Subsystem: Porphyrin, Heme, and Siroheme Biosynthesis

Svetlana Gerdes, Fellowship for Interpretation of Genomes

Introduction

Tetrapyrroles and their derivatives play an essential role in all living organisms. They are involved in many metabolic processes, such as energy transfer, catalysis, and signal transduction. In eukaryotes, the synthesis of tetrapyrroles is restricted to heme, siroheme, chlorophyll and bilins. Prokaryotes additionally form most complicated tetrapyrroles, such as corrinoids, heme d1 and coenzyme F430. An abundant and ubiquitous representative of this group of compounds is heme, a cyclic tetrapyrrole that contains a centrally chelated Fe. The biosynthetic pathway of heme and siroheme can be arbitrary divided into 4 fragments:

A: Biosynthesis of 5-aminolevulinic acid (ALA), the common precursor of all marcocyclic and linear tetrapyrrolesis, can occur via two alternative unrelated routes: the C5-pathway, or the Shemin pathway. The C5 pathway, found in most bacteria, archaea and plants, starts from the C5-skeleton of glutamate, ligated to tRNAGlu. Some alpha-proteobacteria, fungi, and animals synthesize 5-aminolevulinate via Shemin pathway by condensation of succinyl-CoA with glycine.

B. Universal steps in biosynthesis of tetrapyrroles: condensation of 8 molecules of 5-aminolevulinic acid to form Uroporphyrinogen III (Uro-III) - the first cyclic tetrapyrrole intermediate in the pathway. Universally present, very conserved, variations are extremely rare. The corresponding genes form conserved chromosomal clusters in a large number of genomes. Located at the branchpoint of tetrapyrrole biosynthesis, Uro-III can be converted to both Siroheme (via Uro-III methyltransferase, UroM) and protoporphyrin IX (via Uro-III decarboxylase, UroD). Regulation of Uro-III partitioning of into the two main branches, currently poorly understood, can be a fascinating research topic (see below).

C. The three-step biosynthetic route leading from Uroporphyrinogen III to Siroheme. The iron-chelating siroheme is required for the six-electron transfer reactions during assimilatory nitrite or sulfite reduction (Raux et al. 2003) and the corresponding genes are often co-localized with genes encoding nitrite or sulfite reductases. Siroheme biosynthesis genes often cluster also with genes of ALA and Uro-III biosynthesis (the reason for inclusion of siroheme b-sis branch in this Subsystem).

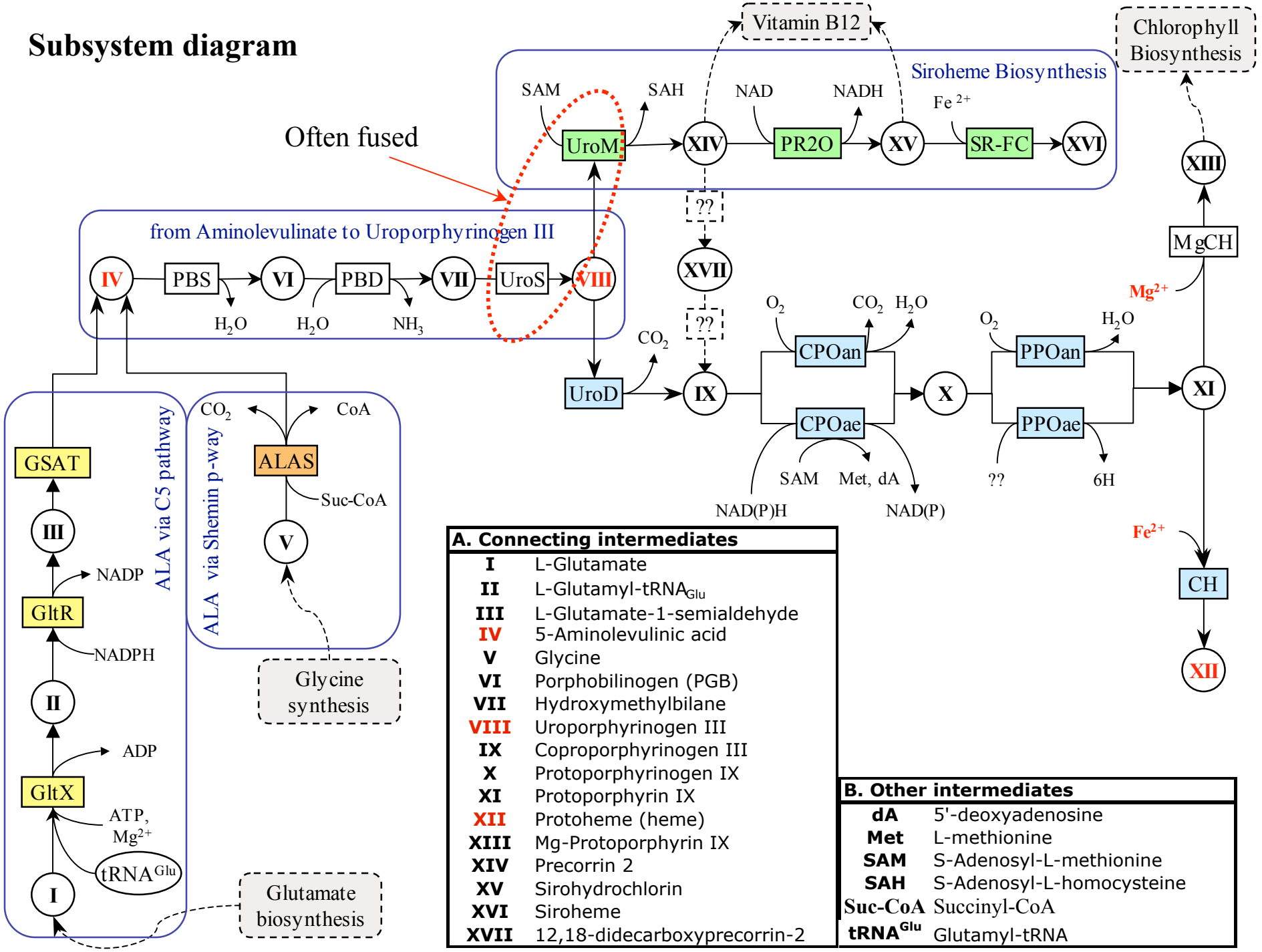
D. Terminal steps of heme biosynthesis: Uroporphyrinogen III to Protoheme. Two forms, one oxygen-dependent and one oxygen-independent are known for each of the two enzymes: Coproporphyrinogen III oxidase (CPO) and Protoporphyrinogen IX oxidase (PPO). Protoporphyrin IX produced in this pathway can incorporate Fe^{2+} or Mg^{2+} in a reaction catalyzed by ferrochelatase or magnesium-chelatase respectively.

A primitive pathway of porphyrin biosynthesis occurs in *Desulfovibrio vulgaris* (ref. 5). It deviates from the known pathway at Uro-III into precorrin-2 and reenters again into coproporphyrinogen III. Apparently, precorrin-2 is converted into 12,18 didecarboxyprecorrin-2 by precorrin-2 decarboxylase. Acetate eliminase subsequently catalyzes the conversion of 12,18-didecarboxyprecorrin-2 to coproporphyrinogen III (ref. 5). This ancient pathway has been replace by the single enzymic process (via UroD - absent in D. *vulgaris*) in the vast majority of other (evolutionary younger) microorganisms.

Functional roles, subsets of roles and alternative forms of enzymes

Subsets of roles	Column	Alternative forms	Abbrev	Functional roles
5-aminolevulinate (ALA) via C5 pathway	1	*glt	**GltX**	Glutamyl-tRNA synthetase (EC 6.1.1.17)
	2		**GltX_m**	Glutamyl-tRNA synthetase, mitochondrial (EC 6.1.1.17)
	3		**GltR**	Glutamyl-tRNA reductase (EC 1.2.1.70)
	4		**GAST**	Glutamate-1-semialdehyde aminotransferase (EC 5.4.3.8)
ALA via Shemin pathway	5		**ALAS**	5-aminolevulinate synthase (EC 2.3.1.37)
5-aminolevulinate to Uroporphyrinogen III	6		**PBS**	Porphobilinogen synthase (EC 4.2.1.24)
	7		**PBD**	Porphobilinogen deaminase (EC 2.5.1.61)
	8		**UroS**	Uroporphyrinogen-III synthase (EC 4.2.1.75)
Uroporphyrinogen III to Siroheme	9		**UroM**	Uroporphyrinogen-III methyltransferase (EC 2.1.1.107)
	10		**PR2O**	Precorrin-2 oxidase (EC 1.3.1.76)
	11		**SR-FC**	Sirohydrochlorin ferrochelatase (EC 4.99.1.4)
Uroporphyrinogen III to Heme	12		**UroD**	Uroporphyrinogen III decarboxylase (EC 4.1.1.37)
	13	*CPO	**CPOae**	Coproporphyrinogen III oxidase, aerobic (EC 1.3.3.3)
	14		**CPOan**	Coproporphyrinogen III oxidase, oxygen-independent (EC 1.3.99.22)
	15	*PPO	**PPOae**	Protoporphyrinogen IX oxidase, aerobic (EC 1.3.3.4)
	16		**PPOan**	Protoporphyrinogen IX oxidase, oxygen-independent, HemG (EC 1.3.-.-)
	17		**FC**	Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1)
?	18		**EcHemY**	Homolog of E. coli HemY protein

Subsystem diagram



A. Connecting intermediates	
I	L-Glutamate
II	L-Glutamyl-tRNA$_{Glu}$
III	L-Glutamate-1-semialdehyde
IV	5-Aminolevulinic acid
V	Glycine
VI	Porphobilinogen (PGB)
VII	Hydroxymethylbilane
VIII	Uroporphyrinogen III
IX	Coproporphyrinogen III
X	Protoporphyrinogen IX
XI	Protoporphyrin IX
XII	Protoheme (heme)
XIII	Mg-Protoporphyrin IX
XIV	Precorrin 2
XV	Sirohydrochlorin
XVI	Siroheme
XVII	12,18-didecarboxyprecorrin-2

B. Other intermediates	
dA	5'-deoxyadenosine
Met	L-methionine
SAM	S-Adenosyl-L-methionine
SAH	S-Adenosyl-L-homocysteine
Suc-CoA	Succinyl-CoA
tRNA^{Glu}	Glutamyl-tRNA

Functional variants

Porphyrin, Heme, and Siroheme Biosynthesis

Organism	Variant Code #	*glt	GltR	GAST	ALAS	PBS	PBD	UroS	UroM	PR2O	SIRFC	UroD	*CPO	*PPO	FC
Candidatus Blochmannia floridanus	910	476-1					546	547	154	154	154				
Buchnera aphidicola str. APS	910	73-1					562	586	405	405	405		524-**14**		
Clostridium botulinum ATCC 3502	910	1009-1				1186	1189	1188	1188	1190	1190		1380-**14**, 2902-**14**, 988-**14**		
Wolbachia pipientis wMel	2099	382-1, 687-1			1135	138	474	876				902	??	??	1047
Wolbachia sp. endosymbiont of Drosophila melanogaster	2019	389-1, 697-1			1154	141	483	892				918	1091-**13**	??	1063
Rickettsia sibirica	2039*	1010-1, 226-1			706	1156	1187	??				641	645-**13**, 429-**14**	??	642
Rickettsia prowazekii str. Madrid E	2039*	312-1, 594-1			793	516	450	??				835	832-**13**, 169-**14**	??	834
Chromobacterium violaceum ATCC 12472	3131	1937-1	79	67	803	1648	54	52	1569, 51, 813	813	813	1122	757-**13**, 3648-**14**, 927-**14**	1653-**15**	2480
Escherichia coli K12	1132	2371-1	1196	155		366	3730	3729	3303, 3728	3303	3303	3910	2404-**13**, 3787-**14**	3776-**16**	471
Bacillus subtilis subsp. subtilis str. 168	1121	92-1	2820	870, 2815		2816	2818	1224, 2817, 329	1563, 329	1565	**1564**	1012	2553-**14**	1014-**15**	1013
Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130	1121	3250-1	3400	3403		3402	3401	??	3274	3273	3273	3406	3407-**14**	3409-**15**	3410
Chlamydophila pneumoniae J138	1021	556-1	711	138		741	53	??				888	378-**14**, 887-**14**	886-**15**	600
Tropheryma whipplei TW08/27	1099	536-1	716	709		710	712	711				715	??	??	713
Prochlorococcus marinus subsp. pastoris MED4	1039	1889-1	1551	1878		232	1865	335	479			1776	591-**13**, 915-**14**	??	1834
Desulfovibrio vulgaris subsp. vulgaris str. Hildenborough [B]	1119	2539-1	1455	3153		852	1882	730	730	1457	1457		3043-**14**	??	

A small fragment of the full **Subsystem Spreadsheet** in SEED is shown. Organisms were selected to illustrate several common operational variants of this subsystem (out of >20 known combinations). Multipositional encoding of functional variants (appearing in **Variant code#** column) is described in the next slide. Missing genes inferred by the functional context analysis are shown by "?". Only three major "missing gene" cases are reflected in this table. Other cases are less frequent. At least some of them are due to "technical problems" (eg incomplete genomes, imperfect ORF detection, etc). Several functional roles (marked with "*") aggregate two alternative enzyme families (as defined in slide 2). The occurrence of a specific form in an organism is shown by a role numbers (shown in black bold font), corresponding to those in slide 2.

Variant codes

(used in subsystem spreadsheet above)

The first digits in a multipositional variant code reflects the type of ALA biosynthesis present in an organism:

[1___]: 5-aminolevulinic acid biosynthesis via C5-pathway

[2___]: 5-aminolevulinic acid biosynthesis via Shemin pathway

[3___]: both pathways are present in an organism

[9___]: no known pathway can be asserted in an organism

The second digit shows the presence or absence of Uro-III to Siroheme pathway

[_1__]: siroheme biosynthesis can be asserted,

[_0__]: siroheme biosynthesis cannot be asserted

Variations in the terminal steps of heme biosynthesis (Uro-III to heme) are encoded in the 3d and 4d digits of variant codes.

The third digit describes the type of coproporphyrinogen III oxidase present in a genome:

[__1_]: CPOae, oxygen-dependent cop-III oxidase (EC 1.3.3.3, HemF)

[__2_]: CPOan, oxygen-independent cop-III oxidase (HemN or/and HemZ)

[__3_]: Both forms of CPO are present in an organism.

[__9_]: Both known forms of CPO oxidase are missing.

The fourth digit reflects the type of protoporphyrinogen oxidase (PPO), catalyzing the penultimate step in heme biosynthsis:

[___1]: Organisms containing oxygen-dependent PPOae

[___2]: Organisms containing oxygen-independent PPOan

[___3]: Both forms of PPO, are present in an organism (uncommon)

[___9]: Both known genes for PPO are missing

Open problems, comments

A. The 5-aminolevulinic acid (ALA) biosynthesis

Organisms in the Subsystem spreadsheet in previous slide are grouped by the type of 5-aminolevulinic acid biosynthesis they perform. The top group have no known route of ALA production, hence, must depend on exogenous porphobilinogen (e.g. *Candidatus, Buchnera*) or 5-aminolevulinic acid (as in *C. botulinum*). The loss of ALA biosynthesis enzymes GltR and GAST in *C. botulinum* is an apparently recent event: the corresponding genes are present in other Clostridia - notably, within the same cluster of Uro-III biosynthestic genes. Second group of organisms utilizes Shemin pathway, third - the C5-pathway. *Chromobacterium violaceum* is a rare example of co-occurrence of both pathways.

B. ALA to Uro-III universal pathway

Variations in the ubiquitously present universal pathway are very rare. However, gene encoding Uroporphyrinogen-III synthase (UroS) is missing in a number of genomes, including those of *Caulobacter, Rickettsiae, Leptospira interrogans, Acinetobacter sp., Cytophaga hutchinsonii*, all *Chlamydia and Chlamydophila*, etc. (see SS Spreadsheet in previous slide)

C. Siroheme pathway

Occurrence of fusion events among the genes of porphinoids (=reduced cyclic tetrapyrroles: vitamin B12, siroheme, coenzyme F430, heme d1) biosynthesis is anomaly high. The first dedicated enzyme of the porphinoids branch, UroIII methyltransferase (UroM) is very often fused with UroIII synthase (UroS), catalyzing the previous step.

Open problems, comments continued

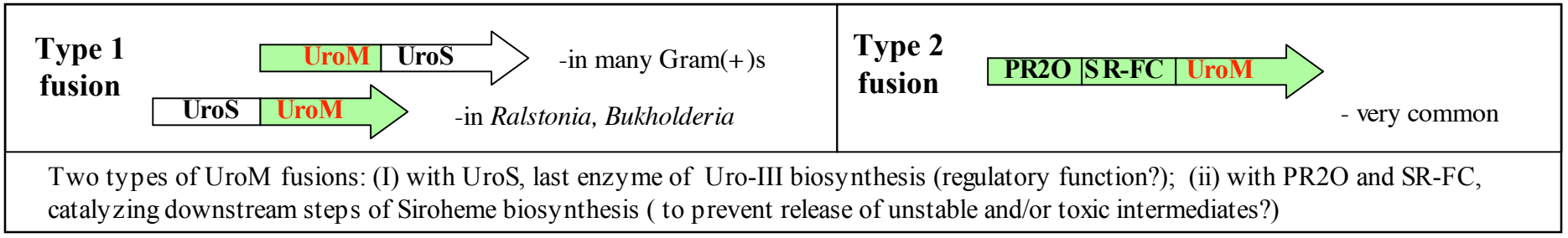
C. Siroheme pathway continued. Interestingly, while in many Gram-Positive bacteria (*Bacillus, Clostridia, Listeria, Fusobacteria*, etc) the order of the two domains in a fused protein is as follows: Uro-III-methyltransferase>>>Uro-III-synthase; it is reversed in *Burkholderiaceae*, indicating that this fusion has occurred independently at least twice in evolution.

The second type of fusion involving UroS is the fusion of all the three steps of sirohaem biosynthesis (methylation, oxidation, Fe-chelation) into a single protein, siroheme synthase, in the majority of siroheme-containing organisms. Interestingly, very often several copies of UroM are present in an organism (see SS spreadsheet):

i. one stand-alone or in a UroS/UroM fusion - always co-localizes with genes involved in Uro-III production (universal pathway
ii. another - within a multi-domain siroheme synthase. These are often located in close vicinity of nitrite reductase (as in *E. coli*) or sulfite (as in *Bacillus*) reductase operons (siroheme is a cofactor in the corresponding processes)
iii. yet another copy of UroM or UroS/UroM fusion can often be found within a cluster of B_{12} biosynthetic genes (in Vit B_{12} producing organisms) - e.g. in *Listeria*.

We believe, the presence of a "dedicated copy" of UroM (often as UroS/UroM fusion) for nearly every branch of porphinoid biosynthesis is a way to regulate partitioning of Uro-III (the branching point intermediate) between siroheme, Vit B12 (corrinoids), F430 versus the porphyrins branch (leading to components with completely saturated ring - hemes and chlorophylls) depending on specific growth conditions. Might be interesting to measure experimentally the differentiation expression of UroM paralogs. Note, that unlike UroS, UroD catalyzing the first step of the porphyrins branch is never involved in fusions with UroS (other ways of regulation?).

On the other hand, the second type of UroS fusion - that with the downstream enzymes of siroheme pathway: precorrin-2 oxidase and sirohydrochlorin ferrochelatase is most like due to (i) instability and/or (ii) cellular toxicity of precorrin-2 and other intermediates of the porphinoid branch.



Two types of UroM fusions: (I) with UroS, last enzyme of Uro-III biosynthesis (regulatory function?); (ii) with PR2O and SR-FC, catalyzing downstream steps of Siroheme biosynthesis (to prevent release of unstable and/or toxic intermediates?)

Open problems, comments continued

D. Terminal steps of heme biosynthesis: Uro-III to Heme

Missing genes (see SS Spreadsheet):

(i) homologs of both known forms of CPO are missing in genomes of several intracellular parasites

(ii) homologs of both known forms of PPO are missing in roughly half of heme-synthesizing microorganisms. Several hypothetical protein families cluster with known genes of this pathway and can be considered gene candidates for the missing functional role (e.g. *hemY* homologs tentatively included in this SS). None of them, however, has a perfect occurrence profile to fill in the PPO gap in all the genomes. More then one (yet unknown) non-orthologous PPO forms exist?

"Out-of-context", superfluous genes:

One or more clear homologs of CPOan (coproporphyrinogen III oxidase, oxygen-independent (EC 1.3.99.22)) seem to be present in several genomes where all other genes of the **Uro-III to Heme** branch are absent (e.g. *Clostridia, Buchnera* - encircled in SS Spreadsheet above). Their function is unclear. Strong functional coupling of some of these homologs with stress-related proteins may indicate their involvement in (oxidative??) stress management or in detoxication/degradation of intermediates of tetrapyrrol biosynthesis

Functional Coupling

Score	Peg	Function
49	fig|1491.1.peg.987	Heat-inducible transcription repressor hrcA
32	fig|1491.1.peg.986	GrpE protein
22	fig|1491.1.peg.985	Chaperone protein dnaK
21	fig|1491.1.peg.989	GTP-binding protein lepA

References

1. N. Frankenberg, J. Moser, and D. Jahn. 2003. Bacterial heme biosynthesis and its biotechnological application. Appl Microbiol Biotechnol 63: 115–127.
2. H. A. Dailey. 2002. Terminal steps of haem biosynthesis. Biochemical Society Transactions 30: 590-595.
3. D.V. Vavilin and W.F.J. Vermaas. 2002. Regulation of the tetrapyrrole biosynthetic pathway leading to heme and chlorophyll in plants and cyanobacteria. Physiologia Plantarum 115: 9–24.
4. M.M. Kolko, L.A. Kapetanovich, and J.G. Lawrence. 2001. Alternative Pathways for Siroheme Synthesis in Klebsiella aerogenes. J. Bact 183: 328–335
5. T. Ishida, L. Yu, H. Akutsu, K. Ozawa, S. Kawanishi, A.Seto, T. Inubushi, and S. Sano. 1998. A primitive pathway of porphyrin biosynthesis and enzymology in *Desulfovibrio vulgaris*. Proc. Natl. Acad. Sci. USA Vol. 95, pp. 4853–485