Supplement:

#### Methods

# Cultivation of *T. elongatus* WT strain and preparation / solubilization of thylakoid membranes

*T. elongatus* WT strain was grown in BG-11 medium (Rippka et al. 1979) at 45°C under illumination of increasing intensity of 50-200 µmol of photons m<sup>-2</sup> s<sup>-1</sup> in a 20L foil fermenter at 5 % CO<sub>2</sub> and then transferred to air level CO<sub>2</sub>. The culture was harvested and the thylakoid membranes were prepared as described in (Kuhl et al. 2000). The thylakoid membranes were suspended in solubilisation buffer (20 mM MES pH 6.0; 25 mM MgCl<sub>2</sub>; 1 % (w/v) β-DM; 1 mM TLCK; 1 mM Pefabloc; 5 mM Na-ascorbate) at a final chlorophyll concentration of 1 mg/ml and incubated for 45 min at 20°C and gentle agitation. An equal volume of dilution buffer (20 mM MES pH 6.0; 1 mM TLCK; 1 mM Pefabloc; 5 mM Na-ascorbate) was added, to get a final concentration of 0.5 % β-DM. Insoluble material was removed by centrifugation at 45.000 g.

#### Purification of NDH-1L by liquid chromatography

The NDH-1L-complex was purified by Ni<sup>2+</sup> affinity chromatography, followed by size exclusion chromatography. The solubilized thylakoids were filtered through a 0.45 µm membrane and then applied to a chelating sepharose fast flow column (GE healthcare, 7.854 ml) with a flow rate of 1 ml/min, previously saturated with 0.1 M NiSO<sub>4</sub> and equilibrated with equilibration buffer (20 mM MES pH 6.0; 0.5 M Mannitol; 0.03 % (w/v)  $\beta$ -DM; 150 mM NaCl; 5 mM Na-ascorbate). The column was washed with equilibration buffer containing 5 mM histidine and the proteins were eluted with a 10-100 mM histidine gradient in equilibration buffer. After concentration of the fractions containing the complex (Amicon concentrator, 15 ml, CutOff 100 kDa), the sample was applied to a Superose 6 size exclusion chromatography column (GE healthcare), previously equilibrated with SEC-buffer (20 mM MES pH 6.0; 0.5 M Mannitol; 0.03 % (w/v)  $\beta$ -DM; 150 mM NaCl; 5 mM Na-ascorbate). After isocratic elution with a flow rate of 0.15 ml/min and a following concentration step, the purified samples were stored at -80 °C.

#### Electrophoresis

5 µg of purified proteins were loaded on a polyacrylamide gradient BN-Gel (3.5 - 16 %) in a Mini Protean 3 electrophoresis unit (BioRad). The electrophoresis was run as described in (Battchikova et al. 2005). Afterwards, the BN-gel lane was cut out, incubated and loaded on to a 14 % SDS-PAGE Gel with 6 M Urea. The proteins were visualized by silver staining (Blum et al. 1987).

## Mass spectrometry analysis of intact proteins

The masses of intact small proteins of the NDH-1 complex were determined by MALDI-ToF mass spectrometry according to (EI-Mohsnawy et al. 2010).

The masses of the polypeptides of NDH-1 complex from the Cramer laboratory were measured by liquid chromatography with electrospray-ionization mass spectrometry according to (Whitelegge 2002).

#### Sample preparation for protein identification via 1D-nLC-ESI-MS/MS

Analysis of 2D-gel spots: Excised protein spots from 2D-Gels were destained and the corresponding proteins were digested *in-gel* with trypsin and/or chymotrypsin (Shevchenko et al. 1996).

Analysis of isolated NDH-1 complexes in solution: The concentrated sample was diluted with 60% methanol, 40% 25 mm ammonium bicarbonate buffer and trypsin and/or chymotrypsin (each 1:100, w/w) were added to the sample. The proteolysis was performed overnight at 37 °C.

## Protein identification via 1D-nLC-ESI-MS/MS

After desalting by ZipTips (Millipore) the samples were resuspended in buffer A (0.1 % formic acid in water) and subjected to 1D-nLC-ESI-MS/MS using an autosampler. An UPLC BEH C<sub>18</sub> column (1.7  $\mu$ m, 75  $\mu$ m x 150 mm, Waters, Milford, MA, USA) and an UPLC Symmetry C<sub>18</sub> trapping column (5  $\mu$ m, 180  $\mu$ m x 20 mm, Waters, Milford, MA, USA) for LC as well as a PicoTip Emitter (SilicaTip, 30  $\mu$ m, New Objective, Woburn, MA, USA) were used in combination with the nanoACQUITY gradient UPLC pump system (Waters, Milford, MA, USA) coupled to an LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA). The analytical column oven was set to 45 °C. For elution of the peptides a multiple step gradient of buffer A to buffer B (0.1 % formic acid in acetonitrile) was applied (0-5 min: 1 %

buffer B; 5-10 min: 5 % buffer B; 10-175 min: 40 % buffer B; 175-200 min: 99 % buffer B; 200-210 min: 1 % buffer B) at a flow rate of 0.4  $\mu$ L/min and a spray voltage of 1.5-1.8 kV. The LTQ Orbitrap was operated by instrument method files of Xcalibur (Rev. 2.0.7). The linear ion trap and orbitrap were operated in parallel, i.e. during a full MS scan on the orbitrap in the range of 300-2000 m/z at a resolution of 60,000. MS/MS spectra of the four most intense precursors were detected in the ion trap. The heated desolvation capillary was set to 200 °C. The relative collision energy for collision-induced dissociation was set to 35 %. Dynamic exclusion was enabled with a repeat count of one and a one-minute exclusion duration window. Singly charged and more than triply charged ions were rejected from MS/MS.

#### **SEQUEST Analysis**

The SEQUEST algorithm was used for MS/MS data interpretation. To obtain reliable protein identification, only peptides with a  $\Delta$ Cn score above 0.1 were used. In addition the cross-correlation scores of double and triple charged peptides had to be greater than 2.5 and 3.5, respectively. As modifications the oxidation of methionine was permitted.

NDH-1 SU	ORF <sup>a</sup>	kDa⁵	TMH℃	$XC^d$	Coverage (%) <sup>e</sup>
NdhF1	tll0720	71.972	16	30.1	1.37
NdhH	tlr1288	45.216		120.3	33.76
NdhD1	tll0719	56.078	12	68.3	9.92
NdhB	tll0045	55.144	14	30.2	6.6
NdhA	tlr0667	41.347	13	60.2	16.09
NdhK	tlr0705	25.742		70.2	24.47
Ndhl	tlr0668	22.415		40.2	20.41
NdhG	tlr0669	21.569	5	30.3	13.00
NdhJ	tlr1430	19.343		10.2	8.33
NdhN	tlr1130	16.636		10.1	7.33
NdhM	tll0447	12.567		40.3	45.05
NdhO	tsl0017	7.867		10.2	18.57

Table S1: NDH-1 subunit analysis after *in-gel* digestion with trypsin

<sup>a</sup> Cyanobase ORF ID.

<sup>b</sup> Calculated molecular weight.

<sup>c</sup> Number of predicted transmembrane helices.

<sup>d</sup> XC score:  $((\Delta Cn^2)+Sp) \cdot Xcorr$ 

<sup>e</sup> Protein coverage given by percentage of identified amino acids.

NDH-1 SU	ORF <sup>a</sup>	kDa⁵	TMH℃	$XC^d$	Coverage (%) <sup>e</sup>
NdhA	tlr0667	41.347	13	50.3	15.04
NdhB	tll0045	55.144	14	40.2	12.43
NdhC	tlr1429	15.003	3	10.3	15.91
NdhD1	tll0719	56.078	12	50.2	17.32
NdhE*	tlr0670	11.133	3	10.3	28.71
NdhF1	tll0720	71.972	16	40.3	9.76
NdhG	tlr0669	21.569	5	40.3	21.50
NdhH	tlr1288	45.216		110.3	34.01
Ndhl	tlr0668	22.415		90.3	45.41
NdhJ	tlr1430	19.343		40.2	32.14
NdhK	tlr0705	25.742		50.2	21.52
NdhL	tsr0706	8.571	2	10.1	11.84
NdhM	tll0447	12.567		60.2	40.54
NdhN	tlr1130	16.636		50.3	46.00
NdhO	tsl0017	7.867		20.2	35.71

Table S2: NDH-1 subunit analysis after digestion with trypsin in solution

<sup>a</sup> Cyanobase ORF ID.
<sup>b</sup> Calculated molecular weight.
<sup>c</sup> Number of predicted transmembrane helices.
<sup>d</sup> XC score: ((ΔCn<sup>2</sup>)+Sp) • Xcorr
<sup>e</sup> Protein coverage given by percentage of identified amino acids.

\* Identified after cleavage with trypsin and chymotrypsin.

Table S3: Identification	of NdhP and	NdhQ by s	pecific pe	ptides
--------------------------	-------------	-----------	------------	--------

	MH+ <sup>a</sup>	∆M (ppm) <sup>b</sup>	P <sup>c</sup>	z <sup>d</sup>	Xcorr <sup>e</sup>	∆Cn <sup>f</sup>	Coverage <sup>g</sup> (%)
NdhP <sup>h</sup>	1782.710	4.8	2.3e <sup>-11</sup>	2	5.145	-	36.36
NdhQ <sup>i</sup>	1299.651	0.9	5.3e⁻⁵	2	2.744	0.575	24.44

<sup>a</sup> Measured masses of the precursor ion. <sup>b</sup> Mass difference of calculated and measured masses in ppm. <sup>c</sup> Peptide probability calculated by the Bioworks software.

<sup>d</sup> Charge of the peptide.

<sup>a</sup> Charge of the peptide.
<sup>e</sup> Cross-correlation score calculated by the sequest algorithm.
<sup>f</sup> ΔCn value calculated by the sequest algorithm.
<sup>g</sup> Protein coverage given by percentage of identified amino acids.
<sup>h</sup> After digestion with trypsin.
<sup>i</sup> After digestion with chymotrypsin.

	Calculated	Measured.average	∆M (Da) <sup>c</sup>	TMH <sup>d</sup>	Modification <sup>e</sup>
	average	mass <sup>b</sup>			
	mass <sup>a</sup>				
NdhP	4902.67	4902.65	0.02	1	N-Formyl
NdhQ	4710.53	4710.52	0.01	1	Minus Met-1

## Table S4: Masses of intact NdhP and NdhQ

<sup>a</sup> Calculated average masses (MH+).

<sup>b</sup> Measured average masses (MH+).

<sup>c</sup> Mass difference of calculated and measured masses.

<sup>d</sup> Number of predicted transmembrane helices.

<sup>e</sup> Predicted common modifications based on the measured masses.

Gene and translated amino acid sequences of the novel NDH-1 subunits

ndhP

Sequence:

ATGGATGCTGTGATTAGCGTAAAGCCCATTTTGCTGGCTATGACGCCTGTATTTA TTCTGTTGTGTTTGTTTTTTGGCACCCGCAATGGCTTCTACGACACGGATCAATA CCACGGTAACGGTTCTGCCCAC

Genomic region: 1189596-1189465

Translated amino acid sequence:

MDAVISVKPILLAMTPVFILLCLFFGTRNGFYDTDQYHGNGSAH

#### ndhQ

Sequence:

ATGGCCACGGATTTTAATCGCGGCATTATGAAGTTTGATGGTGCCGACAGCCCG GCGATGATTGCGATTTCTGCGGTCTTGATTCTTGGCTTTATTGCAGGACTGATTT GGTGGGCACTCCACACCGCTTACGCC Genomic region: 105314-105448

Translated amino acid sequence: MATDFNRGIMKFDGADSPAMIAISAVLILGFIAGLIWWALHTAYA

#### Figure S1



Position of the *ndhP* and *ndhQ* genes within the genome of *T. elongatus*. Open reading frames were predicted with Glimmer (http://bioinformatics.biol.rug.nl/websoftware/orf/orf\_start.php) and included into the genome of *T. elongatus* as annotated in Cyanobase (http://genome.kazusa.or.jp/cyanobase).





Result of MALDI-ToF mass spectrometry of intact NDH-1 complexes. Samples ware analyzed in the low molecular mass range according to (EI-Mohsnawy et al. 2010) and the measured masses were assigned to small NDH-1 subunits with respect to common posttranslational modifications.



<u>NdhP</u>



Prediction of transmembrane helices for NdhP and NdhQ. The translated amino acid sequences of *ndhP* and *ndhQ* were used to predict transmembrane helices with TMHMM (http://www.cbs.dtu.dk/services/TMHMM/). The red bar indicates the position of the TMH within the sequence.

# Figure S4

# <u>NdhP</u>

	10	20	30	40	50
NdhP T_elongatus/1-44	· · · · · · · · · · MDAVISVK	PILLAM	PVFILLCLFFG	RNGFYDTDQ	YH <mark>GNGSAH</mark>
tr B8HVZ5 B8HVZ5_CYAP4/1-40		LILVCM	PVFILLCLLFG	KNGFYDTDN	YH <mark>GNGSAH</mark>
tr B0C803 B0C803_ACAM1/1-40	MDLK	TILV <mark>G</mark> L <mark>T</mark>	PIFIVLCLFF <mark>G</mark> 1	KNGFYDSDD	YH <mark>GNGSAH</mark> · · · ·
tr Q0QM02 Q0QM02_FREDI/1-40		(LILV <mark>g</mark> l <mark>t</mark>	VIFTVLCLFFGT	KNGFYDSDN	YH <mark>GNGSAH</mark>
tr B3DFB3 B3DFB3_MICAN/1-40	· · · · · · · · · · · · · · · · M <mark>D</mark> I K	LILLAL	AVF <mark>T</mark> V <mark>S</mark> CLFF <mark>G</mark> 1	RNGFYDSDN	YD <mark>GNGSAH</mark>
tr B1XNQ7 B1XNQ7_SYNP2/1-47	· · · · · · MYSY <mark>P</mark> N <mark>PMD</mark> IK	(LLLLAL <mark>T</mark>	G V F T V A C L F F G T	QNGFYDSDD	YH <mark>GNGSAH</mark>
tr D7E4C2 D7E4C2_NOSA0/1-40		(LILV <mark>g</mark> l <mark>t</mark>	VIFTFTCLFFG1	KNGFYDSDN	YH <mark>GNGSAH</mark>
tr B7K8C9 B7K8C9_CYAP7/1-49	····MFN <mark>GG</mark> VISI <mark>MD</mark> AK	(LVML <u>I</u> L <mark>T</mark>	GLFIVSCLFFG1	KNGFYDSDN	YH <mark>GNGSAH</mark> · · · ·
tr Q10WS7 Q10WS7_TR/E//1-40	<mark>MD</mark> VK	(LILV <mark>g</mark> l <mark>t</mark>	F L F <mark>T</mark> I <mark>G</mark> C L F F <mark>G</mark> 1	<mark>QNGFYDTD</mark> D	YH <mark>GNGSAH</mark> · · · ·
tr E0U6P3 E0U6P3_9CHRO/1-40	<mark>MD</mark> AK	CIVMIIL <mark>T</mark>	G L F I I <mark>S</mark> C L F F <mark>G</mark> 1	KNGFYDSDN	YH <mark>GNGSAH</mark> · · · ·
tr C7QSH4 C7QSH4_CYAP0/1-40	<mark>MD</mark> VK	(LILVIL <mark>T</mark>	G L F I I <mark>S</mark> C L F F <mark>G</mark> 1	KNGFYDSDN	YD <mark>GNGSAH</mark> ····
tr B7JYA3 B7JYA3_CYAP8/1-40	<mark>MD</mark> V K	(LILV <u>I</u> L <mark>T</mark>	GLF <u>I</u> I <mark>S</mark> CLFF <mark>G</mark> 1	KNGFYDSDN	YD <mark>GNGSAH</mark> ····
tr D4ZT45 D4ZT45_SPIPL/1-40	<mark>MD</mark> I K	(LILV <mark>G</mark> L <mark>S</mark>	VVF <mark>SIA</mark> CIFF <mark>G</mark> 1	Q N G F Y D S D D	YH <mark>GNGSAH</mark> · · · ·
tr A3IP20 A3IP20_9CHRO/1-40	<mark>MD</mark> V K	(LILVIL <mark>T</mark>	ALF <mark>T</mark> V <mark>S</mark> CLFF <mark>G</mark> 1	KNGFYDSDN	YD <mark>GNGSAH</mark> · · · ·
tr B1WX88 B1WX88_CYAA5/1-46	MRRKSIMDV	(LVLVIL <mark>T</mark>	ALF <mark>T</mark> V <mark>S</mark> CLFF <mark>G</mark> 1	KNGFYDSDN	YD <mark>GNGS</mark> AH
tr Q7NEQ5 Q7NEQ5_GLOVI/1-39		(LVILV <mark>I</mark> A	IAF I PLALFFA	RNGFYNTDR	YH <mark>GNGSAH</mark>
tr Q3AMW7 Q3AMW7_SYNSC/1-48	· · · · · MMDAATSSFNLG	TVLLA <mark>S</mark> I	VLF <mark>P</mark> LACLFF <mark>G</mark> 1	RGGYYNTDQ	YD <mark>GNGT</mark> AH
tr D6PGP3 D6PGP3_9BACT/1-47	· · · · · · MDAATSS <mark>F</mark> NLG	TVLLA <mark>S</mark> V	VLF <mark>P</mark> LACLFF <mark>G</mark> 1	RGGYYNTDQ	YD <mark>GNGT</mark> AH
tr B4WKT7 B4WKT7_9SYNE/1-40	· · · · · · · · · · · · · · · · MD I K	(LVFFIL <mark>T</mark>	G L F S V A C L F F G 1	RNGFYDSEN	YH <mark>GNGSAH</mark>
tr D3EP09 D3EP09_UCYNA/1-40	· · · · · · · · · · · · · · · · MD   K		TLF <mark>T</mark> VSCLFF <mark>G</mark> 1	KNGFYDSDD	YK <mark>gngtah</mark>
tr D0CL66 D0CL66_9SYNE/1-53	MSSGQVMNAATSSFNLG	TVLLA <mark>S</mark> I	VLF <mark>P</mark> LACLFF <mark>G</mark> 1	RGGYYNTDQ	Y D G N G T A H · · · ·
tr Q061C7 Q061C7_9SYNE/1-47	· · · · · · MDAALSGFNLG	TVLLA <mark>S</mark> I	VLF <mark>P</mark> LACLFF <mark>G</mark>	RGGYYNTDK	Y D G N G T A H · · · ·
tr Q31QG9 Q31QG9_SYNE7/1-46	· · · · · · · · MRS <mark>P</mark> RT <mark>MDF</mark> K		I PFTLATLYFG1	RNGFYDSDD	Y H G N G T A H
tr Q2JWH2 Q2JWH2_SYNJA/1-40		(LVLV <mark>G</mark> VA	LVL <mark>S</mark> LASF <b>Y</b> FG1	RNGFYDTDK	YHG <mark>NGSAH</mark>
tr Q3AUM5 Q3AUM5_SYNS9/1-47	· · · · · · MEAALAGENLG	TVLLA <mark>S</mark> I	VLF <mark>P</mark> LACLFF <mark>G</mark>	RGGYYNTDK	Y D G N G T A H
tr B4VXU8 B4VXU8_9CYAN/1-40			ILFTVSALIFG	KNGFYDSDN	YHGNGSAH · · · ·
tr[A4CSW3[A4CSW3_SYNPV/1-47	····MDAALSGFNLG	TVLLFGS	GLEVLATLEEG	RGGYYNTDQ	YDGNGTAH · · · ·
tr[A5GiK1]A5GiK1_SYNPW/1-47	·····MDAALSGFNLG	TVLLFGS	GLEVLTILFEG	RGGYYNTDK	YDGNGTAH
thA328M/[A328M/_9SYNE/1-47	···· MDAALHSFNLG	TVLLFGS	GLEVLILLYEG	RGGYYNTDQ	YDGNGTAH
MA3YX61 A3YX61_9SYNE/1-47	·····MSNALSSFNLG	TVLLAGS	GLECLATLYEG	RGGYYDSDD	YDGNGTAH
MBS/P25/BS/P25_9CHRO/1-45	MESSOFILA	TULEU	GEF CLATEFF G	KUDEVECEN	YVCDCCAUDVVD
MABG4L6JA8G4L6_PROM2/1-44			C F V L L I V F F G I	KNDFYESEN	Y KODOCAHOVKR
10000011000001_95111E/1-47	MEAALSGENLG		UPEVIL TVEE	KUREYEREN	YVODGCAHDVVD
				PNGYYDTDK	YOGDGCAHDVKR
++0318431031843 PROM9/1-44				KNDEVEREN	VKGDGCAHDVKR
+4A2BOX5IA2BOX5_PROMS/1-44	MDL T	Т I I E I I S		KNDEVESEN	YKGDGCAHDVKR
HOUDAOLOUDAO SVN S2/1-47	MDAAL SGENLG	TVLVEGS	GLEVIATEVEGI	PGGYYNTDK	
+1A3PCP8 A3PCP8 PROM0/1-44	MDL T		I PEVI I TVVEGI	KNDEVESDN	KGDGCAHDVKR
*148W/D6L48W/D6_PROMP/1-44		VVLELIS	MPEVILTAVEGI	KNDEVESEN	KGDGCAHDVKR
tdA2BWT6IA2BWT6_PROM5/1-44		AVLELIS	I PEVILITAYEGI	KNDEYESEN	YKGDGCAHDVKR
PSSM2 253/1-44		TELILAA	LPFVGLTLFFG	KNGYYDSDD	YQGDGCAHDVKR
·					
		a di kacamatan di ka			
Conservation					
		6997655	56+495769*++	7759*7995	4*8*7**
	-			العمر العرابا	
Quality		_			
Quanty					
		_			
	<b></b>				
Consensus					
	MDAALSGMDLK	TVILIT	GLEVLACIEEGI	KNGEYDSDN	VDGNGSAHDVKR

#### <u>NdhQ</u>

			10		20			30			40			50
Ndh Q  7/1-45	MA	TDFNR	G I M K	FDGAD	SPAI	MIAI	SAV	LIL	GFI	AGL	IWWAL	HTA	YA-	
tr B8HQ97 B8HQ97_CYAP4/1-47	M S	SDFDR	GIM <mark>K</mark>	(FKGAD	RPS	MIAI	SAI	LLL	GSI	GLL	IWWSL	NTA	YAL	N
tr B5W910 B5W910_SPIMA/1-46	M <mark>P</mark>	SDLDR	G I M <mark>K</mark>	(FKGAD	SPT	VVIV	SSL	LIL	G <mark>S</mark> I	GLL	IWWAL	QTA	YSF	
tr D8G423 D8G423_9CYAN/1-46	M -	SDLNR	GIM <mark>K</mark>	(FKGAD	SPT	ATAV	SAI	AIL	GGI	SFL	IWWAL	QSA	YAL	s
tr C7QLH3 C7QLH3_CYAP0/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	K P A I	LVAV	SAI	LVL	GAI	TAL	IFWAL	TTA	YSV	G
tr B7K4W1 B7K4W1_CYAP8/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	K P A I	LVAV	SAI	LVL	GAI	TAL	IFWAL	ТТА	y s v	G
tr D7DVG7 D7DVG7_NOSA0/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	SPK	ννтν	STV	LLL	G <mark>S</mark> I.	AAL	ILWAL	QSA	YAL	s
tr A0ZK00 A0ZK00_NODSP/1-46	M -	SDLNR	GIM <mark>K</mark>	(FKGAD	SPK	V V T I	STV	LLL	G <mark>S</mark> I.	AAL	ILWAL	QAA	YAL	N
tr Q2JND7 Q2JND7_SYNJB/1-45	M -	A D <mark>Y N R</mark>	GIM <mark>K</mark>	FKGAD	SPI	VVLI	SAG	IVA	GVV	SAL	IWWAL	HFA	YAA	
tr D4TTM0 D4TTM0_9NOST/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	SPK	LVMV	STV	LVL	G <mark>S</mark> I.	AIL	LIWAL	RSA	YAL	G
tr D4TLB5 D4TLB5_9NOST/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	SPK	LVMV	STV	LVL	G <mark>S</mark> I.	AIL	LIWAL	RSA	YAL	G
tr B1XP04 B1XP04_SYNP2/1-45	M -	SDLNR	GIM <mark>K</mark>	FDGAD	K <mark>P L</mark>	VVAV	SAV	LVL	GAL	AAL	V I W <mark>G</mark> L	ТТА	YSF	
tr\D4ZUJ7\D4ZUJ7_SPIPL/1-48	MY <mark>MP</mark>	SDLDR	GIM <mark>K</mark>	(FKGAD	SPT	VVIV	SSL	LIL	G <mark>S</mark> I	GLL	I <mark>g</mark> wa l	QTA	YSF	
tr B7KK31 B7KK31_CYAP7/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	K P A	IVAV	SAI	LVL	G <mark>S</mark> I	AL	LIWAL	KV A	YVV	s
tr B1WQ72 B1WQ72_CYAA5/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	K P VI	LVAI	SAF	LVL	GAI	I <mark>G</mark> L	LIWAL	KAA	Y T V	G
tr Q110V2 Q110V2_TR/E//1-46	M	SDLDR	GIM <mark>K</mark>	(FKGAD	TPR	ATAT	SAI	LIL	G <mark>S</mark> I	VELI	L F W <mark>G</mark> L	NTA	Y T V	G
tr A3/ZB6 A3/ZB6_9CHRO/1-46	M -	SDLNR	GIM <mark>K</mark>	FEGAD	K P F I	LVAI	SAF	LVL	GAI	I <mark>G</mark> L	I I WAL	NAA	ΥTI	s
tr Q31QZ6 Q31QZ6_SYNE7/1-45	<mark>M</mark> S	- DL <mark>N</mark> R	GIM <mark>K</mark>	(F <mark>Q</mark> GAD	NPL	AIGL	SAV	LIL	G <mark>S</mark> I	GLL	I L W <mark>g</mark> L	NAA	YSF	
tr B0C254 B0C254_ACAM1/1-50	<mark>M</mark> S	SDFNK	<u>G I M</u> K		NPI	TVAL	S A I	LIF	G <mark>S</mark> I	GLL	I <mark>g</mark> w <mark>s</mark> L	ETA	Y L V	GQLG
tr D3ER21 D3ER21_UCYNA/1-46	M -	SDLNR	G <mark>S</mark> M K	FEGAD	NPVI	LVAI	SAF	LVF	GFI	GAL	I VWAN	IN N A	YVI	н
tr B0JM16 B0JM16_MiCAN/1-44	M -	SDLNR	GIM <mark>K</mark>	FEGAD	K P A1	VVAI	SSI	TV I	G <mark>S</mark> I	IAL	LWAN	IKVA	Y I -	
tr A8YLF6 A8YLF6_MICAE/1-44	M -	SDLNR	GIM	FEGAD	K P A	VVAI	SSI	TV I	G <mark>S</mark> I	IAL	LWAN	IKVA	Y I -	
tr E0UE36 E0UE36_9CHRO/1-46	<mark>M</mark> -	SDLNR	G I M <mark>K</mark>	FEGAD	K P V	IVAV	SAA	LVL	GGI	VGL	I I WA I	KVA	YVV	N
tr B4VXI4 B4VXI4_9CYAN/1-38			M <mark>K</mark>	(FKGAD	SPL	ATVI	зτν	LVL	GGI	AFL	LWWAL	QTA	Y N V	G
tr A0YRS0 A0YRS0_LYNSP/1-45	M -	SDLNR	G I M <mark>K</mark>	FKGAD	SPV	VVLA	SSI	VVL	GGI	G F L 1	VAWAL	QTA	YSF	
						-		- 1						
					<b>.</b> .		<b>.</b>							
Conservation	_	ومر ال												
	1 -	14123	30**	* 5 7 * *	5 2	5957	* 7 6	796	* 49	56*	95*89	44*	51	
			_									. 🗖		
Quality					<b>.</b>	<b>.</b>								
Quality														
							_		_	_	_	_	_	
						-								
Consensus					ч.									_
	M S	SDLNR	GIMK	FEGAD	SPA	VVA+	SAI	LVL	GSI	+ A L	IIWAL	QTA	YAV	G

Multiple sequence alignment of proteins similar to NdhP and NdhQ. The multiple sequence alignment was performed with the programmes CLUSTAL-X (Thompson et al. 1997) and Jalview (Waterhouse et al. 2009). The residue colour code is as follows: orange, Gly; yellow, Pro; green, Thr, Ser, Asn and Gln; red, Lys and Arg; blue, Trp, Met, Val, Ile, Ala, Leu and Phe; pink, Cys; cyan, His and Tyr; and magenta, Glu and Asp. The red bar indicates the position of the TMH within the sequence.

# Figure S5

df6/1-174	1 MAEAFTSFTFTNLHIPSSYNHSPKQNSGPNHGYWLSNVNEKRERNLMRGSLCVRKALPHDLPLMAVMV(	68
ldhP_syn/1-40		
ldhP T_elongatus/1-44		
ndf6/1-174	69 QQIEGMRDIITEKHVW <mark>H</mark> LSDKAI <mark>K</mark> NV <mark>YMFYIMFT</mark> CWGCLYFG <mark>S</mark> A <mark>KDPFYDSEEYRGDGG</mark> DGTGYWVYE	136
NdhP_syn/1-40	1 · · · · · · · · · · · · · · · MD · · · ·	40
NdhP T_elongatus/1-44	1 · · · · · · · · · · · · · · MD AV I SV <mark>KP ILLAMTPVF ILLCLEFGT · RNGFYD TDOY</mark> HGNG <mark>S</mark> AH · · · · · · ·	44
df6/1-174	137 TQEDIEEKARAELWREELIEEIEQKVGGLRELEEAVTK	174
ldhP_syn/1-40		
ldhP T_elongatus/1-44		

Sequence comparison of NdhP and NDF6 (AT1G18730). The sequence alignment was performed with the programmes CLUSTAL-X (Thompson et al. 1997) and Jalview (Waterhouse et al. 2009). The residue colour code is as follows: orange, Gly; yellow, Pro; green, Thr, Ser, Asn and Gln; red, Lys and Arg; blue, Trp, Met, Val, Ile, Ala, Leu and Phe; pink, Cys; cyan, His and Tyr; and magenta, Glu and Asp. The red bar indicates the position of the TMH within the sequence.

#### References:

- Battchikova, N., P. P. Zhang, S. Rudd, T. Ogawa and E. M. Aro (2005). "Identification of NdhL and Ssl1690 (NdhO) in NDH-1L, and NDH-1M complexes of Synechocystis sp PCC 6803." Journal of Biological Chemistry 280(4): 2587-2595.
- Blum, H., H. Beier and H. J. Gross (1987). "Improved Silver Staining of Plant-Proteins, Rna and DNA in Polyacrylamide Gels." <u>Electrophoresis</u> **8**(2): 93-99.
- El-Mohsnawy, E., M. J. Kopczak, E. Schlodder, M. Nowaczyk, H. E. Meyer, B. Warscheid, N. V. Karapetyan and M. Rogner (2010). "Structure and Function of Intact Photosystem 1 Monomers from the Cyanobacterium Thermosynechococcus elongatus." <u>Biochemistry</u> 49(23): 4740-4751.
- Kuhl, H., J. Kruip, A. Seidler, A. Krieger-Liszkay, M. Bunker, D. Bald, A. J. Scheidig and M. Rogner (2000). "Towards structural determination of the water-splitting enzyme Purification, crystallization, and preliminary crystallographic studies of photosystem ii from a thermophilic cyanobacterium." Journal of Biological Chemistry 275(27): 20652-20659.
- Rippka, R., J. Deruelles, J. B. Waterbury, M. Herdman and R. Y. Stanier (1979). "Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria." <u>Journal of General Microbiology</u> 111(Mar): 1-61.
- Shevchenko, A., M. Wilm, O. Vorm and M. Mann (1996). "Mass spectrometric sequencing of proteins from silver stained polyacrylamide gels." <u>Analytical Chemistry</u> 68(5): 850-858.
- Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins (1997). "The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools." <u>Nucleic Acids Research</u> **25**(24): 4876-4882.

- Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2-a multiple sequence alignment editor and analysis workbench." <u>Bioinformatics</u> **25**(9): 1189-1191.
- Whitelegge, J. P. (2002). "Plant proteomics: BLASTing out of a MudPIT." <u>Proceedings of the National Academy of Sciences of the United States of America</u> **99**(18): 11564-11566.