

Subsystem: Biotin biosynthesis

Dmitry Rodionov

Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

Biotin (vitamin H) is an essential cofactor for a class of important metabolic enzymes, biotin carboxylases and decarboxylases (Perkins and Pero 2001). Biotin biosynthetic pathway is widespread among microorganisms. The well-studied systems of biotin biosynthesis from *Escherichia coli*, *Bacillus subtilis*, and *Bacillus sphaericus* differ in the first step. *B. subtilis* and *B. sphaericus* use pimeloyl-CoA synthase encoded by the *bioW* gene to synthesize pimeloyl-CoA from pimelic acid. In *E. coli*, pimeloyl-CoA is synthesized from L-alanine and/or acetate via acetyl-CoA, instead of pimelic acid (Ifuku et al. 1994), and products of the *bioC* and *bioH* genes are required for the pimeloyl-CoA synthesis in *E. coli* (EC numbers have not been assigned). The pathway from pimeloyl-CoA to biotin is similar in *E. coli* and bacilli and uses products of the *bioF*, *bioD*, *bioA*, and *bioB* genes. Based on analysis of co-occurrence of the biotin biosynthetic genes and *bioY* in complete genomes, as well as on the distribution of the BirA binding sites, it has been predicted that the transmembrane protein BioY is involved in biotin transport (Rodionov et al., 2002).

Based on positional analysis, multiple functional variants of the pimeloyl-CoA precursor synthesis were predicted in different species (Rodionov et al., 2002):

1. as in *Escherichia coli*. BioC and BioH
2. as in *Haemophilus influenzae*. BioC and BioG
3. as in *Neisseria spp.*. Two copies of BioC, one is coupled to BioH and another one is linked to BioG.
4. as in *Bacillus subtilis*. The pimeloyl-CoA synthase BioW.
5. as in some cyanobacteria. BioC and BioK.
6. as in some rhizobia. BioC and BioZ.

bioG: The *bioG* gene always forms an operon with *bioC* and other biotin synthesis genes in these genomes; furthermore, in *Bacteroides fragilis* there is a single gene encoding a fused protein BioC-BioG. Most gamma-proteobacteria except *Pasteurellaceae* possess the *bioC-bioH* gene pair, whereas all *Pasteurellaceae* have *bioC-bioG*. Notably, *Neisseria meningitidis* has both *bioC-bioH* and *bioC-bioG* gene pairs, and the latter likely has been acquired from *Haemophilus influenzae* or a closely related bacterium, as the respective genes are highly similar. The phylogenetic tree of the BioC family has a separate branch for the proteins associated with BioG.

bioK: Another *bioC*-linked gene found in some cyanobacteria within the *bioFKCDA* operons.

bioZ: Another *bioC*-linked gene found in rhizobia, where it is the last gene in the biotin biosynthesis operon. It was recently shown that the *bioZ* gene from the *bioABFDZ* operon of *Mesorhizobium loti* can complement *bioH* of *E. coli* (Sullivan et al. 2001).

The observed diversity of enzymes for the first step of the biotin biosynthesis (pimeloyl-CoA synthesis) can reflect either frequent non-orthologous gene displacements, or possible use of different substrates for the biotin biosynthesis.

Fig. 1. Biotin biosynthesis and uptake. Subsystem diagram.

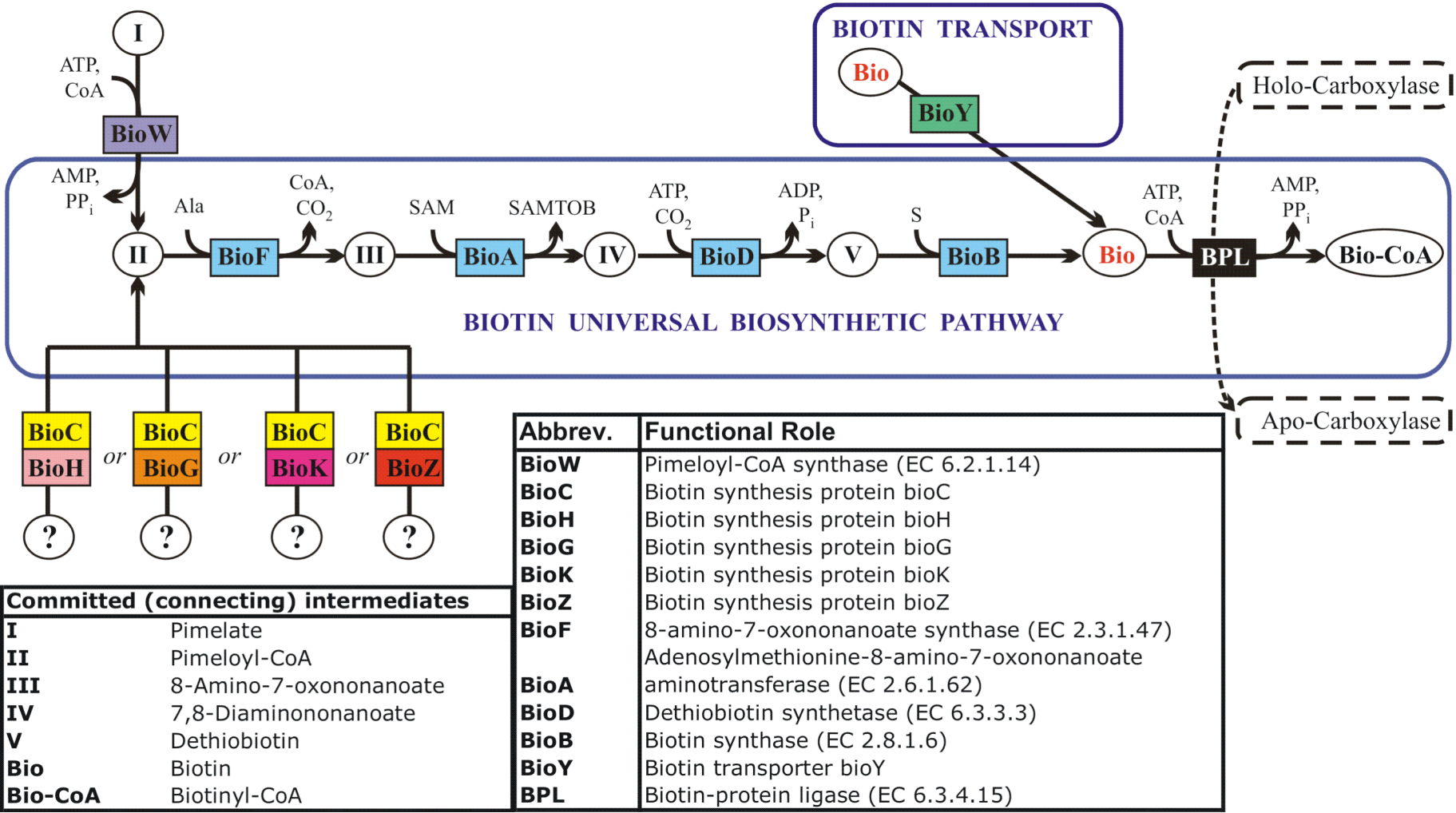


Fig. 2. Biotin biosynthesis and uptake. Subsystem spreadsheet.

Organism	Variant Code	Pymeloyl-CoA synthesis						Biotin synthesis				Biotin uptake
		BioC	BioH	BioG	BioK	BioZ	BioW	BioF	BioA	BioD	BioB	BioY
<i>Escherichia coli</i> K12	1	763	3347					762	760	764	761	
<i>Pseudomonas aeruginosa</i> PAO1	1	504	503					502	421	505	501	
<i>Bacillus cereus</i> G9241	11	3613	3612					3611	3609	3610	3614	2412, 3393
<i>Bacillus halodurans</i> C-125	11	3909	3908					3907	782	783	1748	3358
<i>Haemophilus influenzae</i> Rd KW20	2	1478		1479				1480	1481	1379	984	
<i>Bacteroides fragilis</i> NCTC9343	2	2669		2669				2670	2671	2668	2671	
<i>Fusobacterium nucleatum</i> subsp. <i>Vincentii</i>	22	374		373				372	1657	1656	1655	607
<i>Neisseria meningitidis</i> ser. C (str. FAM18)	3	153, 2143	2142	154				155	971	970	2211	
<i>Methanocaldococcus jannaschii</i> DSM 2661	4						1334	1335	1337	1336	1333	
<i>Bacillus subtilis</i> subsp. <i>subtilis</i> str. 168	44						3027	3025	3026	3024	3023	1037, 3209
<i>Prochlorococcus marinus</i> MED4	55	702			703			704	700	701	1161	184
<i>Synechococcus</i> sp. WH 8102	55	627			628			629	624	626	928	1075
<i>Brucella melitensis</i> 16M	66	182				2839		2836	2838	2837	2835	1430, 318
<i>Mesorhizobium loti</i> MAFF303099	66	2657				4525		4528	4526	4527	4529	5848, 5940
<i>Deinococcus radiodurans</i> R1	7											2649
<i>Thermotoga maritima</i> MSB8	7											792
<i>Streptococcus pneumoniae</i> TIGR4	7											727
<i>Chlamydia trachomatis</i> D/UW-3/CX	7											366

1. as in *Escherichia coli*. BioC and BioH
2. as in *Haemophilus influenzae*. BioC and BioG
3. as in *Neisseria* spp.. Two copies of BioC, one is coupled to BioH and another one is linked to BioG.
4. as in *Bacillus subtilis*. The pimeloyl-CoA synthase BioW.
5. as in some *cyznobia*. BioC and BioK.
6. as in some *rhizobia*. BioC and BioZ.

Functional variant #7: Absence of the *de novo* biotin biosynthetic pathway and the presence of the predicted biotin uptake gene *bioY*.

Functional variants #11, 22, 44, 55, 66 are the same as #1, 2, 4, 5, 6, respectively, but also include additional genes for biotin uptake (*bioY*)