Advances in Microbiology, **2013, 3, 350-365**
http://dx.doi.org/10.4236/aim.2013.34049 Published Online August 2013 (http://www.scirp.org/journal/aim)

Genomic Analysis of *Anabaena variabilis* Mutants PK17 and PK84 That Are Characterised by High Production of Molecular Hydrogen

Sergey V. Shestakov[1*], Lidia E. Mikheeva[1], Andrey V. Mardanov[2], Nikolai V. Ravin[1,2], Konstantin G. Skryabin[1,2]
[1]Faculty of Biology, Moscow State University, Moscow, Russia
[2]Centre "Bioengineering", Russian Academy of Sciences, Moscow, Russia
Email: [*]shestakovgen@mail.ru

Received June 7, 2013; revised July 7, 2013; accepted July 15, 2013

ABSTRACT

The use of cyanobacteria for producing molecular hydrogen is one of the desirable tasks of photobiotechnology. Some years ago, we isolated several chemically induced mutants of the cyanobacterium *Anabaena variabilis* ATCC 29413 that exhibited a high level of H_2-production; but the genetic nature of these mutants remained unresolved. To reveal mutations that could be responsible for enhancement of H_2-production in two independent mutants, PK17 and PK84, the pyrosequencing of their entire genomes was performed. The results were analyzed on the basis of comparison with the complete genome sequence of the reference strain *Anabaena variabilis* ATCC 29413. The genomes of mutants PK17 and RK84 contain 107 and 104 point deviations from the reference genome, respectively. The most probable reason for the increase of H_2-production in mutant PK17 is the mutation identified in the gene *hupL* encoding the large subunit of uptake hydrogenase. A high level of H_2-production in mutant PK84 could be the result of a mutation in a conserved part of the gene *hypF*, which participates in the post-translation maturation of hydrogenase complexes.

Keywords: Cyanobacteria; Hydrogen Metabolism; Hydrogenases; Pyrosequencing; Mutants; *Anabaena*

1. Introduction

The development of method of exploiting cyanobacteria for H_2-production on the basis of bioconversion of solar energy is a desirable task of photobiotechnology [1-3]. In heterocystous cyanobacteria, molecular hydrogen is produced by nitrogenase and reutilised by uptake (Hup) hydrogenase in heterocysts. Bidirectional (Hox) hydrogenase, which functions in heterocysts and vegetable cells, catalyses a reversible evolution/uptake reaction [3,4]. Hydrogen metabolism is a part of a complex gene network that coordinates processes of photosynthetic and respiration electron transport, carbon assimilation, redox cycle, sulfur and metal metabolism [5]. Therefore, investigation of the influence of mutations in genes involved in this network on the level of H_2-production is important for the genetic creation of efficient producers of molecular hydrogen.

Some years ago, we isolated two novel mutants of *Anabaena variabilis* ATCC 29413 that exhibited high H_2-production in the air atmosphere. The mutants were obtained by chemical mutagenesis of the wild-type strain using N-methyl-N-nitro-N-nitrosoguanidine (NTG) and selection of slower growing colonies on agar medium with limited concentrations of ammonium or nitrate [6,7]. One of them, PK84 with reduced activity of both (Hup and Hox) hydrogenases, was successively used for cultivation in photobioreactors [8,9]. The genetic nature of the mutations has remained unresolved. Proceeding from a key function of uptake hydrogenase in the recycling of hydrogen released by nitrogenase, a number of H_2-producing mutants of heterocystous cyanobacteria were created by insertional mutagenesis of *hupSL*-genes [10-14]. However, most of these mutants were characterised by a lower degree of H_2-production in comparison with the mutant PK84 [4,9].

By using a current high-throughput technique of pyrosequencing, we have attempted to reveal mutations that might be responsible for increased levels of hydrogen production in the PK17 and PK84 mutants. The results of re-sequencing of the genomes have been analysed on the

[*]Corresponding author.

basis of comparison with the complete genome sequence of the reference strain *Anabaena variabilis* ATCC 29413 (US DOE Joint Genome Institute, http://genome.jgi.doe.gov/anava/anava.info.html).
The information obtained suggests that the most probable reason for the enhancement of H_2-production in PK17 is a mutation in the *hupL* gene encoding the large subunit of uptake hydrogenase, while a high level of H_2-production in mutant PK84 is associated with mutation of the hydrogenase accessory gene *hypF*, which is involved in the maturation of active complexes of both hydrogenases.

2. Materials and Methods

2.1. Strains and Growth Conditions

The wild-type strain of the filamentous heterocystous cyanobacterium *Anabaena variabilis* ATCC 29413 was kindly provided by Dr. C. P. Wolk in 1977. Since that time, this strain was cultivated in our laboratory with periodical passages on agar slices and plates. This strain will be hereafter named as *A. variabilis* strain PW to distinguish it from the reference wild-type strain *Anabaena variabilis* ATCC 29413 (*A. variabilis* strain TT), which was derived from the laboratory of Dr. T. Thiel for sequencing experiments in the US DOE Joint Genome Institute in 2005.
(http://genome.jgi.doe.gov/anava/anava.info.html)

Two previously obtained H_2-producing mutants of *A. variabilis* ATCC 29413, PK17 and PK84 [6,7], were chosen for whole-genomic analysis. Levels of their nitrogenase activity and H_2-production are shown in **Table S1**.

For isolating of genomic DNA for pyrosequencing cultures were grown in BG11 medium [15] under fluorescent light (40 $\mu E \cdot m^{-2} \cdot s^{-1}$) at 28˚C - 30˚C in 1-L flasks containing 400 ml of magnetically stirred culture. Under these conditions the liquid cultures grew logarithmically with a generation time of approximately 20 - 24 h and after 7 - 10 days of growth contain 80 - 100 $\mu g \cdot ml^{-1}$ of protein. The total protein concentration in culture suspension was determined after the alkaline lysis of cells by a modification of the Lowry method.

2.2. Isolation of genomic DNA

Cultures of the wild-type *A. variabilis* strain PW and mutants PK17 and PK84 grown in BG11 medium for 7 - 10 days were harvested by centrifugation at 2800 g for 20 min, washed with 10 mM Tris·HCl buffer at pH 8.0 and frozen at −20˚C overnight. The samples of genomic DNA were isolated according to the manufacturer's specifications in the QIAamp DNA Stool Mini Kit (QIAGEN, Germany). The concentration of DNA was determined on a NanoDrop (ND 1000) micro volume spectrophotometer (PEQLAB, Germany). All DNA preparations were characterised by an average fragment length not less than 20 kb and had a OD_{260}/OD_{280} ratio of about 2.05 ÷ 2.1.

2.3. Genome Sequencing and Analysis

Genomes of *Anabaena variabilis* ATCC 29413 strain PW, and mutants PK17 and PK84 were sequenced on a Roche GS FLX genome sequencer using the Titanium protocol for a shotgun genome library. The GS FLX runs resulted in the generation about 106 - 120 MB per genome, which corresponds to about 15 - 20 times genome coverage. For each analysed genome, the reads were mapped on the reference genome of *Anabaena variabilis* ATCC 29413 strain TT, available in GenBank (NC_007413, NC_007410, NC_007411 and NC_007412 for the chromosome, plasmid A, plasmid B and plasmid C, respectively) using the GS Reference mapper program (Roche, Switzerland). Only point differences supported by more then 80% of reads were recorded. Genome rearrangements were not found.

3. Results and Discussion

3.1. Mutations Revealed in Genomes of Mutants PK17 and PK84

Whole-genomic pyrosequencing has shown that *A. variabilis* ATCC 29413 strain PW studied in our laboratory differs from the reference *A. variabilis* ATCC 29413 strain TT in 14 positions in the chromosome (**Tables 1**, **S2**). One mutation was revealed in plasmid pAvaA, one of three plasmids reported in the *A. variabilis* ATCC 29413 strain TT (EMBL/GeneBank/DDBJ database). In strain PW and mutants PK17 and PK84, we found an additional plasmid named pAvaD (27501 bp), which structure that is described separately [16].

In comparison with the genome of the reference strain TT, the genome of mutant PK17 (**Tables 1**, **S3**) contains 107 differences including 97 in chromosome. The genome of mutant PK84 has 104 differences (including 96 in chromosome) from the reference strain TT (**Tables 1**, **S4**). A number of mutations, 26 in PK17 and 41 in PK84, were found in noncoding regions or in sites of synonymic substitutions of the chromosomes. Several mutations were revealed in plasmids A and C. Only two deviations from the reference sequence were present simultaneously in the genomes of *A. variabilis* ATCC 29413 strain PW and both mutants (**Tables S2-S4**), suggesting that most of the sequence variations between *A. variabilis* ATCC 29413 strains PW and TT had accumulated in the course of independent laboratory cultivation of these strains apparently after obtaining the PK17 and PK84 mutants. The number of variations between the genomes of strains PW and TT is similar to that found between the labora-

Table 1. Comparison of the genomes of the wild-type *A. variabilis* ATCC 29413 strain PW and NTG-induced mutants with the reference genome of *A. variabilis* ATCC 29413 strain TT.

Strain	Sequencing volume (Mb)	Differences in sequences					
		chromosome					plasmids
		Total	protein-coding genes			noncoding regions	
			missense mutations	nonsense mutations	synonymous substitutions		
Wild-type (PW)	118	14	8	1	-	5	1
PK 17	106	97	69	2	17	9	10
PK 84	120	96	54	1	28	13	8

tory strains of *Synechocystis* sp. PCC 6803, derived from a common ancestor and cultivated independently [17].

The genomes of strains PK17 and PK84 contain missense and nonsense mutations in 74 and 58 putative genes, respectively (**Tables S3**, **S4**). In the PK17 mutant, nonsynonymous substitutions were localised in 23 genes encoding hypothetical proteins or products with unknown functions. The genome of mutant PK84 contains mutations in 17 genes of this type (**Table 2**). Nonsynonymous substitutions were revealed mostly in genes coding for histidine- and serine/threonine protein kinases, other proteins of signal transduction systems, transporters, various types of transferases and peptidases, enzymes of secondary metabolism and biosynthesis of cell walls. A number of mutations occurred in genes controlling processes of DNA repair, recombination, restriction-modification, translation etc. (**Tables 2**, **S3**, **S4**).

It should be noted that PK17 and PK84 are independently isolated mutants induced by NTG, causing multiple mutations. Therefore, it is not surprising that almost all mutations occurred in different genes (except only two of them) in the genomes of these two mutants. The absence of visible phenotypic alterations in both mutants indicates that mutations in most of the genes are located in functionally insignificant sites or in genes expressed under specific cultural conditions, for example, related to stress-induced adaptive responses. It is also possible that mutations are compensated for by other functionally similar genes. The mutant PK17 showed a decreased rate of growth (nearly 50% of wild-type strain) in nitrogen-free liquid medium and in medium supplemented with NH_4, presumably because of its possible deficiency in glutamate synthase activity (**Table S3**). Preliminary data shows the approximately 50% decreasing of GOGAT enzyme activity in PK17 cells comparing the wild type level. The growth rate of the mutant PK84 in nitrogen-free liquid medium did not differ from the wild type strain [6,7].

3.2. Mutations in Genes Encoding Hup Hydrogenase and *HypF* Accessory Protein

Some missense mutations that might be related to an increase of H_2-production in mutants PK17 and PK84 are represented in **Table 3**. The most probable gene-candidate for the mutation responsible for the enhancement of H_2-production in PK17 is the gene *hupL*; it codes for the large subunit of uptake hydrogenase (**Table 3**). A mutation in this gene leads to the replacement of the polar amino acid cysteine with aromatic tyrosine at position 446, which is located in a highly conserved part of the enzyme, as determined by alignment analysis (**Figure S1**, **Table S5**). This site is situated near the C-terminal part of the protein and may be functionally important for its conformational state. It cannot be excluded that a mutation in this site prevents the cleavage of 16 terminal amino acids by a specific hupW-endonuclease essential for the maturation of active uptake hydrogenase [18].

The obviously predictive reason leading to the high level of H_2-production in PK84 is a mutation in the gene *hypF*, the product of which plays an essential role in the post-translational modification of uptake hydrogenase as well as bidirectional hydrogenase. Inactivation of hydrogen uptake activity in the *hypF*-deficient mutant resulted in a dramatic increase in the hydrogen evolution capacity of purple sulfur phototrophic bacterium *Thiocapsa roseopersicina* under nitrogen-fixing conditions [19]. Deletion and insertion mutants of *hypF*, *hypA*1, *hypB*1, *hypC*, and other *hyp* genes of the cyanobacterium *Synechocystis* sp. PCC 6803 showed no hydrogenase activity [20]. The disruption of the *hypF* gene in the cyanobacterium *Anabaena* sp. PCC 7120 also resulted in low Hup and Hox activities without negative effects on growth rate and nitrogenase activity [21].

The substitution of a negatively charged aspartic acid to polar asparagine at position 374 was identified in the middle part of the Hyp F protein of mutant PK84 (**Table 3**). The mutation is localised in a highly conserved region according to alignment analysis (**Figure S2**, **Table S6**). Evident information about the function of this site is absent, but its location in the YrdC-like domain of the *HypF* protein (region of 213-391 amino acid sequence) indicates its possible participation in RNA binding [22, 23] and/or conversion of a carbamoyl moiety to a carbamoyl adenylate intermediate and allowing the invol-

Table 2. Substitutions in the genomes of H_2-producing mutants in comparison with the wild-type *A. variabilis* ATCC 29413 strain TT.

Mutant	Protein-coding genes containing nonsynonymous substitutions			
	Proteins with unknown functions and hypothetical proteins	Protein kinases	Metabolic and transport systems	Other genes*
PK17	23	10	31	10
PK84	17	3	32	6

*including genes related to DNA repair, recombination, restriction-modification etc.

Table 3. Mutations in the PK17 and PK84 genomes that might be related to elevated levels of H_2-production.

Mutant	Gene	Gene product	Amino acid substitution
PK17	*hupL Ava*_4595	Large subunit of uptake hydrogenase	Cys446Tyr
	*ftr Ava*_1641	Ferredoxin-thioredoxin reductase (β-chain)	Asp12Asn
PK84	*hypF Ava*_4602	Hydrogenase maturation protein *HypF*	Asp374Asn
	*ftr Ava*_1641	Ferredoxin-thioredoxin reductase (β-chain)	Val81Ile

vement of *HypF* in the conversion of carbamoyl phosphate to the CO and CN ligands of [NiFe] cluster of hydrogenases. The blockage of such a function must cause repression of the activity of both hydrogenases. Actually, the level of phenazine-methosulfate dependent H_2-uptake and activity of bidirectional hydrogenase measured by methylviologen-dependent H_2 production were significantly decreased in mutant PK84 [7].

3.3. Mutations that Might be Related to the Network of Hydrogen Metabolism

As mentioned above, the genomes of the PK17 and PK84 mutants contain a large number of mutations in various genes (**Tables S3**, **S4**). It is possible that some of these mutations contribute to alterations in H_2 photoproduction. A high level of hydrogen release might be caused by combination of *hupL* and *hypF* deficiency with other mutations that affect ferredoxin-driven electron transfer to nitrogenase and/or activity of enzymes controlling reductant formation. Hypothetically, this proposal concerns missense mutations in two genes, *ftr*, coding for the β-chain of ferredoxin-thioredoxine reductase (**Table 3**), and *Ava*_3964, coding for hypothetical protein without close homologs in the genomes of cyanobacteria as well as in other prokaryotes. They are possible candidates because mutations in only these genes were found in both mutants. The presence of mutations in *ftr* and *Ava*_3946 in these mutants may be just a matter of chance. However, it cannot be excluded that fixation specifically of these mutations in the genomes of both mutants resulted from a particular selection procedure for a decreased ability to reutilize molecular hydrogen produced by nitrogenase. In this case, the consideration of a possible involvement of *ftr* and *Ava*_3964 in the control of H_2-production deserves attention.

Ferredoxin-thioredoxin reductase (FTR) reduces thioredoxins through a disulfide-dithiol interexchange system. Thioredoxins participate in the regulation of several key enzymes of carbon metabolism (including the Calvin-Benson cycle) and some proteins that control redox status and, thus, are involved in the network of hydrogen metabolism [24,25]. Reduced thioredoxin activates uptake hydrogenase at the post-translational level [26]. It was also known that the disulfide redox cascade participates in the assembly of the hydrogenase maturation complex [27]. The active site of the FTR β-chain (13 kDa) comprises a [FeS] cluster and an adjacent redox-active disulfide responsible for interaction with thioredoxins. A mutation in the FTR protein of mutant PK84 was found at position 81 (substitution of valine to isoleucine) within the highly conserved part of the β-chain (**Figure S3**, **Table S7**). Mutation in PK17 occurred at position 12 (substitution of negative charged aspartic acid to polar asparagine), which is also localised in the conserved part. It is possible that both mutations in different sites of FTR can potentially lead to a lack/decrease in the activity of FTR, thus, they affect a regulatory function of thioredoxins and correspondingly a gene-metabolic network of hydrogen production.

In order to examine a hypothesis about a connection between the functions of the *ftr* and H_2-photoproduction in *A. variabilis*, experiments for the construction and studies of mutants impaired in this gene and double mutants containing also *hupL*$^-$ or *hypF*$^-$ mutations have been started.

4. Acknowledgements

This work was supported by the Ministry of Education

and Science of Russian Federation (contract 8272) and a Grant for Leading Scientific Schools "NSh-376.2012.4".

REFERENCES

[1] R. C. Prince and H. S. Kheshgi, "The Photobiological Production of Hydrogen: Potential Efficiency and Effectiveness as a Renewable Fuel," *Critical Reviews in Microbiology*, Vol. 31, No. 1, 2005, pp. 19-31. doi:10.1080/10408410590912961

[2] D. Dutta, D. De, S. Chaudhuris and S. K. Bhattacharya, "Hydrogen Production by Cyanobacteria," *Microbial Cell Factories*, Vol. 4, 2005, pp. 36-46. doi:10.1186/1475-2859-4-36

[3] P. Tamagnini, E. Leitao, P. Oliveira, D. Ferreira, F. A. L. Pinto, D. J. Harris, T. Heidorn and P. Lindblad, "Cyanobacterial Hydrogenases: Diversity, Regulation and Application," *FEMS Microbiology Reviews*, Vol. 31, No. 6, 2007, pp. 692-720. doi:10.1111/j.1574-6976.2007.00085.x

[4] H. Bothe, O. Schmitz, M. G. Yates and W. E. Newton, "Nitrogen Fixation and Hydrogen Metabolism in Cyanobacteria," *Microbiology Molecular Biology Reviews*, Vol. 74, No. 4, 2010, pp. 529-551. doi:10.1128/MMBR.00033-10

[5] S. V. Shestakov and L. E. Mikheeva, "Genetic Control of Hydrogen Metabolism in Cyanobacteria," *Russian Journal of Genetics*, Vol. 42, No. 11, 2006, pp. 1272-1284. doi:10.1134/S1022795406110093

[6] L. E. Mikheeva, O. A. Koksharova and S. V. Shestakov, "Mutant of the Cyanobacterium *Anabaena variabilis* ATCC 29413 that Produces Hydrogen," *Vestnik of Moscow University*, *Section Biology*, No. 2, 1994, pp. 54-57 (in Russian).

[7] L. E. Mikheeva, O. Schmitz, S. V. Shestakov and H. Bothe, "Mutants of the Cyanobacterium *Anabaena variabilis* Altered in Hydrogenase Activity," *Zeitschrift für Naturforschung*, Vol. 50C, 1995, pp. 505-510.

[8] V. B. Borodin, A. Tsygankov, K. K. Rao and D. O. Hall, "Hydrogen Production by *Anabaena variabilis* PK84 under Simulated Outdoor Conditions," *Biotechnology and Bioengineering*, Vol. 69, No. 5, 2000, pp. 479-485. doi:10.1002/1097-0290(20000905)69:5<478::AID-BIT2>3.0.CO;2-L

[9] J. Liu, V. E. Bukatin and A. A. Tsygankov, "Light Energy Conversion into H_2 by *Anabaena variabilis* Mutant PK84 Dense Cultures Exposed to Nitrogen Limitation," *International Journal of Hydrogen Energy*, Vol. 31, No. 11, 2006, pp. 1591-1596. doi:10.1016/j.ijhydene.2006.06.025

[10] T. Happe, K. Schutz and H. Bohme, "Transcriptional and Mutational Analysis of the Uptake Hydrogenase of the Filamentous Cyanobacterium *Anabaena variabilis*," *Journal of Bacterioogy*, Vol. 182, No. 6, 2000, pp. 1624-1631. doi:10.1128/JB.182.6.1624-1631.2000

[11] H. Masukawa, M. Mochimaru and H. Sakurai, "Disruption of the Uptake Hydrogenase Gene, But Not of the Bidirectional Hydrogenase Gene, Leads to Enhanced Photobiological Hydrogen Production by Nitrogen-Fixing Cyanobacterium *Anabaena* sp. PCC 7120," *Applied Microbiology and Biotechnology*, Vol. 58, No. 5, 2002, pp. 618-624. doi:10.1007/s00253-002-0934-7

[12] P. Lindberg, K. Schutz, T. Happe and P. Lindblad, "A Hydrogen-Producing Hydrogenase-Free Mutant Strain of *Nostoc punctiforme* ATCC 29133," *International Journal of Hydrogen Energy*, Vol. 27, No. 11-12, 2002, pp. 1291-1296. doi:10.1016/S0360-3199(02)00121-0

[13] P. Lindblad, K. Christensson, P. Lindberg, A. Fedorov, F. Pinto and A. Tsygankov, "Photoproduction of H_2 by Wild Type *Anabaena* PCC 7120 and a Hydrogen Uptake Deficient Mutant: From Laboratory Experiments to Outdoor Culture," *International Journal of Hydrogen Energy*, Vol. 27, No. 11-12, 2002, pp. 1271-1281. doi:10.1016/S0360-3199(02)00111-8

[14] F. Yoshino, H. Ikeda, H. Masukawa and H. Sakurai, "High Photobiological Hydrogen Production Activity at *a Nostoc* sp. PCC 7422 Uptake Hydrogenase-Deficient Mutant with High Nitrogenase Activity," *Marine Biotechnology*, Vol. 9, No. 1, 2007, pp. 101-112. doi:10.1007/s10126-006-6035-3

[15] R. Rippka, J. Deruelles, B. Waterbury, M. Herdman and R. Y. Stanier, "Generic Assignment, Strain Histories and Properties of Pure Cultures of Cyanobacteria," *Journal of General Microbiology*, Vol. 11, No. 1, 1979, pp. 1-61. doi:10.1099/00221287-111-1-1

[16] A. V. Mardanov, A. V. Beletsky, V. M. Gumerov, E. A. Karbysheva and L. E. Mikheeva, "The New Low-Copy Number Plasmid of Cyanobacterium *Anabaena variabilis*," *Genetika*, Vol. 49, No. 8, 2013, pp. 921-929 (in Russian).

[17] Y. Kanesaki, Y. Shiwa, N. Tajima, M. Suzuki, S. Watanabe, N. Sato, M. Ikeuchi and H. Yoshikawa, "Identification of Substrain-Specific Mutations by Massively Parallel Whole-Genome Resequencing of *Synechocystis* sp. PCC 6803," *DNA Research*, Vol. 19, No. 1, 2012, pp. 67-79. doi:10.1093/dnares/dsr042

[18] R. Wunschiers, M. Batur and P. Lindblad, "Presence and Expression of Hydrogenase Specific C-Terminal Endopeptidases in Cyanobacteria," *BMC Microbiology*, Vol. 3, 2003, pp. 8-20. doi:10.1186/1471-2180-3-8

[19] B. Fodor, G. Rakhely, A. Kovacs and K. L. Kovacs, "Transposon Mutagenesis in Purple Sulfur Photosynthetic Bacteria: Identification of *hypF*, Encoding a Protein Capable of Processing [NiFe] Hydrogenases in Alpha, Beta, and Gamma Subdivisions of the Proteobacteria," *Applied and Environmental Microbiology*, Vol. 67, No. 6, 2001, pp. 2476-2483. doi:10.1128/AEM.67.6.2476-2483.2001

[20] D. Hoffmann, K. Gutekunst, M. Klissenbauer, R. Schulz-Friedrich and J. Appel, "Mutagenesis of Hydrogenase Accessory Genes of *Synechocystis* sp. PCC 6803. Additional Homologues of hypA and hypB are not Active in Hydrogenase Maturation," *FEBS Journal*, Vol. 273, No. 19, 2006, pp. 4516-4527. doi:10.1111/j.1742-4658.2006.05460.x

[21] M. Wakai, S. Davar, H. Masukawa and H. Sakurai, "Effects of Disruption of *HypF* Gene on Photobiological Hydrogen Production in *Anabaena* sp. PCC 7120," *Abstracts of* 13*th International Congress on Photosynthesis*,

Montréal, 28 August-3 september 2004, p. 769.

[22] S. Petkun, R. Shi, Y. Li, A. Asinas, C. Munger, L. Zhang, M. Waclawek, B. Soboh, G. Sawers and M. Cygler, "Structure of Hydrogenase Maturation Protein *HypF* Reaction Intermediates Shows Two Active Sites," Structure, Vol. 19, No. 12, 2011, pp. 1773-1783. doi:10.1016/j.str.2011.09.023

[23] K. A. Harris, V. Jones, Y. Bilbille, M. A. Swairjo and P. F. Agris, "YrdC Exhibits Properties Expected of a Subunit for a tRNA Threonylcarbamoyl Transferase," *RNA*, Vol. 17, No. 9, 2011, pp. 1678-1687. doi:10.1261/rna.2592411

[24] S. Dai, C. Schwendtmayer, K. Johansson, S. Ramaswamy, P. Schurmann and H. Eklund, "How Does Light Regulate Chloroplast Enzymes? Structure—Function Studies of the Ferredoxin/Thioredoxin System," *Quarterly Reviews of Biophysics*, Vol. 33, No. 1, 2000, pp. 67-108. doi:10.1017/S0033583500003607

[25] M. Lindal and F. J. Florencio, "Thioredoxin-Linked Processes in Cyanobacteria Are as Numerous as in Chloroplasts, But Targets Are Different," *Proceedings of National Academy Sciences of USA*, Vol. 100, No. 26, 2003, pp. 16107-16112. doi:10.1073/pnas.2534397100

[26] H. Papen, T. Kentemich, T. Schmulling and H. Bothe, "Hydrogenase Activities in Cyanobacteria," *Biochemie*, Vol. 68, No. 1, 1986, pp. 121-132. doi:10.1016/S0300-9084(86)81077-X

[27] S. Watanabe, R. Matsumi, T. Arai, H. Atomi, T. Imanaka and K. Miki, "Crystal Structures of [NiFe] Hydrogenase Maturation Proteins HypC, HypD, and HypE: Insights into Cyanation Reaction by Thiol Redox Signaling," *Molecular Cell*, Vol. 27, No. 1, 2007, pp. 29-40. doi:10.1016/j.molcel.2007.05.039

Supplementary Material

Table S1. The nitrogenase activity and H_2-evolution in the wild-type PW strain and mutants PK 17 and PK 84 in the course of nitrogenase derepression[*].

Strain	C_2H_4-formation and H_2-evolution activities in nitrogen-free medium (μmol·mg protein^{-1}·h^{-1})[**]			
	$Bg11_0$ medium		$Bg11_0$ medium + 5 mM fructose	
	C_2H_4	H_2	C_2H_4	H_2
Wild-type	1.75	0.03	2.92	0.09
PK 84	1.82	1.75	2.60	2.36
PK 17	1.52	1.67	2.45	2.24

[*]After 5 days of growth in the medium with 20 mM $NaNO_3$, the cells were transferred into N-free medium with or without fructose and incubated under the light for 24 h in flasks. Activities were determined after an additional 20-h incubation under argon in 14-ml hermetic glass reaction vials with 2-ml samples of cultures. [**]Nitrogenase activity was measured by an acetylene-reduction assay with 10% acetylene (v/v). H_2-evolution was measured in parallel probes without acetylene. The gas phase was analysed by gas chromatography as described earlier [6].

Table S2. Identified differences between nucleotide sequences of wild-type *Anabaena variabilis* ATCC 29413 (strain PW) and the reference genome of *Anabaena variabilis* ATCC 29413 (strain TT).

Position in reference genome	Refer. nt.	Nt. in WT	Refer. A.A	A.A in WT	Gene ID and annotation
			Chromosome		
			non-synonymous substitutions		
1229795	C	A	R	I	Ava_1020 circadian oscillation regulator
1278583	CC	-	P	Q	Ava_1048 Metallophosphoesterase
1750901	T	G	I	S	Ava_1419 Protein of unknown function DUF152
2607807	G	A	T	I	Ava_2109 serine/threonine protein kinase with Chase2 sensor
2673156	A	G	*	W	Ava_2160 conserved hypothetical protein
2673883	C	A	C	*	Ava_2161 Peptidase U62, modulator of DNA gyrase
4652055	T	G	Y	D	Ava_3737 Filamentous hemagglutinin-like protein
4832872	G	A	R	C	Ava_3878 possible Sensor with Chase2 domain
5665083	G	T	S	I	Ava_4518 two component transcriptional regulator, LuxR family
			substitutions in non-coding regions		
3553286	T	C			
3934464	T	C			
3553286	T	C			
4823571	-	A			
5496561	C	T			
			Plasmid A		
127716	T	C	C	R	Ava_B0112 Recombination protein MgsA

[*]mutation resulted in appearance of stop codon.

Table S3. Identified differences between nucleotide sequences of mutant PK17 (strain PW) and the reference genome of *Anabaena variabilis* ATCC 29413 (strain TT).

Position in reference genome	Refer. nt.	Nt. in RK17	Refer. A.A	A.A in RK17	Gene ID and annotation
			Chromosome		
			non-synonymous substitutions		
174337	G	A	G	D	Ava_0130 Peptidase M48, Ste24p
464675	G	A	E	K	Ava_0360 D-alanine—D-alanine ligase
506643	C	T	R	H	Ava_0391 conserved hypothetical protein
508779	C	T	V	I	Ava_0393 Protein of unknown function DUF815
603880	C	T	A	T	Ava_0481 conserved hypothetical protein
808405	G	T	G	V	Ava_0641 SSU ribosomal protein S12P methylthiotransferase
949894	C	T	A	V	Ava_0776 conserved hypothetical protein
1073596	C	T	A	V	Ava_0877 conserved hypothetical protein
1270060	C	T	E	K	Ava_1044 Polysaccharide biosynthesis protein
1270267	C	T	V	I	Ava_1044 Polysaccharide biosynthesis protein
1291056	C	T	G	D	Ava_1060 Protein of unknown function DUF820
1294646	G	A	E	K	Ava_1063 Predicted transcriptional regulator
1410190	C	T	R	*	Ava_1152 conserved hypothetical protein
1421476	G	A	S	N	Ava_1161 conserved hypothetical protein
1558179	C	T	E	K	Ava_1267 Phospho-N-acetylmuramoyl-pentapeptide-transferase
1588926	G	A	D	N	Ava_1294 glutamate synthase (ferredoxin)
1656294	C	T	T	I	Ava_1343 PAS/PAC sensor hybrid histidine kinase
2062288	C	T	D	N	Ava_1641 Ferredoxin thioredoxin reductase, beta chain
2062711	G	A	T	I	Ava_1642 Protein of unknown function DUF58
2141901	C	T	A	V	Ava_1708 Phosphoesterase, RecJ-like protein
2208923	G	A	R	*	Ava_1772 conserved hypothetical protein
2431239	C	T	S	N	Ava_1955 ABC transporter, transmembrane region
2457440	G	A	A	T	Ava_1980 Serine/Threonine protein kinase and Signal Transduction Histidine Kinase (STHK) with GAF sensor
2457599	G	A	E	K	Ava_1980 Serine/Threonine protein kinase and Signal Transduction Histidine Kinase (STHK) with GAF sensor
2616522	G	A	A	V	Ava_2115 phenylalanyl-tRNA synthetase, alpha subunit
2622233	G	A	A	T	Ava_2120 conserved hypothetical protein
2869577	C	T	A	V	Ava_2308 NADH dehydrogenase subunit M
2889880	G	A	P	S	Ava_2328 GAF sensor signal transduction histidine kinase
2954008	G	A	G	E	Ava_2372 exonuclease RecJ
2963964	C	-	L	F	Ava_2384 conserved hypothetical protein
2969219	C	T	E	K	Ava_2389 Glycosyl transferase, family 2
3024252	C	T	S	F	Ava_2434 Glycosyl transferase, group 1
3049359	G	A	P	S	Ava_2457 conserved hypothetical protein

Continued

3051104	G	A	A	V	Ava_2458 Amino acid-binding ACT
3058767	G	A	A	V	Ava_2463 conserved hypothetical protein
3060762	C	T	A	T	Ava_2466 Periplasmic Sensor Signal Transduction Histidine Kinase
3073678	G	A	S	N	Ava_2477 phosphate ABC transporter substrate-binding protein, PhoT family
3158185	C	T	A	T	Ava_2557 Secretion protein HlyD
3195255	G	A	L	F	Ava_2586 DevC protein
3244399	C	T	P	S	Ava_2620 Phycobilisome protein
3251051	C	T	L	F	Ava_2629 WD-40 repeat-containing protein
3471049	C	T	A	T	Ava_2808 TPR repeat protein
3473303	G	A	A	T	Ava_2809 serine/threonine protein kinase
3601538	G	A	G	D	Ava_2897 Cobalt transport protein
3762828	G	A	G	S	Ava_3037 tRNA-i(6)A37 thiotransferase enzyme MiaB
3820362	C	T	L	F	Ava_3078 WD-40 repeat-containing protein
3919219	G	A	T	M	Ava_3157 conserved hypothetical protein
3999129	C	T	G	S	Ava_3216 2-isopropylmalate synthase
4341283	G	A	G	D	Ava_3499 Restriction modification system DNA specificity domain protein
4652055	T	G	Y	D	Ava_3737 Filamentous hemagglutinin-like protein
4659759	G	A	A	T	Ava_3744 conserved hypothetical protein
4673669	C	T	E	K	Ava_3753 Sucrose synthase, glycosyl transferase, Group 1
4674485	C	T	G	S	Ava_3753 Sucrose synthase, glycosyl transferase, Group 1
4740651	G	A	G	D	Ava_3806 HAD-superfamily hydrolase subfamily IIIA
4906799	C	T	P	S	Ava_3957 NifZ
4913200	C	T	V	I	Ava_3964 hypothetical protein
4928504	G	A	R	H	Ava_3975 Cache sensor hybrid histidine kinase
4933213	G	A	P	L	Ava_3980 3-deoxy-D-arabinoheptulosonate-7-phosphate synthase
5148709	C	T	D	N	Ava_4132 conserved hypothetical protein
5148906	C	T	G	D	Ava_4132 conserved hypothetical protein
5159021	C	T	A	V	Ava_4137 TPR repeat protein
5420964	C	T	T	I	Ava_4326 PAS/PAC sensor hybrid histidine kinase
5451653	C	T	V	I	Ava_4345 multi-sensor signal transduction histidine kinase
5454912	C	T	V	I	Ava_4348 UvrB/UvrC protein
5581912	G	A	S	F	Ava_4457 GAF sensor signal transduction histidine kinase
5750505	C	T	C	Y	Ava_4595 Nickel-dependent hydrogenase, large subunit
5895345	C	T	A	T	Ava_4716 Serine/Threonine protein kinase and Signal Transduction Histidine Kinase (STHK) with GAF sensor
5956686	C	T	A	V	Ava_4745 Beta-ketoacyl synthase
5966766	C	T	P	L	Ava_4747 Beta-ketoacyl synthase
6248240	C	T	S	N	Ava_4967 TonB-dependent receptor
6322227	G	A	V	M	Ava_5032 Pentapeptide repeat protein

Continued

			synonymous substitutions		
28580	G	A	Q	Q	Ava_0030 DNA Topoisomerase I
711078	G	A	V	V	Ava_0566 conserved hypothetical protein
1516837	G	A	N	N	Ava_1231 transcriptional regulator, TetR family
2227612	C	T	K	K	Ava_1785 Succinylglutamate desuccinylase/aspartoacylase
2642064	C	T	V	V	Ava_2136 conserved hypothetical protein
3544391	G	A	H	H	Ava_2854 DNA Topoisomerase IV subunit A
4211143	C	T	I	I	Ava_3387 conserved hypothetical protein
4307688	C	T	V	V	Ava_3474 Uroporphyrinogen-III Synthase/Uroporphyrinogen-III C-methyltransferase
4315777	A	G	E	E	Ava_3481 methionyl-tRNA formyltransferase
4331354	C	T	L	L	Ava_3491 hypothetical protein
4930151	G	A	K	K	Ava_3976 Response Regulator Receiver Signal Transduction Histidine Kinase
5194177	G	A	S	S	Ava_4160 VCBS
5446110	G	A	I	I	Ava_4342 ABC transporter, transmembrane region
5519033	G	A	R	R	Ava_4402 conserved hypothetical protein
5605419	C	T	G	G	Ava_4476 Predicted signal transduction protein Containing Nacht domain
5951115	G	A	S	S	Ava_4742 Beta-ketoacyl synthase
6232751	C	T	S	S	Ava_4953 molybdopterin synthase subunit MoaD
			substitutions in non-coding regions		
1092351	C	T			
2492604	C	T			
3355310	C	T			
3450066	G	A			
4560437	G	A			
4923183	G	A			
5011152	A	G			
5565011	G	A			
5722150	C	T			
			Plasmid A		
71661	G	A			
99793	C	T	Y	Y	Ava_B0090 conserved hypothetical protein
117956	G	A	P	L	Ava_B0106 ATPase, E1-E2 type
127716	T	C	C	R	Ava_B0112 Recombination protein MgsA
128987	G	A	R	R	Ava_B0112 Recombination protein MgsA
223097	G	A	R	C	Ava_B0199 hypothetical protein
277355	C	T	D	N	Ava_B0245 hypothetical protein
290839	G	A	P	S	Ava_B0266 conserved hypothetical protein
			Plasmid C		
55486	C	T	A	V	Ava_C0035 conserved hypothetical protein
260242	C	T	T	I	Ava_C0223 conserved hypothetical protein

*mutation resulted in appearance of stop codon.

Table S4. Identified differences between nucleotide sequences of mutant PK84 (strain PW) and the reference genome of *Anabaena variabilis* ATCC 29413 (strain TT).

Position in reference genome	Refer. nt.	Nt. in RK84	Refer. A.A	A.A in RK84	Gene ID and annotation
			Chromosome		
			non-synonymous substitutions		
345125	C	T	A	V	Ava_0267 conserved hypothetical protein
419871	C	T	A	T	Ava_0323 probable glycosyl transferase
510357	C	T	A	V	Ava_0394 Protein of unknown function UPF0136
567360	C	T	A	V	Ava_0442 conserved hypothetical protein
571957	G	A	A	T	Ava_0445 4Fe-4S ferredoxin, iron-sulfur binding protein
695118	G	A	V	M	Ava_0549 RNAse R
796132	G	A	R	W	Ava_0632 Protein of unknown function DUF512
812693	C	T	D	N	Ava_0644 Aldo/keto reductase
924795	C	T	A	T	Ava_0748 NADH dehydrogenase subunit M
937783	G	A	S	L	Ava_0762 serine/threonine protein kinase
1221405	G	A	G	D	Ava_1011 Short-chain dehydrogenase/reductase SDR
1251578	C	T	L	F	Ava_1029 multi-sensor hybrid histidine kinase
1406576	G	A	E	K	Ava_1149 histidine kinase
1418275	G	A	A	V	Ava_1158 Restriction modification system DNA specificity domain protein
1610249	G	A	V	I	Ava_1311 Plasmid stabilization system
1615402	G	A	T	I	Ava_1316 Heat shock protein Hsp70
1759217	C	T	T	I	Ava_1425 UvrD/REP helicase
1853436	C	T	E	K	Ava_1505 SSU ribosomal protein S1P
1970927	G	A	A	V	Ava_1595 conserved hypothetical protein
2062081	C	T	V	I	Ava_1641 Ferredoxin thioredoxin reductase, beta chain
2283176	C	T	S	F	Ava_1836 glutathione synthase
2326042	T	C	V	A	Ava_1873 Peptidoglycan-binding domain 1
2401227	C	T	G	S	Ava_1932 hypothetical protein
2552940	GCCA	CC	HM	HR	Ava_2063 conserved hypothetical protein
2647889	C	T	A	V	Ava_2140 Peptidase S8 and S53, subtilisin, kexin, sedolisin
2676532	C	T	T	I	Ava_2162 Cl-channel, voltage gated
2872203	C	T	A	V	Ava_2311 Prolyl-tRNA synthetase
3207952	G	A	R	H	Ava_2591 Short-chain dehydrogenase/reductase SDR
3332996	C	T	A	V	Ava_2696 ATP-dependent DNA helicase RecQ
3400890	G	A	R	C	Ava_2759 Pentapeptide repeat protein
3439699	C	T	A	V	Ava_2787 Ser-tRNA(Thr) hydrolase/threonyl-tRNA synthetase
3490779	C	T	A	T	Ava_2819 8-amino-7-oxononanoate synthase
3610790	C	T	L	F	Ava_2905 Polysaccharide pyruvyl transferase

Continued

3611462	C	T	P	S	Ava_2905 Polysaccharide pyruvyl transferase
3632773	C	T	A	V	Ava_2926 Phycobilisome protein
3643490	G	A	G	D	Ava_2939 Phycobilisome linker polypeptide
3655425	G	A	L	F	Ava_2949 LytB protein
3673448	G	A	G	E	Ava_2964 Protein of unknown function DUF1400
4099729	C	T	S	F	Ava_3286 ABC transporter-like protein
4107800	G	A	P	S	Ava_3293 Protein of unknown function DUF820
4194155	G	A	E	K	Ava_3370 conserved hypothetical protein
4441148	C	T	P	S	Ava_3566 conserved hypothetical protein
4533131	C	T	T	I	Ava_3636 phosphoribosyl formylglycinamidine synthase Subunit II
4605170	G	A	S	N	Ava_3695 L-aspartate aminotransferase apoenzyme
4652055	T	G	Y	D	Ava_3737 Filamentous hemagglutinin-like protein
4691689	C	T	L	F	Ava_3769 phosphoglucosamine mutase
4912914	C	T	G	D	Ava_3964 hypothetical protein
5002719	G	A	S	N	Ava_4018 conserved hypothetical protein
5119624	G	A	A	T	Ava_4109 Amino acid adenylation
5367385	G	A	R	K	Ava_4281 conserved hypothetical protein
5587287	G	A	A	V	Ava_4463 Apolipoprotein N-acyltransferase
5757914	G	A	D	N	Ava_4602 Hydrogenase maturation protein *HypF*
5785521	G	A	A	V	Ava_4625 GTP-binding protein, HSR1-related protein
5961512	G	A	W	*	Ava_4746 Amino acid adenylation
6239037	G	A	S	N	Ava_4960 conserved hypothetical protein
			synonymous substitutions		
83431	C	T	R	R	Ava_0064 signal transduction histidine kinase
263186	G	A	G	G	Ava_0202 conserved hypothetical protein
410378	G	A	Q	Q	Ava_0316 Peptidase M1, membrane alanine aminopeptidase
445649	C	T	S	S	Ava_0346 TPR repeat protein
454958	C	T	G	G	Ava_0354 protein serine/threonine phosphatase
525567	G	A	S	S	Ava_0409 Protein of unknown function UPF0118
591515	C	T	D	D	Ava_0465 amino acid/amide ABC transporter substrate-binding protein, HAAT family
798463	C	T	S	S	Ava_0634 conserved hypothetical protein
891335	G	A	P	P	Ava_0721 bacterial peptide chain release factor 1 (bRF-1)
1216586	G	A	C	C	Ava_1007 imidazole glycerol phosphate synthase subunit hisF
1227413	G	A	G	G	Ava_1016 circadian clock protein KaiC
1414921	G	A	N	N	Ava_1156 Type I site-specific deoxyribonuclease HsdR
1445420	G	A	D	D	Ava_1175 histidine kinase
1616559	C	T	Q	Q	Ava_1316 Heat shock protein Hsp70

Continued

2291343	C	T	R	R	Ava_1841 WD-40 repeat-containing protein
2310976	C	T	P	P	Ava_1856 Glycosyl hydrolase, BNR repeat protein
2552949	A	G	G	G	Ava_2063 conserved hypothetical protein
2620397	C	T	L	L	Ava_2118 FMN adenylyltransferase/riboflavin kinase
3221837	C	T	L	L	Ava_2599 nucleoside ABC transporter ATP-binding protein
3628098	C	T	L	L	Ava_2922 uroporphyrinogen-III C-methyltransferase
3643218	C	T	S	S	Ava_2939 Phycobilisome linker polypeptide
3917437	C	T	I	I	Ava_3155 conserved hypothetical protein
3920286	G	A	S	S	Ava_3158 conserved hypothetical protein
4488395	G	A	R	R	Ava_3597 O-antigen polymerase
4961502	G	A	P	P	Ava_3989 Major facilitator superfamily MFS_1
5254255	C	T	G	G	Ava_4195 conserved hypothetical protein
5501964	T	C	G	G	Ava_4391 serine/threonine protein kinase
6004344	G	A	N	N	Ava_4776 Mg-protoporphyrin IX monomethyl ester (oxidative) cyclase
			substitutions in noncoding regions		
594288	C	T			
840593	C	T			
1405494	G	A			
1747358	G	A			
2745073	C	T			
3296936	C	T			
3717778	C	T			
4122703	G	A			
4446973	C	T			
4983345	C	T			
5482672	C	T			
6066400	G	A			
6163021	C	T			
			Plasmid A		
57566	C	T	V	V	Ava_B0045 conserved hypothetical protein
127716	T	C	C	R	Ava_B0112 Recombination protein MgsA
219899	G	A	L	L	Ava_B0197 Osmosensitive K+ channel Signal Transduction Histidine Kinase, sensor domain protein
292610	G	A	Q	*	Ava_B0268 conserved hypothetical protein
			Plasmid C		
5478	C	T	S	N	Ava_C0002 Amino acid adenylation
103839	G	A			
211246	G	A			
231797	C	T	A	T	Ava_C0195 conserved hypothetical protein

*mutation resulted in appearance of stop codon.

Table S5. The organisms, proteins and % of identity of enzymes indicated in the Figure S1.

Entry name	Organism	Protein length	Identity, %
Q3M494_ANAVT	*Anabaena variabilis* (strain ATCC 29413/PCC 7937)	531	100
Q8YZ11_NOSS1	*Nostoc* sp. (strain PCC 7120/UTEX 2576)	531	98
O68307_NOSP7	*Nostoc punctiforme* (strain ATCC 29133/PCC 73102)	531	91
Q84GM3_9NOST	*Anabaena siamensis* TISTR 8012	531	86
Q841J7_9CHRO	*Gloeothece* sp. PCC 6909	531	83
B1WTU9_CYAA5	*Cyanothece* sp. (strain ATCC 51142)	539	80
Q846P6_9CYAN	*Lyngbya majuscula* CCAP 1446/4	537	83
Q4BUZ7_CROWT	*Crocosphaera watsonii* WH 8501	531	79
Q10Z54_TRIEI	*Trichodesmium erythraeum* (strain IMS101)	534	78
D3EMX6_UCYNA	Cyanobacterium UCYN-A	531	77
A5E9A2_BRASB	*Bradyrhizobium* sp. (strain BTAi1/ATCC BAA-1182)	532	72
F7YAW7_MESOW	*Mesorhizobium opportunistum* (strain LMG 24607/HAMBI 3007/WSM2075)	532	72

Table S6. The organisms, proteins and % of identity of enzymes indicated in the Figure S2.

Entry name	Organism	Protein length	Identity, %
Q3M487_ANAVT	*Anabaena variabilis* (strain ATCC 29413/PCC 7937)	786	100
Q8YYZ7_NOSS1	*Nostoc* sp. (strain PCC 7120/UTEX 2576)	785	91
B2J5X8_NOSP7	*Nostoc punctiforme* (strain ATCC 29133/PCC 73102)	781	77
D4TFZ5_9NOST	*Cylindrospermopsis raciborskii* CS-505	802	66
Q4A561_9CYAN	*Lyngbya majuscula* CCAP 1446/4	788	65
Q117W3_TRIEI	*Trichodesmium erythraeum* (strain IMS101)	785	63
A5E9B2_BRASB	*Bradyrhizobium* sp. (strain BTAi1/ATCC BAA-1182)	824	52
Q93TY4_THIRO	*Thiocapsa roseopersicina*	806	52
Q20XQ9_RHOPB	*Rhodopseudomonas palustris* (strain BisB18)	785	46
Q4BYT6_CROWT	*Crocosphaera watsonii* WH 8501	768	43
B8HTU1_CYAP4	*Cyanothece* sp. (strain PCC 7425/ATCC 29141)	818	63
A0ZAS0_NODSP	*Nodularia spumigena* CCY 9414	780	74
F7YAV7_MESOW	*Mesorhizobium opportunistum* (strain LMG 24607/HAMBI 3007)	807	51
Q3J0L9_RHOS4	*Rhodobacter sphaeroides* (strain ATCC 17023/2.4.1)	758	43
Q8KX20_SYNP2	*Synechococcus* sp. (strain ATCC 27264/PCC 7002/PR-6)	766	44
B0JKD2_MICAN	*Microcystis aeruginosa* (strain NIES-843)	754	43
C2LGU4_PROMI	*Proteus mirabilis* ATCC 29906	766	41
H4JBR2_ECOLX	*Escherichia coli* DEC1D	750	42

Table S7. The organisms, proteins and % of identity of enzymes indicated in the Figure S3.

Entry name	Organism	Protein length	Identity, %
Q3MCM2_ANAVT	*Anabaena variabilis* (strain ATCC 29413/PCC 7937)	121	100
Q8YPX5_NOSS1	*Nostoc* sp. (strain PCC 7120/UTEX 2576)	121	98
B2IWB5_NOSP7	*Nostoc punctiforme* (strain ATCC 29133/PCC 73102)	124	92
B1WYY2_CYAA5	*Cyanothece* sp. (strain ATCC 51142)	152	76
Q4C4P6_CROWT	*Crocosphaera watsonii* WH 8501	123	74
D4TKN9_9NOST	*Cylindrospermopsis raciborskii* CS-505	122	85
B0JX64_MICAN	*Microcystis aeruginosa* (strain NIES-843)	119	79
Q115H7_TRIEI	*Trichodesmium erythraeum* (strain IMS101)	121	71
B1XQ57_SYNP2	*Synechococcus* sp. (strain ATCC 27264/PCC 7002)	119	70
FTRC_ORYSJ	*Oryza sativa subsp. japonica* (Rice)	146	74
A3QQP9_MESVI	*Mesostigma viride*	136	67
D3ENK0_UCYNA	Cyanobacterium UCYN-A	121	70
FTRC1_SPIOL	*Spinacia oleracea* (Spinach)	144	66
Q6LBH9_MAIZE	*Zea mays* (Maize)	115	72
FTRC_ARATH	*Arabidopsis thaliana* (Mouse-ear cress)	146	66
Q94IK0_SOLTU	*Solanum tuberosum* (Potato)	148	66
FTRC_SOYBN	*Glycine max* (Soybean) (*Glycine hispida*)	144	66
FTRC2_SPIOL	*Spinacia oleracea* (Spinach)	148	65

```
412  RCLREFVLNDPWYIKPKEKDGRGWGATEASRGSLCHWIDIEGGKIKNYQVIAATTWNVGP  471   Q3M494_ANAVT
412  RCLREFVLNDPWYIKPKEKDGRGWGATEASRGSLCHWIDIEGGKIKNYQVIAATTWNVGP  471   Q8YZ11_NOSS1
412  RCLREFVLNDPWYIKPKEKDGRGWGATEASRGSLCHWIDIEGGKIKNYQVIAATTWNIAL  471   O68307_NOSP7
412  HCLREFKLNDPWYIKPKEKDGKGWGATEAARGALCHWVDIEEGKIKQYQVIAPGTWNIGP  471   Q84GM3_9NOST
412  RCLREFQLKEPWYIKPTEKDGRGWGATEASRGALCHWIEVQGGKIKNYQIIAPTTWNVGP  471   Q841J7_9CHRO
420  RCLREFKLNDPWYIKPTEKDGRGWGGTEASRGALCHWVEVEKGKIKNYQVIAPTTWNVGP  479   B1WTU9_CYAA5
418  RCLREFKLREPWYIKPTPKDGRGWGATEASRGALCHWVEIENGKIKNYQILAPSTWNIGP  477   Q846P6_9CYAN
412  RCLREFKLNDPWYIKPTEKDGRGWGGTEASRGALCHWVEVEKGKIKNYQVIAPTTWNVGP  471   Q4BUZ7_CROWT
415  HCLREFNLRDPWYIKPQEKDGRGWGATNASRGSLCHWVEIEGGKINNYQIIAPTTWNVGP  474   Q10Z54_TRIEI
412  RCLREFKLNDPWYIKPKEKDGRGWGGTEASRGALCHWVEVEKGVIKNYQIIAPSTWNIGP  471   D3EMX6_UCYNA
413  HALREFRLNDPWYVKPTEKDGRGWGATEAIRGALCHWIEVQGGKIKNYQIITPTTWNVGP  472   A5E9A2_BRASB
413  HALREFRLNDPWYVKPTEKDGRGWGATEAIRGALCHWIEVQNGKIKNYQIIAPTTWNVGP  472   F7YAW7_MESOW
     :.**** *.:***:**  ***:***.*:* **:****:::: * *::**:::  ***:.

472  RDSEGVRGPIEEALIGTPIEDSRDPVEVGHVARSFDSCLVCTVHAHDAKTGEELARFRTA  531   Q3M494_ANAVT
472  RDSEGVRGPIEEALIGTPIEDSRDPVEVGHVARSFDSCLVCTVHAHDAKTGEELARFRTA  531   Q8YZ11_NOSS1
472  RDGEGIRGPIEEALIGTPIYDSSDPVEVGHVARSFDSCLVCTVHAHDAKTGEELARFRTA  531   O68307_NOSP7
472  RDGAGQRGPVEEALIGTPIEDPTDPVEVGHVARSFDSCLVCTVHAHDAKTGKELARFRTA  531   Q84GM3_9NOST
472  RDGQGVRGPIEEALIGTPVEDLRDPVEVGHVARSFDSCLVCTVHAHNAKTGEELARFRTN  531   Q841J7_9CHRO
480  RDGNHTRGPIEEALVGIPIEDTSNPVEVGHVARSFDSCLVCTVHAHDAKTGKELARFRTN  539   B1WTU9_CYAA5
478  RDGEGVRGPIEEALIGTPIADKNDPVEVGHVARSFDSCLVCTVHAHDAKTGEELARFRTA  537   Q846P6_9CYAN
472  RDGTHTRGPIEEALVGIPIEDSTNPVEVGHVARSFDSCLVCTVHAHDAKTGKELARFRTN  531   Q4BUZ7_CROWT
475  RDGQGVRGPIEEALVGIPIADVNDPVEVGHVARSFDSCLVCTVHAHDAKTGEELARFRTS  534   Q10Z54_TRIEI
472  RDGKDVRGPIEEALIGIPIEDMSNPVEVGHVARSFDSCLVCTVHAHDAKTGKELARFRTN  531   D3EMX6_UCYNA
473  RSDRDELGPIEQALIGTPVADVNDPVEVGHVCRSYDSCLVCTVHAHHASTGKELARFRTA  532   A5E9A2_BRASB
473  RSDGGELGPIEQALIGTPIADPSDPVEVGHVCRSYDSCLVCTVHAHDASTGTELARFRTA  532   F7YAW7_MESOW
     *.     **:*:**:* *: *  :********.**:***********.*.** *******
```

Figure S1. Alignment of the amino acid sequences of the C-terminal regions of various HupL proteins. Amino acids highlighted in grey are analogues of Cys 446 of HupL of *A. variabilis*. Residues that are identical (*), conserved (:), or semiconserved (.) in the sequences are indicated below the sequences. Conserved aminoacids of the active sites are shown bold and underlined. The C-end segment that is cleaved by the HupW-endopeptidase is shown in bold italics. (Abbreviations after enzyme sequences are shown in the Table S5).

Figure S2. Alignment of the amino acid sequences in the middle part of various *HypF* proteins. Amino acids highlighted in grey are analogues of Asp -374 of Hyp F of *A.variabilis*. Residues that are identical (*), conserved (:), or semiconserved (.) in the sequences are indicated below the sequences. (Abbreviations after enzyme sequences are shown in the Table S6).

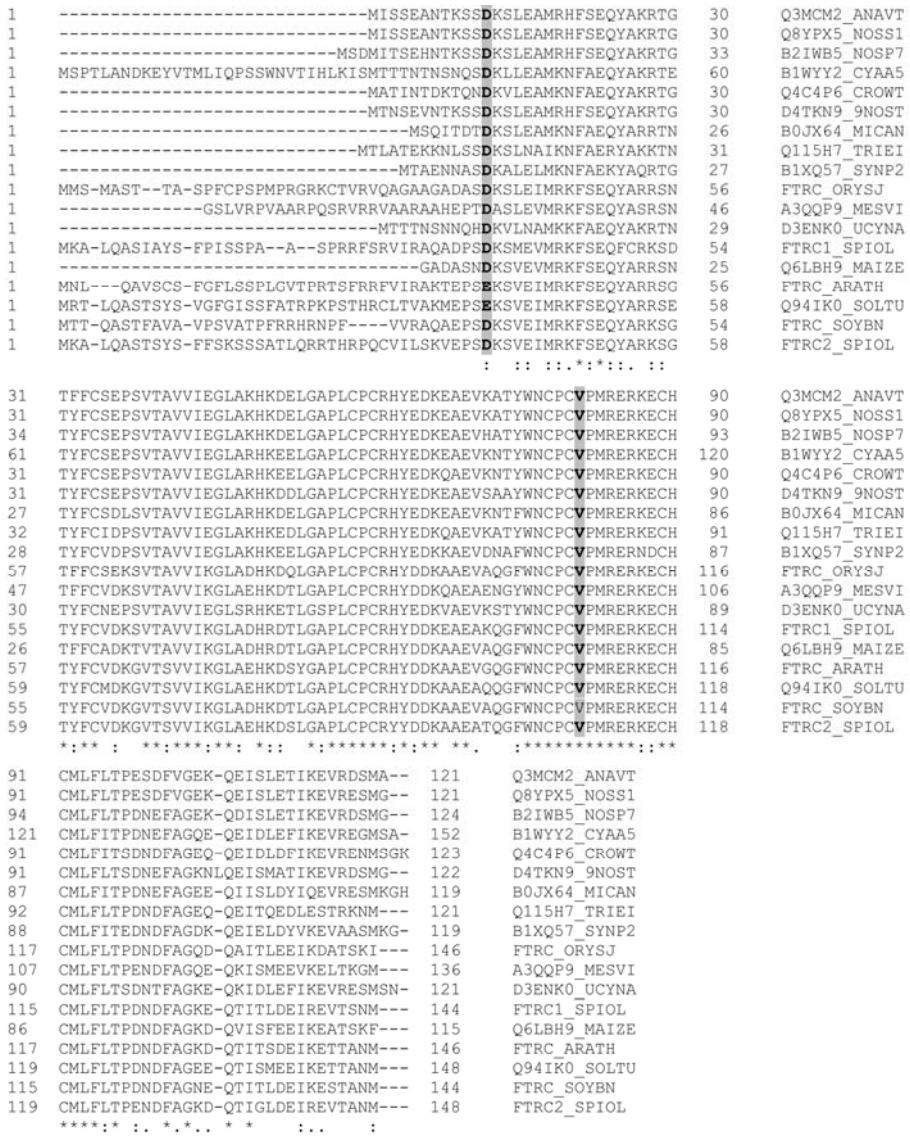

Figure S3. Alignment of the amino acid sequences of various ferredoxin-thioredoxin reductases, beta chain proteins. Amino acids highlighted in grey are analogues of Asp 12 and Val 81 of *A. variabilis* Ftr. Residues that are identical (*), conserved (:), or semiconserved (.) in the sequences are indicated below the sequences. (Abbreviations after enzyme sequences are shown in the Table S7).

