THE ROLE OF GENE DUPLICATION IN THE EVOLUTION OF PURINE NUCLEOTIDE SALVAGE PATHWAYS

ARTURO BECERRA and ANTONIO LAZCANO*
Facultad de Ciencias, UNAM, Apdo. Postal 70-407, Cd. Universitaria, 04510 México, D.F., México

(Received 5 November, 1996)

Abstract. Purine nucleotides are formed *de novo* by a widespread biochemical route that may be of monophyletic origin, or are synthesized from preformed purine bases and nucleosides through different salvage pathways. Three monophyletic sets of purine salvage enzymes, each of which catalyzes mechanistically similar reactions, can be identified: (a) adenine-, xanthine-, hypoxanthine- and guanine-phosphoribosyltransferases, which are all homologous among themselves, as well as to nucleoside phosphorylases; (b) adenine deaminase, adenosine deaminase, and adenosine monophophate deaminase; and (c) guanine reductase and inosine monophosphate dehydrogenase. These homologies support the idea that substrate specificity is the outcome of gene duplication, and that the purine nucleotide salvage pathways were assembled by a patchwork process that probably took place before the divergence of the three cell domains (Bacteria, Archaea, and Eucarya). Based on the ability of adenine PRTase to catalyze the condensation of PRPP with 4-aminoimidazole-5-carboxamide (AICA), a simpler scheme of purine nucleotide biosynthesis is presented. This hypothetical route requires the prior evolution of PRPP biosynthesis. Since it has been argued that PRPP, nucleosides, and nucleotides are susceptible to hydrolysis, they are very unlikely prebiotic compounds. If this is the case, it implies that many purine salvage pathways appeared only after the evolution of phosphorylated sugar biosynthetic pathways made ribosides available.

Abbreviations: AICA, 4-aminoimidazole-5-carboxamide; AICN, 4-aminoimidazole-5-carbonitrile; AICAR, 5-amino-4-imidazolecarboxamide ribotide; PRPP, 5-phospho-α-D-ribosyl-1-pyrophosphate; PRTase, phosphoribosyltransferase; ADA, adenine deaminase; AMP, adenosine 5'-monophosphate; IMP, inosine 5'-monophosphate; GMP, guanosine 5'-monophosphate; GDP, guanosine 5'-diphosphate; GTP, guanosine 5'-triphosphate; XMP, xanthosine 5'-monophosphate; HMP, hypoxanthosine 5'-monophosphate; IMPDH, inosine monophosphate dehydrogenase; PNPase, purine nucleoside phosphorylase.

1. Introduction

The first detailed attempt to explain the origin of metabolic pathways is due to Horowitz (1945), who suggested that biosynthetic routes are the outcome of the stepwise, sequential acquisition of enzymes in reverse order as found in extant pathways. Twenty years after this so-called retrograde hypothesis was first suggested, the discovery of operons prompted Horowitz (1965) to propose that clusters of genes involved in biosynthetic routes were the result of early tandem duplication events.

An alternative interpretation of the role of gene duplication in the establishment of anabolic routes was developed by Waley (1969), Ycas (1974), and Jensen

* Author for all correspondence.

Origins of Life and Evolution of the Biosphere **28:** 539–553, 1998.

(1976). According to this scheme, biosynthetic routes were assembled by a patchwork mechanism involving primitive enzymes of broad substrate specificity. Gene duplication and subsequent divergence of the new sequences led to the diversification of function and narrowing of specificity.

The *de novo* purine ribonucleotide biosynthesis is a highly conserved anabolic route whose ample phylogenetic distribution (Henderson and Paterson, 1973) suggest that it may have already been present in the last common ancestor of the three cell domains (Bacteria, Archaea, and Eucarya). In contrast, there is a wide variety of salvage pathways (Figure 1), which participate not only in the regulation of the intracellular concentration of nucleotides, but also in the reutilization of free intra- or extracellular bases and purine nucleosides (Berens *et al*., 1995).

Salvage pathways are essential for a group of distantly related organisms in which the *de novo* purine route has apparently been lost. This group, which is strictly dependent on exogenous purines and their corresponding nucleosides and nucleotides, includes eubacteria such as *Haemophilus influenzae* and the mycoplasma, amitochondrial eukaryotes like *Giarda lamblia* and *Trichomonas vaginalis*, protists such as the kinetoplastida, and parasitic helmiths and specialized animal tissue like that of brain (Tham *et al*., 1993; Berens *et al*., 1995; Tatusov *et al*., 1996). It has been suggested that the absence of the *de novo* biosynthesis of purine nucleotides among several species of parasitic protists is an indication that this pathway never evolved among the protozoa (Hitchings, 1982). However, the 18S rRNA phylogenies of the same set of organisms shows that they are a paraphyletic group (Sogin, 1994) that may have undergone multiple independent losses of purine anabolism. This possibility is supported by the observation that one of the most frequent metabolic defficiencies found in parasitic protists is precisely the secondary loss of purine nucleotide biosynthesis (Wang, 1991).

Since it is generally assumed that the first organisms were derived from the preformed organic compounds available in the primitive environment (Oparin, 1938), it is tempting to assume that their growth and reproduction depended on the heterotrophic uptake of nucleotides and other raw material present in the primitive soup (see, for instance, Pennisi, 1996). However, the problems associated with the prebiotic availability of nucleotides argue against this possibility, and suggest that the corresponding salvage pathways evolved after the development of purine-nucleotide biosynthesis. Recent sequence comparisons have shown that several of the salvage enzymes belong to large families involved primarily in nucleotide metabolism (Mushegian and Koonin, 1994; Bork *et al*., 1995; Holm and Sander, 1997). However, no attempt has been made to discuss these evolutionary relationships from the viewpoint of the emergence and development of metabolic pathways, nor in terms of the prebiotic availability of the enzyme substrates.

It has been proposed that the PRTase-mediated attachment of purines to 5-phospho-a-D-ribosyl-1-pyrophosphate (PRPP) to form the corresponding 5'-nucleotides, which today is part of a salvage pathway (Figure 1), may have participated in an ancient form of purine nucleotide biosynthesis (Zubay, 1993). Here

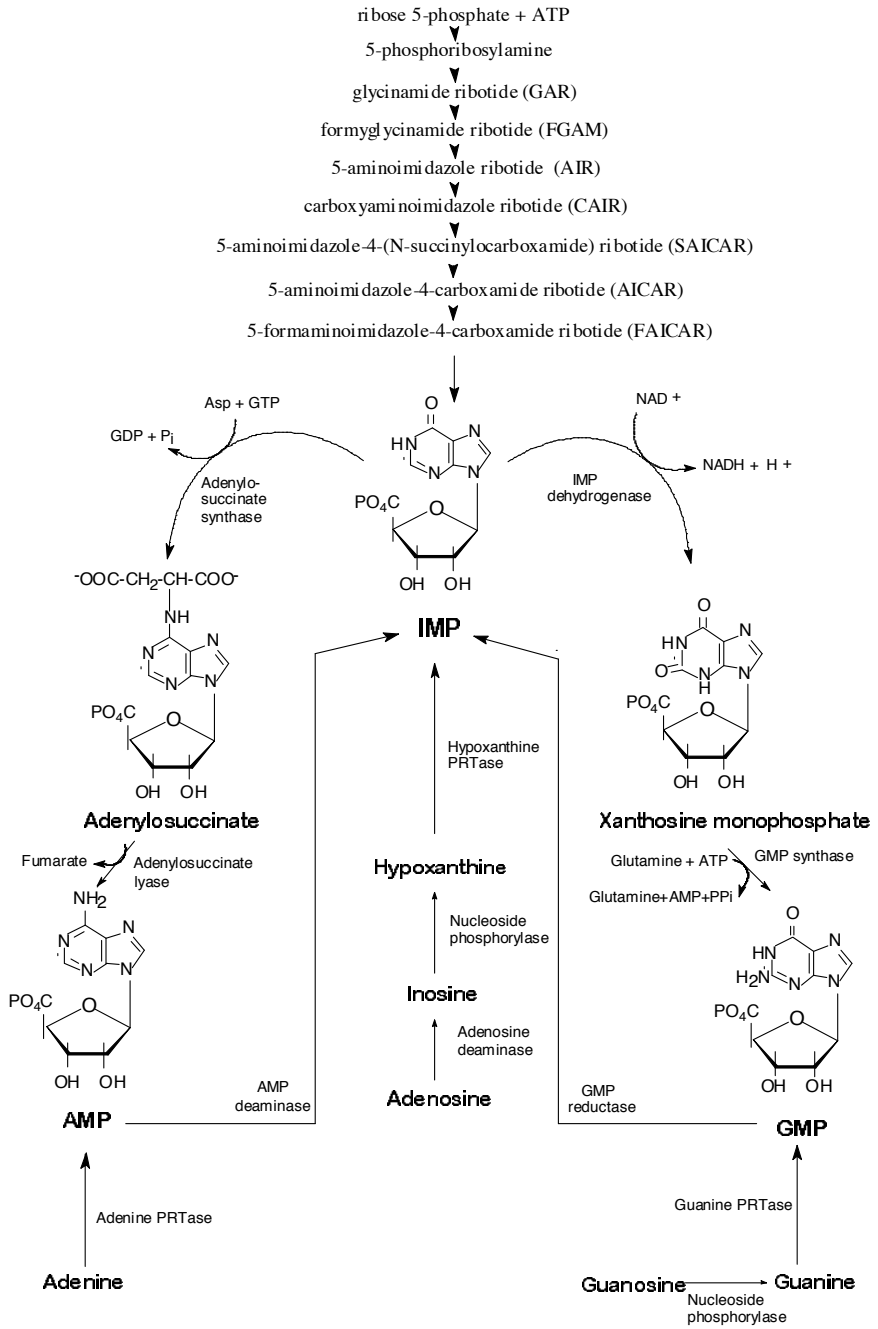


Figure 1. Purine nucleotide metabolism. Both the *de novo* and the major salvage pathways routes are shown.

Table I
Biological distribution of PRTases and nucleoside phosphorylases

Organism	APRTase	XPRTase	HGPRTase	PNPase
Bacteria				
Escherichia coli	*	*	*	*
Haemophilus influenzae	*	*	*	*
Salmonella typhimurum	*	*		?
Lactococcus lactis	*		*	?
Mycoplasma genitalium	*		*	*
Mycoplasma pneumoniae	*		*	*
Synechocystis sp.	*			
Archaea				
Methanococcus voltae	*	*	*	?
Methanococcus jannaschii	*			*
Eucarya				
Giardia lamblia	*			
Tritrichomonas foetus	*			*
Kinetoplastida	*	*	*	*
Leishmania donavani	*		*	?
Plasmodium falciparum	*		*	*
Toxoplasma gondii	*	*	*	*
Eimeria tenella	*			*
Schistosoma mansoni	*		*	*
mammalian cells	*		*	*

APRTase: adenine phosphoribosyltransferase; GPRTase: guanine phosphoribosyltransferase; XPRTase: xanthine phosphoribosyltransferase; HGPRTase: hypoxanthine-guanine phosphoribosyltransferase; HGXPRTase: hypoxanthine-guanine-xanthine phosphoribosyl-transferase; PNPase: nucleoside phosphorylase. GPRTase has only been reported in *G. lamblia*, and HGXPRTase only in *T. foetus* and *E. tenella* (based on Berens *et al*., 1995).
* Present; ? no information available.

we suggest an additional possibility, based on a semi-enzymatic route involving the PRTase-mediated condensation of PRPP with the prebiotic reagent 5-amino-4-imidazolecarboxamide, to produce 5-amino-4-imidazolecarboxamide ribotide (AICAR). This hypothetical route, which would require the prior evolution of PRPP biosynthesis, omits the first eight steps of the extant purine biosynthesis, all of which involve highly unstable intermediates.

Knowledge of the biological distribution and diversity of salvage pathways is still fragmentary (Tables I and II). Nevertheless, as discussed in this paper, analysis

Table II

Purine nucleotide salvage pathways in organisms whose entire genomes have been sequenced (January, 1997)

	PRTase[a]	Nucleoside[b] phosphorylase	Adenine deaminase[c]
Bacteria			
Haemophilus influenzae	*	*	
Escherichia coli	*		*
Mycoplasma genitalium	*	*	
Mycoplasma pneumoniae	*	*	
Synechocystis sp	*		
Archaea			
Methanococcus jannaschii	*	*	*
Eucarya			
Saccharomyces cerevisiae	*	*	*

[a] Direct conversion of base into a ribonucleotide.
[b] Reversible conversion of bases to nucleosides.
[c] Interconversion by base alterations.
The pirS48560 sequence from *S. cerevisiae* has a 50.7% identity value in 274 aa segment of human nucleoside phosphorylase.

of the available sequences supports the hypothesis that purine nucleotide salvage routes were shaped to a considerable extent by a patchwork mechanism.

2. Materials and Methods

Amino acid and nucleotide sequences were extracted from the GenBank, EMBL, SWISS-PROT, and PIR databases. Additional information was retrieved from web sites which contain the complete genome sequences of *Escherichia coli* K12 (http://susi.bio.uni-giessen.de/usr/local/www/html/e cdc.html), *Haemophilus influenzae* (http://www.tigr.org/tdb/mdb/hidb/.html), *Mycoplasma genitalium* (http://www.tigr.org/ tdb/mdb/mgdb/.html), *Mycoplasma pneumoniae* (http://www.zmbh.uni-heidelberg.de/M_pneumoniae/MP_Home.html), *Synechocystis* sp. strain PCC 6803 (http://www.kazusa.or.jp:/cyano/cyano.html), *Methanococcus jannaschii* (http://www.tigr.org/tdb/mdb/mjdb/mjdb.html.), and *Saccharomyces cerevisiae* (http://genome-www.stanford.edu/ Saccharomyces/). Sequences were compared one against every other one using the Pearson algorithm, which is part of the FASTA program. These sequences were also compared with the BLAST (Basic Local Alignment Search Tool) algorithm that is available in the web site http://www.ncbi.nlm.nih.gov/BLAST/. This allowed the rapid identification of homologous protein sequences by focusing on regions shared by a pair of sequences in which a high

density of identities is present (Pearson and Lipman, 1988; Pearson, 1990). Multiple amino acid sequence aligments were constructed using the MACAW program, which produces aligments of ungapped blocks detected by pairwise comparisons of the sequences in a given set (Schuler *et al*., 1991). Consensus sequences derived from these comparisons were used to confirm the identification of the reported salvage enzymes, and to perform database searches of possible additional unidentified salvage enzymes. Results of these searches are shown in Tables I and II.

3. Results and Discussion

3.1. Purine phosphoribosyltransferases

Purine phosphoribosytranferases (PRTases) are pentosyltransferases (EC 2.4.2) that catalyze the kinetically irreversible conversion of purines into ribonucleotides by a pyrophosphate cleavage reaction

$$\text{purine} + 5 - \text{phospho} - \alpha - \text{D ribosyl} - 1 - \text{pyrophosphate} \rightarrow$$

$$5' - \text{NMP} + \text{pyrophosphate} \qquad (1)$$

Purine phosphoribosyltransferases are part of a larger family of widely distributed enzymes which participate in the biosyntheses of histidine, tryptophan, purines, pyrimidines, and NAD, as well as in nucleotide salvage pathways. PRTases form a monophyletic group of enzymes with a considerable degree of divergence, but their evolutionary relationships with other enzymes that participate in nucleotide salvage pathways such as nucleoside phosphorylases (PNPases), is supported by the presence of three relatively large motifs, one of which contains a conserved nucleoside-binding site with a ribose-binding region (Mushegian and Koonin, 1994).

Both PNPase and purine PRTase activities have been identified in a wide range of organisms of the three cell domains (Table I). The phylogenetic distribution of PRTase genes follows a complex pattern in which independent secondary losses and duplications appear to have taken place in widely separated lineages. For instance, in *H. influenzae* two non-contiguous sequences, *HI0674* and *HI0692*, encoding identical xanthine-guanine- PRTases have been identified (Fraser *et al*., 1995; Casari *et al*., 1995). Hypoxanthine- (HPRTase) and guanine-phosphoribosyltransferase (GPRTase) biochemical activities have been detected in cell extracts of the euryarchaeota *Methanococcus voltae* (Bowen *et al*., 1996), but the corresponding genes are absent in the closely related *M. jannaschii*, in whose genome the only identifiable purine phosphoribosyltranferase gene corresponds to adenine PRTase (Bult *et al*., 1996). No adenine PRTase activity has been found in the amitochondrial protist *Tritrichomonas foetus*. Purine PRTase activities also appear to be absent in the phylogenetically distant *Trichomonas vaginalis* and *Entoamoeba histolytica* (Berens *et al*., 1995).

3.2. ADENINE DEAMINASE, ADENOSINE DEAMINASE, ADENOSINE MONOPHOSPHATE DEAMINASE, AND ADENYLOSUCCINATE SYNTHASE

Adenosine deaminase (EC 3.5.4.4) is a well-studied monomeric enzyme that catalyzes the irreversible biosynthesis of inosine by the hydrolytic deamination of adenosine (Equation (3)). This reaction is mechanistically equivalent to (a) the adenine deaminase (ADA, EC 3.5.4.2) mediated synthesis of hypoxathine from adenine (Equation (2)); and (b) the formation of IMP from AMP (Equation (4)), catalyzed by adenosine monophosphate deaminase (EC 3.5.4.6). The products of these interconversion reactions are utilized as part of the general salvage pathway scheme (Figure 1):

$$\text{adenine} + H_2O \xrightarrow{\text{adenine deaminase}} \text{hypoxanthine} + NH_3 \quad (2)$$

$$\text{adenosine} + H_2O \xrightarrow{\text{adenosine deaminase}} \text{inosine} + NH_3 \quad (3)$$

$$\text{AMP} + H_2O \xrightarrow{\text{AMP deaminase}} \text{IMP} + NH_3 \quad (4)$$

The evolutionary relatedness between adenosine deaminase and AMP deaminase had long been suspected due to their common reaction mechanisms (Frieden *et al*., 1980), and was confirmed by sequence analysis of eubacterial and eukaryotic adenosine- and AMP deaminases (Chang *et al*., 1991). The reactions shown in Equations (3) and (4) are mechanistically identical to that of Equation (2), and suggest a monophyletic origin of all three different deaminases.

Three dimensional analysis of conversion patterns has shown that the similarities that adenosine deaminase shares with urease and phosphotriesterase (Jabri *et al*., 1995) can be extended to include AMP- and cytosine deaminases, and also dihydroorotases, allantoinases, hydantoinases, and imidazolonepropionases, all of which are part of the urease superfamily (Holm and Sander, 1997). Members of this superfamily are found in a wide variety of synthetic and degradative pathways, many of which involve nucleotides as intermediates or end-products (Holm and Sander, 1997). Thus, it is possible that they are the descendants of a protein of low specificity which catalyzed a wide range of primitive biochemical reactions involving heterocyclic compounds.

Deaminases may be relative latecomers in evolution. At 85 °C the half-lives of adenine and adenosine due to hydrolytic deaminations are 1.7 and 2.6 yr, respectively (Frick *et al*., 1987). These reactions can be enhanced by inorganic catalysts, as shown by the efficient clay-mediated hydrolytic deamination of both adenine and adenosine into hypoxanthine and inosine, respectively, under putative prebiotic

conditions (Strasak and Sersen, 1991). Under physiological conditions spontaneous deamination of adenine proceeds at such high rates that a specific enzyme mechanism involving an N-glycosylase has evolved to excise IMP from DNA and reduce its mutagenic consequences (Singer and Kusmierek, 1982). This raises the possibility that deaminases and other hydrolytic enzymes (including cytosine deaminase, N-glycosylases, RNaseH, and the exonuclease domain of DNA polymerase I) were not essential during the very early stages of metabolic evolution; the high spontaneous rates of the corresponding reactions would guarantee the availability of substrates required for the operation of these particular salvage pathways.

3.3. Inosine Monophosphate Dehydrogenase and Guanosine Monophosphate Reductase

Inosine monophosphate dehydrogenase (IMPDH, EC 1.1.1.205) is a biosynthetic NAD^+-dependent dehydrogenase which converts IMP into XMP by catalyzing the formation of a carbonyl group at C2 of inosinate (Equation (5)). This reaction is followed by the guanylate synthase-mediated displacement of this carbonyl oxygen in the keto form of XMP by an amino group, to produce GMP (Figure 1). In both the *Escherichia coli* and *H. influenzae* genomes the genes encoding these two enzymes are adjacent (data not shown). This colinearity may be due to the physiological link between their products, i.e., the requirement for the simultaneous availability of the two enzymes for the reaction sequence IMP $\rightarrow$ XMP $\rightarrow$ GMP to take place.

$$\text{inosine } 5' - \text{monophosphate} + NAD^+ + H_2O \longrightarrow$$

$$\text{xanthosine } 5' - \text{monophosphate} + NADH \qquad (5)$$

The opposite reaction, i. e., the reconversion of GMP into IMP (Equation (6)), is catalyzed by GMP reductase (EC 1.6.6.8), a $NADP^+$-dependent salvage deamination enzyme

$$\text{guanosine } 5' - \text{phosphate} + NADH \longrightarrow$$

$$\text{inosine } 5' - \text{phosphate} + NH_3 + NADP^+ \qquad (6)$$

Like PRTases, GMP reductase has a peculiar phylogenetic distribution that defies a simple explanation. It is absent in insects (Becker, 1974) and in rodents (Kanno *et al.*, 1989), but in humans GMP reductase deficiencies may be lethal (Henikoff and Smith, 1989). The *guaC* gene encoding GMP reductase is present in *E.coli* (Andrew and Guest 1988), but absent in the *H. influenzae* (Fleischmann *et al.*, 1995), *M. genitalium* (Fraser *et al.*, 1995), *M. jannaschii* (Bult *et al.*, 1996), and *S. cerevisiae* genomes. GMP reductase appears to be absent also among the apicomplexa, but has been detected in *Tritrichomonas foetus*, several kinetoplastida, and mammals (Beck *et al.*, 1994; Berens *et al.*, 1995).

The evolutionary relationship between IMPDH and GMP reductase was first reported by Andrews and Guest (1988). Recent database searches have shown that both enzymes are part of a β/α-barrel protein subclass in which a conserved phosphate-binding site is found near their C-terminus (Bork *et al*., 1995). Because IMPDH and GMP reductase are homologues and catalyze sequential steps in different purine metabolic pathways (Weber *et al*., 1992; Bork *et al*., 1995), it could be argued that they represent a particular case of the Horowitz retrograde hypothesis. The Horowitz hypothesis is not proben by the existence of sucessive biochemical reactions catalyzed by homologous enzymes (Fani *et al*., 1995), since during the early evolution of a given pathway the reactions could have been mediated by a primitive, less specific enzyme. This possibility is consistent with the homology of IMPDH and GMP reductase with other β/α barrel enzymes, many of which use heterocyclic substrates (Bork *et al*., 1995). Due to its hydrolytic instability, it is probable that GMP was absent from the primitive soup. Thus, GMP reductase may have appeared after its biosynthetic homologue IMPDH and the development of purine nucleotide anabolism.

4. A Hypothesis on the Origin of Purine Nucleotide Biosynthesis

PRPP and other phosphorylated sugars are unlikely components of the primitive soup (Larralde *et al*., 1995). Furthermore, it has been argued that ribose was not a significant component of the prebiotic environment (Shapiro, 1988; Larralde *et al*., 1995). Accordingly, purine PRTase activity could not have appeared when life originated but only after riboside-synthesizing metabolic routes had evolved. It is unlikely that the biosynthesis of purines was developed from prebiotic components, since all intermediates are ribosides which are unstable to hydrolysis.

On the other hand, it is generally agreed that purines and their prebiotic precursors were present in the primitive environment. This includes 5-amino-4-imidazolecarboxamide (AICA), which is a key intermediate in a potentially prebiotic synthesis of guanine and hypoxanthine (Sanchez *et al*., 1968). Two prebiotic routes to AICA can be envisoned. Sunlight irradiation of the HCN tetramer, diaminomaleonitrile produces 4-aminoimidazole-5-carbonitrile in relatively good yields (Ferris and Orgel, 1966) (Equation (7)).

Under basic conditions diaminomaleonitrile reacts with formamide and also produces 4-aminoimidazole-5-carbonitrile (AICN) (Equation (8)), whose hydrolysis also produces AICA (Equation (9)) in good yields of $\sim$20% (Oró and Kimball, 1962; Lowe *et al*., 1963; Sanchez *et al*., 1968).

AICA is not known to play any direct role in extant organisms. It is, however, an alternate substrate for adenine PRTase, an enzyme which not only catalyzes the formation of AMP from PRPP and adenine (Equation (1)), but also the direct condensation of PRPP with the prebiotic reagent AICA, to yield 5-amino-4-imidazolecarboxamide ribotide AICAR (K_{eq} = 9.7, pH 8) (Flaks *et al*., 1957),

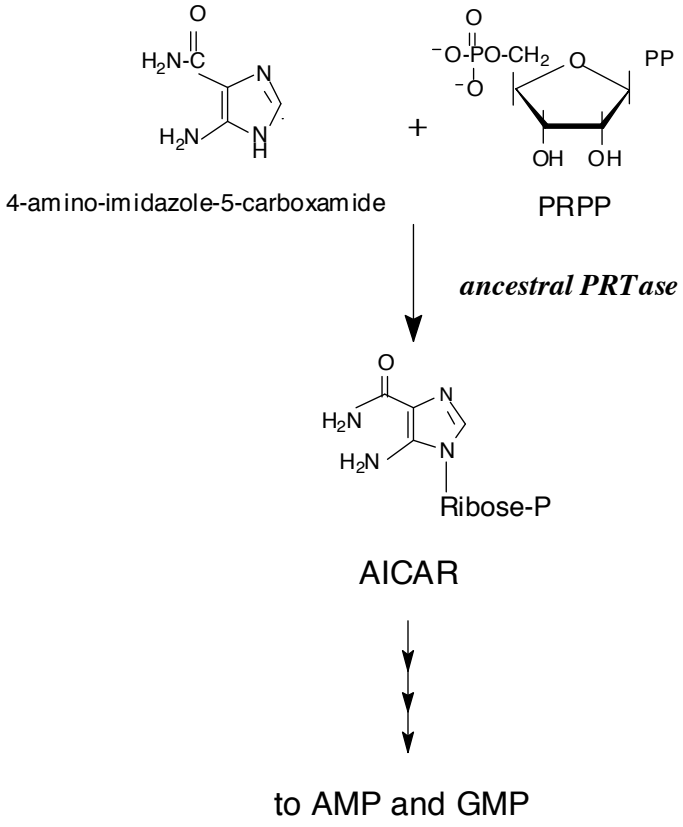
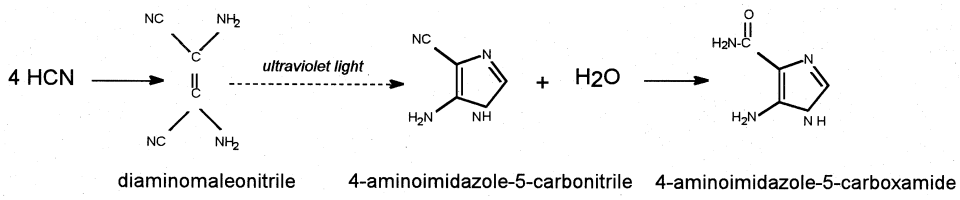


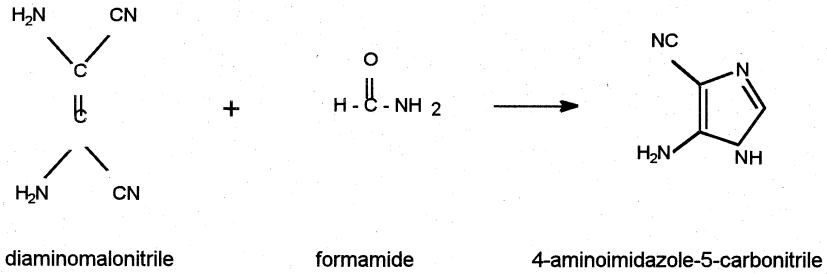
Figure 2. A hypothetical ancestral semi-enzymatic AICA-dependent biosynthesis of purine nucleotides.



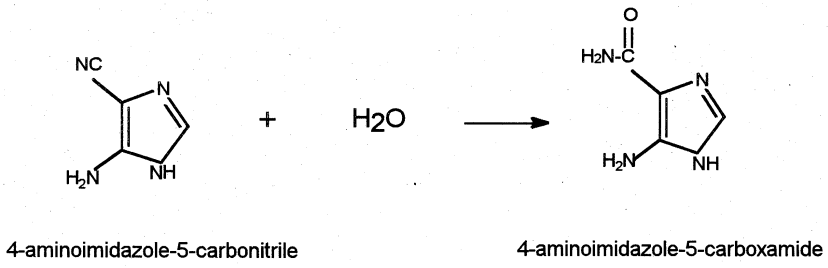
Equation (7).

which is an intermediate in the *de novo* biosynthesis of purine nucleotides (Figure 1). As shown in Figure 2, this raises the possibility of a simpler version of purine biosynthesis, in which an ancestral PRTase played a key role as an anabolic enzyme. The feasibility of this simpler pathway is currently being investigated.

The hypothetical route summarized in Figure 2 implies that a primitive version of the anabolic route of purine appeared when prebiotic reagents such as AICA were still available. This possibility is supported by the likelihood that important amounts of HCN and its derivatives were available in the primitive environment.



Equation (8).



Equation (9).

The half-life of AICA is 114 days at 100 °C and pH 7, and 108 days at 100 °C and pH 8 (Sanchez *et al.*, 1968). No data on the half-lives of AICA have been determined for lower temperatures, but it is known to be approximately a hundred-times more stable than its corresponding nitrile, AICN, whose half-life at O °C and pH 8 is 2×10^3 yr (Sanchez *et al.*, 1968). AICA and other HCN condensation products may have been available due to eutectic freezing (Miller and Orgel, 1974), or because their formation was enhanced by different prebiotic catalysts such as glyconitrile (Schwartz and Goverde, 1982). We do not speculate on the evolution of PRPP biosynthesis, and assume that the origin of phosphorylated sugar metabolism is related to the emergence of the ribose-phosphate backbone of RNA. This remains an open question, since the nature of the backbone of the first genetic polymer and the origin of the RNA world itself are still unknown (Lazcano and Miller, 1996).

5. Conclusions

Enzymes that participate in purine biosynthesis exhibit 35 to 65% of amino acid sequence similarity between distantly related organisms belonging to the Bacteria and Eucarya domains(Johnson *et al.*, 1987; Henikoff and Smith, 1989). By contrast, enzymes involved in purine nucleotide salvage pathways have complex phylogenetic histories. Availability of several complete cellular genomes has allowed insights into the versatility of salvage strategies followed by different organisms (Table II).

Whether a simple underlying pattern exists or not is not known. For instance, while PRTase and nucleoside phosphorylase activities appear to be essential, and have not been lost in *H. influenzae, M. genitalium*, and *M. pneumoniae* adenine deaminase is clearly disposable (Table II).

It is possible that the major salvage pathways were established during early stages of metabolic evolution, perhaps prior to the divergence of the three main cell lines (Mushegian y Koonin, 1996a). Unlike other macromolecules such as 16S rRNA, the peculiarities of the phylogenetic distribution of the salvage pathway enzymes show that these are not good universal molecular markers. Nevertheless, they provide a good model for the study of metabolic evolution. Direct uptake of purines (Hitchings, 1982) and pyrimidines from the prebiotic soup by the first organisms may be considered as a primitive form of salvage pathway, but it is unlikely that this included their ribose-derivatives. As argued here, the problems with the prebiotic synthesis of nuclosides and nucleotides suggest that the salvage pathways involving PRTases, deaminases, GMP reductases, and other enzymes that use nucleosides and nucleotides as substrates are not truly primordial, but evolved only after the biosynthesis of ribosides made them available.

Since comparative genomic analysis has demonstrated that gene order is not conserved in prokaryotic evolution (Mushegian and Koonin, 1996b; St. Jean and Charlebois, 1996; Tatusov *et al*., 1996; Watanabe *et al*., 1997), the dispersal in the prokaryotic genomes of homologous genes encoding pathway enzymes (data no shown) cannot be used to disprove the Horowitz hypothesis. However, analysis of the enzymatic sequences of the salvage enzymes does not support the retrograde hypothesis. With the exception of the pair formed by IMPDH and GMP reductase, we have found no other case in which sequential enzymatic steps in salvage pathways are catalyzed by homologous enzymes. This single example cannot be considered definitive proof of the retrograde hypothesis, since alternative explanations based on the existence of primitive less-specific enzymes are equally plausible, and are in fact supported by the homology of salvage enzymes with many other proteins.

The discovery that a significant portion of bacterial genomes is the outcome of ancient paralogous duplications (Fleischmann *et al*., 1995, Fraser *et al*., 1995; Koonin *et al*., 1995; Labedan and Riley, 1995), i.e., of gene duplications that took place prior to the divergence of the three major cell domains, is consistent with the hypothesis of the patchwork assembly of metabolic routes, and may be invoked to explain the evolution of salvage pathways. Each of the sets of enzymes discussed here is formed by homologous proteins that catalyze mechanistically similar reactions. This supports the idea that ancient pathways were mediated by enzymes of broad specificity (Waley, 1969; Ycas, 1974; Jensen, 1976), which may have participated in metabolic routes that today are not directly connected, such as the histidine and pyrimidine biosynthetic pathways.

Recognition of the role of ancient duplication events does not yield answers to questions related to the emergence of the original starter types, i.e., of the

enzymes that did not arise in this manner. In some cases, the starter types may stem from slow non-enzymatic reactions where the protein improved on a previously sluggish process (Lazcano and Miller, 1996), as in the case of the photochemical decarboxylation of orotic acid which yields uracil described by Ferris and Joshi (1979). Primitive pathways may have existed in which only a few steps were mediated by enzymes. In other cases, semi-enzymatic syntheses may have taken place (Miller and Lazcano, in prep.). One example may be the model of purine nucleotide biosynthesis presented here, in which prebiotic reagents like AICA, and biological catalysts such as adenine PRTase, both participate. In this regard, it is interesting to note that several enzymes including nitrogenase (Silver and Postgate, 1973), urease (Estermair *et al*., 1992), and adenine PRTase (Flaks *et al*., 1957), catalyze reactions involving HCN and/or its derivatives (cyanides, acetylene, cyanamide, dicyanamide, AICA), all of which may be prebiotic reagents. Since these compounds rarely participate in contemporary biochemical process, their use as alternate substrates by different enzymes raises the possibility that these are vestigial activities from a time in which metabolism depended on semi-enzymatic processes (Miller and Lazcano, in prep.).

Acknowledgements

This paper was presented at the Orleans ISSOL'96 Meeting, July 5–12, 1996. We thank Stanley L. Miller for calling to our attention to the paper by Estermaier *et al*. (1972), and Chris Sander for sharing with us his results prior to publication. We are indebted to two anonymous reviewers for their critical reading and many suggestions, which helped to improve an earlier version of this paper. Economic support from the ISSOL Travel Fund is gratefully acknowledged (A.B.). This paper was completed during a leave of absence of one of us (A. L.) as Visiting Professor at the Institut Pasteur (Paris), during which he enjoyed the hospitality of Professor Henri Buc and his associates at the Unité de Physicochimie des Macromolécules Biologiques. Support from the Manlio Cantarini Foundation is gratefully acknowledged. A.L. is an Affiliate of the NSCORT (NASA Specialized Center for Research and Training) in Exobiology at the University of California, San Diego.

References

Andrews, S. C. and Guest J. R.: 1988, *Biochem. J.* **255**, 35.
Beck, J. T., Zhao, S. and Wang, C. C.: 1994, *Exp. Parasitol.* **78**, 101.
Becker, J.-L.: 1974, *Biochemie* **56**, 1249.
Berens, R. L., Krug, E. C. and Marr, J. J.: 1995, in Marr, J. J. and Müller, M. (eds.), *Biochemistry and Molecular Biology of Parasites*, Academic Press, London, pp. 89–117.
Bork, P., Gellerich, J., Groth, H., Hooft, R. and Martin, F.: 1995, *Protein Science* **4**, 268.
Bowen, T. L., Lin, W. C. and Whitman, W. B.: 1996, *J. Bacteriol.* **178**, 2521.

Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Relch, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S. M., Weldman, J. F., Fuhrmann, J. L., Nguyen, D., Utterback, T. R., Kelley, J. M., Peterson, J. D., Sadow, P. W., Hanna, M. C., Cotton, M. D., Roberts, K. M., Hurst, M. A., Kaine, B. P., Borodovsky, M., Klenk, M. P., Fraser, C. M., Smith, H. O., Woese, C. R. and Venter, J. C.: 1996, *Science* **273**, 1058.

Casari, G., Andrade, M. A., Bork, P., Boyle, S. J., Ouzounis, C., Schneider, R., Tamames, J., Valencia, A. and Sander, C.: 1995, *Nature* **376**, 647.

Chang, Z., Nygaard, P., Chinault, A. C. and Kellems, R. E.: 1991, *Biochemistry* **30**, 2273.

Estermaier, L. M., Sieber, A. H., Lottspeich, F., Matern, D. H. M. and Hartmann, G. R.: 1992, *Angew. Chem. Int. Ed. Engl.* **31**, 620.

Fani, R., Lio, P. and Lazcano, A.: 1995, *J. Mol. Evol.* **41**, 760.

Ferris, J. P. and Joshi, P. C.: 1979, *J. Org. Chem.* **44**, 2133.

Ferris, J. P. and Orgel, L. E.: 1966, *J. Am. Chem.* Soc. **88**, 1074.

Flaks, J. G., Erwin, M. J. and Buchanan, J. M.: 1957, *J. Biol. Chem.* **228**, 201.

Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., McKenney, K., Sutton, G., FitzHugh, W., Fields, C., Gocayne, J. D., Scott, J., Shirley, R., Spriggs, T., Hedblom, E., Cotton, M. D., Utterback, T. R., Hanna, M. C., Nguyen, D. T., Saudek, D. M., Brandon, R. C., Fine, L. D., Fritchman, J. L., Fuhrmann, J. L., Geoghagen, N. S. M., Gnehm, C. L., McDonald, L. A., Small, K. V., Fraser, C. M., Smith, H. O. and Venter, J. C.: 1995, *Science* **269**, 496.

Fraser, C. M., Gocayne, J. D., White, O., Adams, M. D., Clayton, R. A., Fleischmann, R. D., Bult, C. J., Kerlavage, A. R., Sutton, G., Kelley, J. M., Fritchman, J. L., Weidman, J. F., Small, K. V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T. R., Saudek, D. M., Phillips, C. A., Merrick, J. M., Tomb, J. F., Dougherty, B. A., Bott, K. F., Hu, P. C., Lucier, T. S., Peterson, S. N., Smith, H. O., Hutchinson III, C. A. and Venter, J. C.: 1995, *Science* **270**, 397.

Frick, L., MacNeela, J. P. and Wolfenden, R.: 1987, *Bioorganic Chem.* **15**, 100.

Frieden, C., Kurz, L. C. and Gilbert, H. R.: 1980, *Biochemistry* **19**, 5303.

Henderson, J. F. and Paterson, R. P.: 1973, *Nucleotide Metabolism: an introduction*, Academic Press, New York.

Henikoff, S. and Smith, J. M.: 1989, *Cell* **58**, 1021.

Hitchings, G. H.: 1982, *Adv. Enzyme Regul.* **20**, 375.

Holm, L. and Sander, C.: 1997, *Proteins* (in press).

Horowitz, N. H.: 1945, *Proc. Natl. Acad. Sci. USA* **31**, 153.

Horowitz, N. H.: 1965, in Bryson, V. and Vogel, H. J. (eds.), *Evolving Genes and Proteins*, Academic Press, New York, pp. 15–23.

Jabri, E., Carr, M. B., Hausinger, R. P. and Karplus, P. A.: 1995, *Science* **268**, 998.

Jensen, R. A.: 1976, *Annu. Rev. Microbiol.* **30**, 409.

Johnson, D. H., Edstrom, J. E., Burnett, J. B. and Friedman, T. B.: 1987, *Gene* **59**, 77.

Kanno, H., Huang, I. Y., Kan, Y. W. and Yoshida, A.: 1989, *Cell* **58**, 595.

Koonin, E. V., Tatusov, R. L. and Rudd, K. E.: 1995, *Proc. Natl. Acad. Sci. USA* **92**, 11921.

Labedan, B. and Riley, M.: 1995, *Mol. Biol. Evol.* **12**, 980.

Larralde, R., Robertson, M. P. and Miller, S. L.: 1995, *Proc. Natl. Acad. Sci. USA* **92**, 8158.

Lazcano, A. and Miller, S. L.: 1996, *Cell* **85**, 793.

Lowe, C. U., Rees, M. W. and Markham, R.: 1963, *Nature* **199**, 219.

Miller, S. L. and Orgel, L. E.: 1974, *The Origins of Life on Earth*, Prentice-Hall, Englewood Cliffs.

Mushegian, A. R. and Koonin, E. V.: 1994, *Protein Science* **3**, 1081.

Mushegian, A. R. and Koonin, E. V.: 1996a, *Proc. Natl. Acad. Sci. USA* **93**, 10268.

Mushegian, A. R. and Koonin, E. V.: 1996b, *TIGS* **12**, 289.

Oparin, A. I.: 1938, *The Origin of Life*, Macmillan, New York.

Oró, J. and Kimball, A. P.: 1962, *Arch. Biochem. Biophys.* **96**, 293.

Pearson, W. R.: 1990, in Doolittle, R. F. (ed.), *Methods in Enzymology* **138**, pp. 63–98.

Pearson, W. R. and Lipman, D. I.: 1988, *Proc. Natl. Acad. Sci. USA* **85**, 2444.

Pennisi, E.: 1996, *Science* **272**, 1098.

Sanchez, R. A., Ferris, J. P. and Orgel, L. E.: 1968, *J. Mol. Biol.* **38**, 121.
Schuler, G. D., Altschul, S. F. and Lipman, D. J.: 1991, *Proteins Struct. Funct. Genet.* **9**, 180.
Schwartz, A. W. and Goverde, M.: 1982, *J. Mol. Evol.* **18**, 351.
Shapiro, R.: 1988, *Origins Life Evol. Biosphere* **18**, 71.
Silver, W. S. and Postgate, J. R.: 1973, *J. Theoret. Biol.* **40**, 1.
Singer, B. and Kusmierek, J. T.: 1982, *Ann. Rev. Biochem.* **52**, 655.
Sogin, M. L.: 1994, in S. Bengtson (ed.), *Early Life on Earth: Nobel Symposium No. 84*, Columbia University Press, New York, pp. 181–192.
St. Jean, A. and Charlebois, R. L.: 1996, *J. Bacteriol.* **178**, 3860.
Strasak, M. and Sersen, F.: 1991, *Naturwissenschaften* **78**, 121.
Tatusov, R. L., Mushegian, A. R., Bork, P., Brown, N. P., Hayes, M. S., Borodovsky, M., Rudd, K. E. and Koonin, E. V.: 1996, *Current Biol.* **6**, 279.
Tham, T. N., Ferris, S., Kovacic, R., Montagnier, L. and Blanchard, A.: 1993, *J. Bacteriol.* **175**, 5281.
Waley, S. G.: 1969, *Comp. Biochem. Physiol.* **30**, 1.
Wang, C. C.: 1991, in Toft, C. A., Aeschlimann, A. and Bolis, L. (eds.), *Parasite-Host Association: Coexistence or Conflict?* Oxford University Press, Oxford, pp. 258–273.
Watanabe, H., Mori, H., Itoh, T. and Gojobori, T.: 1997, *J. Mol. Evol.* **44** (Suppl. 1), S57.
Weber, G., Nakamura, H., Natsumeda, Y., Szekeres, T. and Nagai, M.: 1992, *Adv. Enzyme Regul.* **32**, 57.
Ycas, M.: 1974, *J. Theoret. Biol.* **44**, 145.
Zubay, G.: 1993, *Chemtracts – Biochem. Mol. Biol.* **4**, 317.