Subsystem: Folate_Biosynthesis
(Synthesis and Recycling of Tetrahydrofolate)

Valérie de Crécy-Lagard,[1] and Andrew D. Hanson [2]
[1]Department of Microbiology and Department of Microbiology and Cell Science,
[2]Department of Horticultural Sciences
University of Florida, Gainesville, FL 32611

- **I. Introduction**

- In the majority of living cells folates occur as one-carbon substituted tetrahydro-pteroyl-polyglutamate derivatives. These folates donate one-carbon groups during the synthesis of purines, formylmethionyl-tRNA, thymidylate, pantothenate, glycine, serine, and methionine. Folates have importance in human nutrition, health, and disease, and antifolate drugs are commonly used in cancer chemotherapy and as antimicrobials.
- **Plants, fungi, and bacteria** synthesize the precursor DHF *de novo* from GTP and *p*-aminobenzoic acid (pABA). The NADPH-dependent reduction of DHF catalyzed by DHFR is essential both in the *de novo* and in the recycling pathways. DHFR recycles the DHF produced by all organisms that use thymidylate synthase (TS) encoded by *thyA* to convert dUMP to dTMP. A new FAD-dependent TS enzyme (*thyX*) has been recently discovered (Science 2002 297:5-7). In the organisms with a *thyX* gene, DHFR is needed only for the *de novo* pathway and the corresponding gene is going to be essential in rich medium only if these organisms need formyl-methionyl-tRNA.
- **Animal cells** lack key enzymes of the folate biosynthetic pathway and a dietary source of folate is required for normal growth and development.
- The situation in **Archaea** is quite complicated. THF has been shown to exist in halophilic Archaea but most methanogenic Archaea do not use THF but another cofactor tetrahydro-methanopterin as a pterin C1 carrier (J. Bacteriol. 1988 17:4608-12). However, recent studies have shown that some Archaea such as *Methanosarcina barkeri* have THF-dependent enzymes and that their growth is dependent on folic acid or pABA (Arch. Microbiol. 2004 182:313-25).

Subsystem analysis

This subsystem was quite a challenge to encode. Many genes are fused encoding multi-functional proteins, and several enzymes belong to superfamilies of paralogs, making annotations tricky and laborious. Many issues remain to be resolved. The pABA pathway is associated with aromatic amino acid metabolism and has not been expanded in the context of this subsystem.

A subsystem diagram including the list and abbreviations of functional roles and pathway intermediates is provided in Figure 1. A representative section of the subsystem spreadsheet is shown in Figure 2 (modified from the full display available in SEED). Brief notes and comments on some of the revealed problems and conjectures are provided in Section II "Subsystem Notes".

II. Subsystem Notes

1. De novo pathway

Fungi and plants, as well as most bacteria, have the full THF biosynthesis pathway. In *E. coli* and *S. cerevisiae,* genes have been identified for all but one step of the pathway, the conversion of 7,8-dihydroneopterin triphosphate to the corresponding monophosphate (*folQ*); removal of the last phosphate is believed to be mediated by a non-specific phosphatase (J. Biol. Chem. 1974 249:2405-10).

Missing and alternative genes

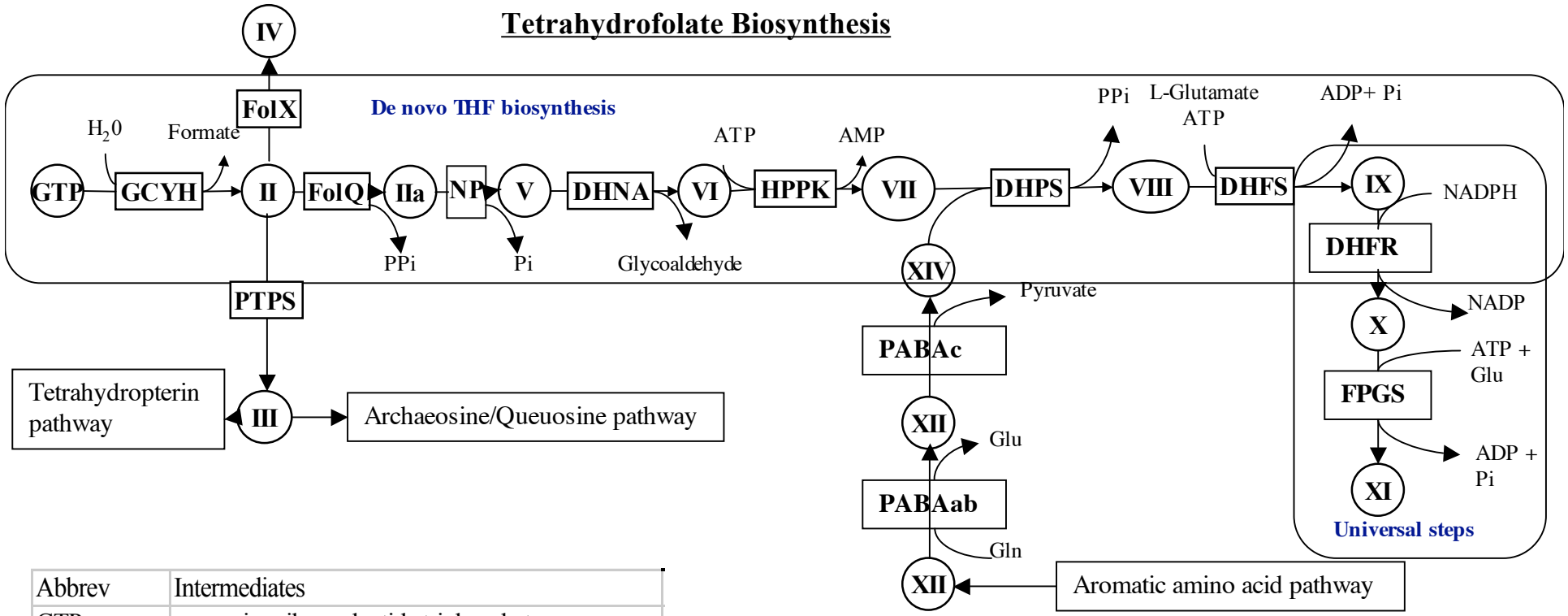
A *folQ* gene was recently identified in *L. lactis* (J. Biol. Chem. 2005 280:5274-80) as part of the *folKEPQC* gene cluster. This gene belongs to a large and functionally heterogeneous Nudix hydrolase superfamily hardly amenable to projection of annotations just by homology. Several Nudix hydrolase gene candidates can be identified in folate gene clusters, and these should be tested experimentally. Other putative phosphohydrolases unrelated to FolQ are found in some of the folate-related gene clusters, e.g., peg1083 in *Clostridium perfringens*.

DHFR

DHFRs are known to be encoded by several gene families. In addition to the most common *folA* family, a pteridine reductase-like *folM* ("DHFR1" in our subsystem; J. Bacteriol. 2003 185:7015-8), and a flavin-dependent reductase ("altDHFR2"; Mol. Microbiol. 2004 54:1307-18) have been identified. DHFR1 enzymes belong to the large superfamily of short-chain dehydrogenases-reductases (SDR) involved in a variety of reactions. A FolM-specific sequence motif TGXXXRXG was used to discriminate DHFR1 subfamily from other SDRs (featuring a TGXXXGXG motif). Even after propagation of all known DHFR families, respective genes are still missing in several organisms **(see variant 6)**. One gene candidate (altDHFR3) is embedded in the *FolEKPB* gene cluster in *Streptomyces coelicolor.* This prediction was made by A. Hanson and it is currently being tested in his laboratory.

Figure 1. Subsystem diagram

Subsystem: Folate_Biosynthesis



Abbrev	Intermediates
GTP	guanosine ribonucleotide triphosphate
II	7,8-dihydroneopterin triphosphate
IIa	7,8-dihydroneopterin monophosphate
III	6-pyruvoyltetrahydropterin
IV	7,8-dihydromonapterin
V	7,8-dihydroneopterin
VI	7,8-dihydrohydroxymethylpterin pyrophospate
VII	7,8-dihydrohydroxymethylpterin
VIII	dihydropteroate
IX	dihydrofolate
X	tetrahydrofolate
XI	tertrahydrofolate polyglutamate
XII	chorismate

Abbrev	Functional Role
GCYHI1	GTP cyclohydrolase I (EC 3.5.4.16) type 1
GCYHI2	GTP cyclohydrolase I (EC 3.5.4.16) type 2
PTPS	6-pyruvoyl tetrahydrobiopterin synthase (EC 4.2.3.12)
FOLQ	Dihydroneopterin triphosphate pyrophosphohydolase
NP	Probably mediated by a non-specific phosphatase
DHNA	Dihydroneopterin aldolase (EC 4.1.2.25)
FolX	Dihydroneopterin triphosphate epimerase
HPPK	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3)
DHPS	Dihydropteroate synthase (EC 2.5.1.15)
DHFS	Dihydrofolate synthase (EC 6.3.2.12)
DHFR	Dihydrofolate reductase (EC 1.5.1.3)
DHFR1	FolM Alternative dihydrofolate reductase 1
DHFR2	Alternative dihydrofolate reductase 2
DHFR3	Alternative dihydrofolate reductase 3
FPGS	Folylpolyglutamate synthase (EC 6.3.2.17)
PABA	Aminodeoxychorismate synthase glutamine amidotransferase component II (EC 6.3.5.8)
PABB	Aminodeoxychorismate synthase component I (EC 6.3.5.8)
PABC	Aminodeoxychorismate lyase (EC 4.1.3.38)

Figure 2. Subsystem sprteadsheet (fragment)

Subsystem: Folate_Biosynthesis

		DHF de novo synthesis							THF-PolyGlu synthesis	
Organism	Variant Code	*folE	folQ	*DHNA	folX	HPPK	DHPS	DHFS	*DHFR	FPGS
Borrelia burgdorferi B31 [B]	-1									
Homo sapiens [E]	1	398-1							18425-10, 977-10	4858
Mycoplasma genitalium G-37 [B]	10								234-10	?
Methanosarcina barkeri [A]	12	3612-2				95			2414-12	
Thermoplasma volcanium GSS1 [A]	13	1214-2		666-5			153	757	?	757
Enterococcus faecium [B]	2							1776	1663-10	1776
Rickettsia conorii str. Malish 7 [B]	3	527-1					37	779	34-10	779
Plasmodium falciparum 3D7 [E]	4	1580-1		?		4134	4134	4846	2808-10	4846
Escherichia coli K12 [B]	5	2128-1	?	3006-4	2277	142	3120	2289	49-10, 1592-11	2289
Geobacter metallireducens [B]	5	1502-2, 3335-2	?	1275-5		1273	3050	474	1416-10	474
Lactococcus lactis subsp. lactis Il1403 [B]	5	1188-1	1190	1187-4		1188	1189	1191	1183-10, 2013-10	1191
Prochlorococcus marinus MED4 [B]	6	1823-1	?	1768-4		158	1481	924	?	924
Chlamydophila pneumoniae J138 [B]	7	?	?	755-4		756	756	?	757-10	?
Mycobacterium microti OV254 [B]	8	?	?	2586-4		2585	2587, 713	744	3027-10	744

GCYHI

GTP cyclohydrolase I is the first step of the folate pathway. It is found in mammals that have a biopterin (BH4) pathway, but it is still a missing gene in many species containing an otherwise complete set of folate biosynthetic genes. A candidate gene family (GCYHI2) was identified by genome context analysis techniques in conjunction with a tRNA modification pathway (see comments to Queuosine-Archaeosine Biosynthesis subsystem). This prediction was made by V. de Crécy, and it is currently being tested in her laboratory. Several organisms are still lacking either form of GCYHI **(see variant 8).**

DHNA

Homologs of *folB* gene (encoding DHNA) appear to be missing in many organisms. Genome and functional context analysis allowed A. Hanson to infer that this role may be played by some homologs of *E. coli* fructose-6-phosphate aldolase ("FSA" in subsystem). This prediction is being tested by the Hanson laboratory. Some organisms still lack any DHNA gene candidates **(variant 4).**

Multiple missing genes:

- A few organisms lack both DHNA and HPPK genes **(variants 3)**

-Chlamydiae lack GCYHI and the DHFS/FPGS **(variants 7)**

- **2. Salvage pathways**
- The most studied salvage pathway is found in mammals and many other eukaryotes have just the DHFR and FPGS enzymes **(see variant 1).** Many relatively poorly studied bacteria also seem to rely on a salvage pathway. Many pathogenic and related bacteria have only the DHFS, FPGS and DHFR genes **(variant 2)**, suggesting a salvage of 7,8-dihydropteroate. However, this compound is not expected to occur in their natural environment leaving us with an open problem for further studies.
- In Borrelia **(nonfunctional variant code "– 1")** no genes of the folate pathway can be detected, in agreement with the presence of the folate-independent thymidylate synthase (TS) of the *thyX* family. In most Mycoplasmas, DHFR and a folate-dependent TS gene are found, but FPGS appears to be a missing gene **(variant 10).** These species may salvage mono- or polyglutamylated folates from the host. A default **variant code "0" is retained** for those genomes where we were unable to rationalize gene patterns. Many of them are due to incomplete sequencing or sequencing mistakes, or even reflect the existence of pseudogenes

- **3. Folate in Archaea.**
- The situation in Archaea is even more complicated. Although many methanogens (e.g., *M. jannaschii*) produce tetrahydromethanopterin instead of folate (hence **variant code "-1"**), they contain homologs of some folate-related genes, likely reflecting a resemblance between some steps in the biosynthesis of these two cofactors. For example, a FolP-like gene family (FolP2, MJ0107) found mostly in Archaea, may be involved in methanopterin biosynthesis (M. Rasche, personal communication). On the other hand, the ThyA-like TS enzyme of *M. jannaschii* was shown to utilize folate rather than methanopterin derivatives as a cofactor (Nature Struct. Biol. 1999 6:750-4. This observation may indicate an existence of: (i) an alternative folate biosynthetic pathway, or (ii) an unidentified methanopterin-related methyl donor. In both cases one should anticipate an existence of an alternative reductase in those genomes that lack DHFR homologs. Several archaeal genomes (*Thermoplasma* and *Ferroplasma,* **variant 13)** containing almost all of the *de novo* pathway genes, lack recognizable homologs of DHFR and HPPK. Finally, one of the several archaeal species containing both ThyA-like TS and DHFR **(variant 12),** was shown to be dependent on folate or pABA for growth. More experimental studies are needed to clarify the biogenesis of methanopterin and folate-related cofactors in Archaea.