Evolution of the Purine Biosynthetic Enzyme, Formylglycinamide Ribonucleotide Amidotransferase (FGARAT).

Allison Batties*, Jessica Bennett, and Jeffrey D. Newman, Biology Department, Lycoming College, Williamsport, PA

Abstract

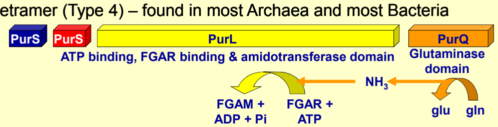
Formylglycinamide ribonucleotide amidotransferase (FGARAT) catalyzes the fourth step in the purine biosynthetic pathway. Two forms of this enzyme have been described in the literature, a monomeric form present in eukaryotes, beta and gamma proteobacteria, and a heterotetrameric form found in most other Bacteria and Archaea. The heterotetrameric enzyme is composed of PurS, PurL and PurQ subunits in a ratio of 2:1:1, while the monomeric enzyme has a domain organization corresponding to a gene fusion of *purS-purS-purL-purQ*. To investigate the evolution of this enzyme, FGARAT sequences were retrieved from the completed genomes of different families of organisms and compared.

Multiple sequence alignments have revealed a heterodimeric form of the enzyme as well as a second monomeric form. The heterodimer consists of a PurQ subunit together with a PurL of intermediate length that corresponded to a *purS-purS-purL* gene fusion. The phylogenetic distribution of the heterodimer suggests that it arose in an ancestor of the delta-proteobacteria and has spread to several other organisms via horizontal gene transfer. The second monomeric FGARAT has a domain organization similar to the previously described form, but also has sequence features that suggest an independent origin. This enzyme form is present in several phylogenetically diverse organisms, thus obscuring its origin. Further insights into the evolution of FGARAT from a tetramer to a dimer and finally to a monomer have been derived from mapping of residues conserved between the different forms onto the recently solved structure of the monomeric FGARAT (Anand et al., 2004).

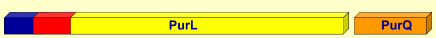
Background

Three forms of FGARAT

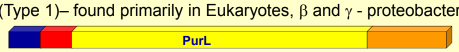
- Heterotetramer (Type 4) – found in most Archaea and most Bacteria



- Heterodimer (Type 2)– found primarily in Planctomycetes and δ - proteobacteria



- Monomer (Type 1)– found primarily in Eukaryotes, β and γ - proteobacteria



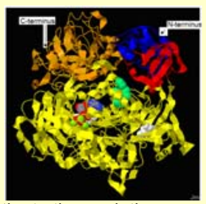


Figure 1. Domain/ Subunit organization of FGAR Amidotransferase Enzymes.
Color coding of subunits and domains correspond to domain colors shown in 3-D structure view of monomeric enzyme found in *Salmonella typhimurium (Anand et al., 2004).*

Goals: 1. Compare FGARAT sequences to investigate the evolution of this enzyme.
2. Identify the location of residues conserved between different enzyme forms.

Hypothesis: Residues conserved between different enzyme forms will be located in regions where the domain organization is shared by those forms.

Methods

All completed genomes (as of 3/1/07) in the NCBI Microbial Genomes Database were searched for FGARAT sequences, either using existing annotation or by performing BLAST searches. After initial compilation of predicted ORF sizes, the FGARAT sequences from at least one representative of each taxonomic family were retrieved from the NCBI Database, and the sequences from multiple subunit FGARATs were assembled to mimic the domain organization of the monomer. Multiple sequence alignments (Clustal W) and phylogenetic trees were generated using the Megalign component of the Lasergene sequence analysis suite (DNASTAR). The atomic coordinates for FGARAT were retrieved from the RCSB Protein Data Bank (PDB ID 1T3T - Anand et al., 2004) and analyzed using Protein Explorer (Martz, 2002)

Sequences Analyzed

Table 1. 100 Representative Sequences used for Phylogenetic Analysis

Organism	Taxonomy	Sub-units	PurL	PurQ	PurS
Gramella forsetii KT0803	Bacteroidetes, Flavobacteriales	1	1225	-	-
Cytophaga hutchinsonii	Bacteroidetes, Sphingobacteriales	1	1231	-	-
Saccharomyces cerevisiae	Eukaryota, Fungi	1	1358	-	-
Homo sapiens	Eukaryota, Metazoa	1	1338	-	-
Dictyostelium discoideum	Eukaryota, Mycetozoa	1	1355	-	-
Arabidopsis thaliana	Eukaryota, Viridiplantae	1	1387	-	-
Rhodoferax ferrireducens	Proteobacteria, Beta, Burkholderiales	1	1409	-	-
Polaromonas naphthalenivorans	Proteobacteria, Beta, Burkholderiales	1	1340	-	-
Burkholderia pseudomallei	Proteobacteria, Beta, Burkholderiales	1	1356	-	-
Thiobacillus denitrificans	Proteobacteria, Beta, Hydrogenophilales	1	1291	-	-
Methylobacillus flagellatus	Proteobacteria, Beta, Methylophilales	1	1297	-	-
Neisseria meningitidis	Proteobacteria, Beta, Neisseriales	1	1320	-	-
Nitrosomonas europaea	Proteobacteria, Beta, Nitrosomonadales	1	1304	-	-
Azoarcus sp. EbN1	Proteobacteria, Beta, Rhodocyclales	1	1310	-	-
Desulfotalea psychrophila	Proteobacteria, Delta, Desulfobacterales	1	1267	-	-
Myxococcus xanthus	Proteobacteria, Delta, Myxococcales	1	1302	-	-
Aeromonas hydrophila	Proteobacteria, Gamma, Aeromonadales	1	1357	-	-
Colwellia psychrerythraea	Proteobacteria, Gamma, Alteromonadales	1	1323	-	-
Baumannia cicadellinicola	Proteobacteria, Gamma, Candidatus Baumannia	1	1297	-	-
Nitrosococcus oceani	Proteobacteria, Gamma, Chromatiales	1	1300	-	-
Salmonella typhimurium	Proteobacteria, Gamma, Enterobacteriales	1	1295	-	-
Escherichia coli	Proteobacteria, Gamma, Enterobacteriales	1	1295	-	-
Coxiella burnetii	Proteobacteria, Gamma, Legionellales	1	1296	-	-
Methylococcus capsulatus	Proteobacteria, Gamma, Methylococcales	1	1288	-	-
Hahella chejuensis	Proteobacteria, Gamma, Oceanospirillales	1	1298	-	-
Haemophilus influenzae	Proteobacteria, Gamma, Pasteurellales	1	1320	-	-
Pseudomonas syringae	Proteobacteria, Gamma, Pseudomonadales	1	1313	-	-
Francisella tularensis	Proteobacteria, Gamma, Thiotrichales	1	1290	-	-
Vibrio fischeri	Proteobacteria, Gamma, Vibrionales	1	1303	-	-
Xanthomonas campestris	Proteobacteria, Gamma, Xanthomonadales	1	1348	-	-
Magnetococcus sp. MC-1	Proteobacteria, Magnetococcus	1	1295	-	-
Clostridium acetobutylicum	Firmicutes, Clostridia	1	1255	-	-
Clostridium perfringens	Firmicutes, Clostridia	1	1266	-	-
Clostridium tetani	Firmicutes, Clostridia	1	1258	-	-
Streptococcus pyogenes	Firmicutes, Lactobacillales	1	1257	-	-
Streptococcus agalactiae	Firmicutes, Lactobacillales	1	1203	-	-
Fusobacterium nucleatum	Fusobacteria, Fusobacterales	1	1249	-	-
Corynebacterium diphtheriae	Actinobacteria, Actinomycetales	1	1238	-	-
Bifidobacterium longum	Actinobacteria, Bifidobacteriales	1	1244	-	-
Methanocorpusculum labreanum	Archaea, Euryarchaeota, Methanomicrobiales	1	1231	-	-
Methanococcus maripaludis	Archaea, Euryarchaeota, Methanococci	2	989	272	-
Methanospirillum hungatei	Archaea, Euryarchaeota, Methanomicrobiales	2	979	282	-
Dehalococcoides ethenogenes	Chloroflexi, Dehalococcoidetes	2	953	255	-
Rhodopirellula baltica	Planctomycetes, Planctomycetales	2	1009	292	-
Anaplasma marginale	Proteobacteria, Alpha, Rickettsiales	2	1016	260	-
Bdellovibrio bacteriovorus	Proteobacteria, Delta, Bdellovibrionales	2	1009	239	-
Desulfovibrio vulgaris	Proteobacteria, Delta, Desulfovibrionales	2	1009	269	-
Geobacter sulfurreducens	Proteobacteria, Delta, Desulfuromonadales	2	996	275	-
Syntrophobacter fumaroxidans	Proteobacteria, Delta, Syntrophobacterales	2	1009	269	-
Acidobacteria bacterium Ellin345	Acidobacteria, Acidobacteriales	4	768	231	80
Solibacter usitatus	Acidobacteria, Solibacteres, Solibacterales	4	742	232	81
Corynebacterium glutamicum	Actinobacteria, Actinomycetales	4	762	223	81
Corynebacterium jeikeium	Actinobacteria, Actinomycetales	4	839	223	84
Streptomyces coelicolor	Actinobacteria, Actinomycetales	4	752	226	90
Acidothermus cellulolyticus 11B	Actinobacteria, Actinomycetales	4	754	225	81
Rubrobacter xylanophilus	Actinobacteria, Rubrobacterales	4	727	220	74
Aquifex aeolicus	Aquificae, Aquificales	4	745	227	77
Pyrobaculum aerophilum	Archaea, Crenarchaeota, Thermoprotei	4	697	212	84
Archaeoglobus fulgidus	Archaea, Euryarchaeota, Archaeoglobi	4	765	271	80
Haloarcula marismortui	Archaea, Euryarchaeota, Halobacteria	4	720	228	84
Methanothermobacter thermautotrophicus	Archaea, Euryarchaeota, Methanobacteria	4	714	214	84
Methanocaldococcus jannaschii	Archaea, Euryarchaeota, Methanococci	4	733	230	83
Methanosarcina acetivorans	Archaea, Euryarchaeota, Methanomicrobia	4	715	232	88
Methanopyrus kandleri	Archaea, Euryarchaeota, Methanopyri	4	724	226	84
Methanococcoides burtonii	Archaea, Euryarchaeota, Methanosarcinales	4	715	231	82
Pyrococcus abyssi	Archaea, Euryarchaeota, Thermococci	4	705	223	84
Picrophilus torridus	Archaea, Euryarchaeota, Thermoplasmata	4	741	248	75
Thermoplasma acidophilum	Archaea, Euryarchaeota, Thermoplasmata	4	759	257	79
Haloquadratum walsbyi	Archaea, Halobacteria, Halobacteriales	4	725	228	91
Salinibacter ruber	Bacteroidetes, Sphingobacteriales	4	754	235	92
Chlorobium tepidum	Chlorobi, Chlorobiales	4	759	234	84
Synechococcus elongatus	Cyanobacteria	4	777	221	74
Gloeobacter violaceus	Cyanobacteria, Gloeobacteria	4	774	232	88
Nostoc sp. PCC 7120	Cyanobacteria, Nostocales	4	782	224	92
Trichodesmium erythraeum	Cyanobacteria, Oscillatoriales	4	775	231	87
Prochlorococcus marinus	Cyanobacteria, Prochlorales	4	803	217	90
Deinococcus radiodurans	Deinococcus-Thermus, Deinococcales	4	747	266	84
Thermus thermophilus	Deinococcus-Thermus, Thermales	4	725	227	84
Bacillus subtilis	Firmicutes, Bacillales	4	742	227	84
Carboxydothermus hydrogenoformans	Firmicutes, Clostridia	4	728	234	81
Thermoanaerobacter tengcongensis	Firmicutes, Clostridia	4	733	224	82
Moorella thermoacetica	Firmicutes, Clostridia, Thermoanaerobacteriales	4	733	236	84
Enterococcus faecalis	Firmicutes, Lactobacillales	4	739	224	83
Caulobacter crescentus	Proteobacteria, Alpha, Caulobacterales	4	739	220	79
Sinorhizobium meliloti	Proteobacteria, Alpha, Rhizobiales	4	743	223	80
Silicibacter pomeroyi	Proteobacteria, Alpha, Rhodobacterales	4	719	222	76
Gluconobacter oxydans	Proteobacteria, Alpha, Rhodospirillales	4	734	233	80
Candidatus Pelagibacter ubique	Proteobacteria, Alpha, Rickettsiales	4	730	227	80
Zymomonas mobilis	Proteobacteria, Alpha, Sphingomonadales	4	734	221	77
Rhizobium etli	Proteobacteria, Alpha, Rhizobiales	4	743	223	80
Anaeromyxobacter dehalogenans	Proteobacteria, Delta, Myxococcales	4	759	218	81
Campylobacter jejuni	Proteobacteria, Epsilon, Campylobacterales	4	728	215	81
Leptospira interrogans	Spirochaetes, Spirochaetales	4	745	219	82
Thermotoga maritima	Thermotogae, Thermotogales	4	603	213	82

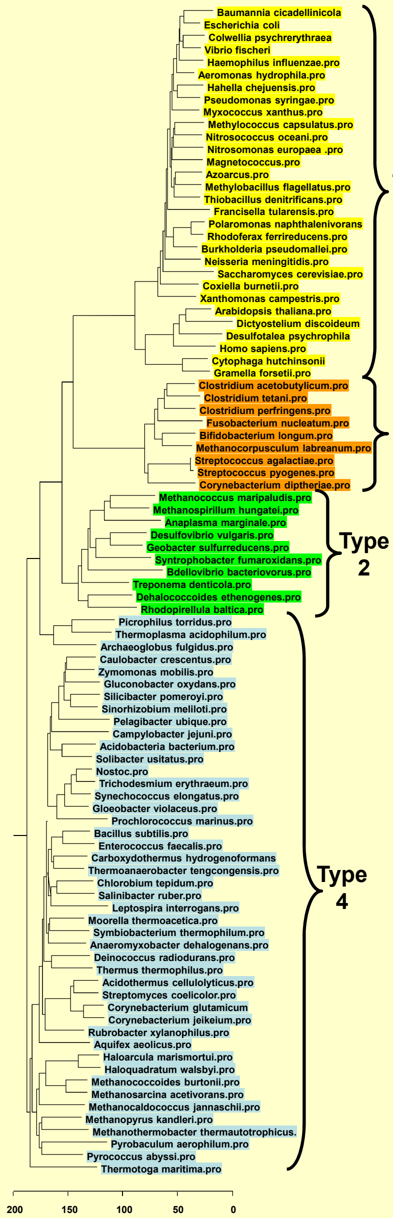


Figure 2. Phylogenetic Tree Based on Clustal W alignment of combined PurS-PurS-PurL-PurQ sequences.

Results & Conclusions

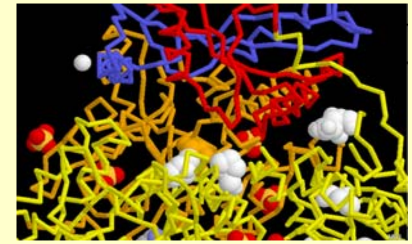


Figure 3. Residues conserved between Type 2 and Type 1b are located at the interface between PurS and PurL domains

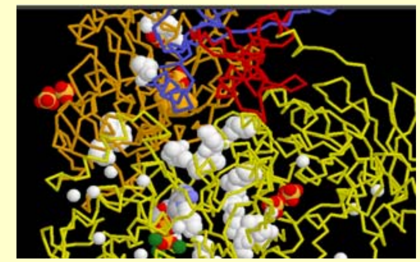


Figure 4. Residues conserved between Type 2 and Type 4 are located at the interface between PurQ and PurL domains

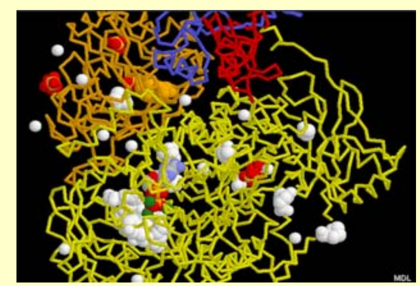


Figure 5. Residues conserved between Type 1a, 2 and Type 4 are located near the ADP binding site of the PurL domain.

Implications

The Salmonella PurL domain contains two active sites, one is functional, and one is vestigial but still binds ADP (Anand et al, 2004). The conservation of residues near the vestigial site in the type 1a, 2 and 4 enzymes but not in the type 1b enzyme, suggests that the type 1b enzyme may no longer bind ADP.

References

- Anand, R., Hoskins, A., Stubbe, J., Ealick, SA. (2004). Domain Organization of *Salmonella typhimurium* Formylglycinamide Ribonucleotide Amidotransferase Revealed by X-ray Crystallography. *Biochemistry* 43:10328-10342.
- Martz, Eric. 2002. Protein Explorer: Easy Yet Powerful Macromolecular Visualization, *Trends in Biochemical Sciences*, 27:107-109.