 2018, **34**, 4118-4120.
2. V. V. Solovyev, A. A. Salamov, I. A. Seledtsov, D. Vorobyev and A. G. Bachinsky, 2011.
3. S. T. Stripp, *Acs Catal*, 2021, **11**, 7845-7862.
4. M. Senger, S. Mebs, J. Duan, F. Wittkamp, U. P. Apfel, J. Heberle, M. Haumann and S. T. Stripp, *Proc Natl Acad Sci U S A*, 2016, **113**, 8454-8459.
5. M. Senger, T. Kernmayr, M. Lorenzi, H. J. Redman and G. Berggren, *Chem Commun (Camb)*, 2022, **58**, 7184-7187.
6. M. Senger, S. Mebs, J. Duan, O. Shulenina, K. Laun, L. Kertess, F. Wittkamp, U. P. Apfel, T. Happe, M. Winkler, M. Haumann and S. T. Stripp, *Phys Chem Chem Phys*, 2018, **20**, 3128-3140.
7. M. Senger, K. Laun, F. Wittkamp, J. Duan, M. Haumann, T. Happe, M. Winkler, U. P. Apfel and S. T. Stripp, *Angew Chem Int Ed Engl*, 2017, **56**, 16503-16506.
8. H. Land, A. Sekretareva, P. Huang, H. J. Redman, B. Nemeth, N. Polidori, L. S. Meszaros, M. Senger, S. T. Stripp and G. Berggren, *Chem Sci*, 2020, **11**, 12789-12801.
9. D. V. Rial and E. A. Ceccarelli, *Protein Expr Purif*, 2002, **25**, 503-507.
10. E. S. Morales, I. L. Parcerisa and E. A. Ceccarelli, *Protein Sci*, 2019, **28**, 800-807.
11. H. Li and T. B. Rauchfuss, *J Am Chem Soc*, 2002, **124**, 726-727.
12. A. Silakov, C. Kamp, E. Reijerse, T. Happe and W. Lubitz, *Biochemistry-Us*, 2009, **48**, 7780-7786.
13. S. Mebs, M. Senger, J. Duan, F. Wittkamp, U. P. Apfel, T. Happe, M. Winkler, S. T. Stripp and M. Haumann, *J Am Chem Soc*, 2017, **139**, 12157-12160.
14. C. Lorent, S. Katz, J. Duan, C. J. Kulka, G. Caserta, C. Teutloff, S. Yadav, U. P. Apfel, M. Winkler, T. Happe, M. Horch and I. Zebger, *J Am Chem Soc*, 2020, **142**, 5493-5497.
15. M. Senger, V. Eichmann, K. Laun, J. Duan, F. Wittkamp, G. Knor, U. P. Apfel, T. Happe, M. Winkler, J. Heberle and S. T. Stripp, *J Am Chem Soc*, 2019, **141**, 17394-17403.