

RESEARCH ARTICLE

Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines

Alberto Vázquez-Salazar¹, Arturo Becerra¹, Antonio Lazcano^{1,2*}

1 Facultad de Ciencias, Universidad Nacional Autónoma de México, Cd. Universitaria, Cd. de México, México, **2** Miembro de El Colegio Nacional, Centro, Cd. de México, México

* alar@ciencias.unam.mx



Abstract

Background

The imidazole group is an ubiquitous chemical motif present in several key types of biomolecules. It is a structural moiety of purines, and plays a central role in biological catalysis as part of the side-chain of histidine, the amino acid most frequently found in the catalytic site of enzymes. Histidine biosynthesis starts with both ATP and the pentose phosphoribosyl pyrophosphate (PRPP), which is also the precursor for the *de novo* synthesis of purines. These two anabolic pathways are also connected by the imidazole intermediate 5-aminoimidazole-4-carboxamide ribotide (AICAR), which is synthesized in both routes but used only in purine biosynthesis. Rather surprisingly, the imidazole moieties of histidine and purines are synthesized by different, non-homologous enzymes. As discussed here, this phenomenon can be understood as a case of functional molecular convergence.

Results

In this work, we analyze these polyphyletic processes and argue that the independent origin of the corresponding enzymes is best explained by the differences in the function of each of the molecules to which the imidazole moiety is attached. Since the imidazole present in histidine is a catalytic moiety, its chemical arrangement allows it to act as an acid or a base. On the contrary, the *de novo* biosynthesis of purines starts with an activated ribose and all the successive intermediates are ribotides, with the key β -glycosidic bondage joining the ribose and the imidazole moiety. This prevents purine ribonucleotides to exhibit any imidazole-dependent catalytic activity, and may have been the critical trait for the evolution of two separate imidazole-synthesizing-enzymes. We also suggest that, in evolutionary terms, the biosynthesis of purines predated that of histidine.

Conclusions

As reviewed here, other biosynthetic routes for imidazole molecules are also found in extant metabolism, including the autocatalytic cyclization that occurs during the formation of creatine from creatine phosphate, as well as the internal cyclization of the Ala-Ser-Gly motif of

OPEN ACCESS

Citation: Vázquez-Salazar A, Becerra A, Lazcano A (2018) Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines. PLoS ONE 13(4): e0196349. <https://doi.org/10.1371/journal.pone.0196349>

Editor: Israel Silman, Weizmann Institute of Science, ISRAEL

Received: February 8, 2018

Accepted: April 11, 2018

Published: April 26, 2018

Copyright: © 2018 Vázquez-Salazar et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Alberto Vázquez-Salazar is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 404507 from CONACYT. Financial support to Arturo Becerra from PAPIIT-UNAM (IN223916) is gratefully acknowledged.

Competing interests: The authors have declared that no competing interests exist.

some members of the ammonia-lyase and aminomutase families, that lead to the MIO cofactor. The diversity of imidazole-synthesizing pathways highlights the biological significance of this key chemical group, whose biosyntheses evolved independently several times.

Introduction

The imidazole group is a five-membered heterocyclic chemical compound containing a tertiary nitrogen and an imino group. It is present in the structure of several key molecules of major biological significance, most notably purines and histidine. The structure of the unsubstituted imidazole ring has two nitrogen atoms in positions 1 and 3 (also named pyrrole- and pyridine-nitrogen atoms, respectively) (Fig 1). The pyrrole nitrogen is a very weak acid with a $pK_a = 14.4$. In contrast, the pK_a of the pyridine nitrogen is 6.9, which allows it to accept a proton under neutral pH conditions [1]. This phenomenon does not occur in the imidazole moiety of purines, but it takes place in the imidazole side-chain of histidine, whose pK_a value is in the range of 6 to 7 as a result of the electron-withdrawing inductive effect of the protonated amino group [2,3]. Bioinformatic analyses have shown that histidine is the most frequently found catalytic residue in the active sites of enzymes [4]. This phenomenon may be explained by the level of ionization of the imidazole moiety of histidine, which is estimated to be between 9–50%, resulting in its well-known ability to act as a general acid or base to donate or accept

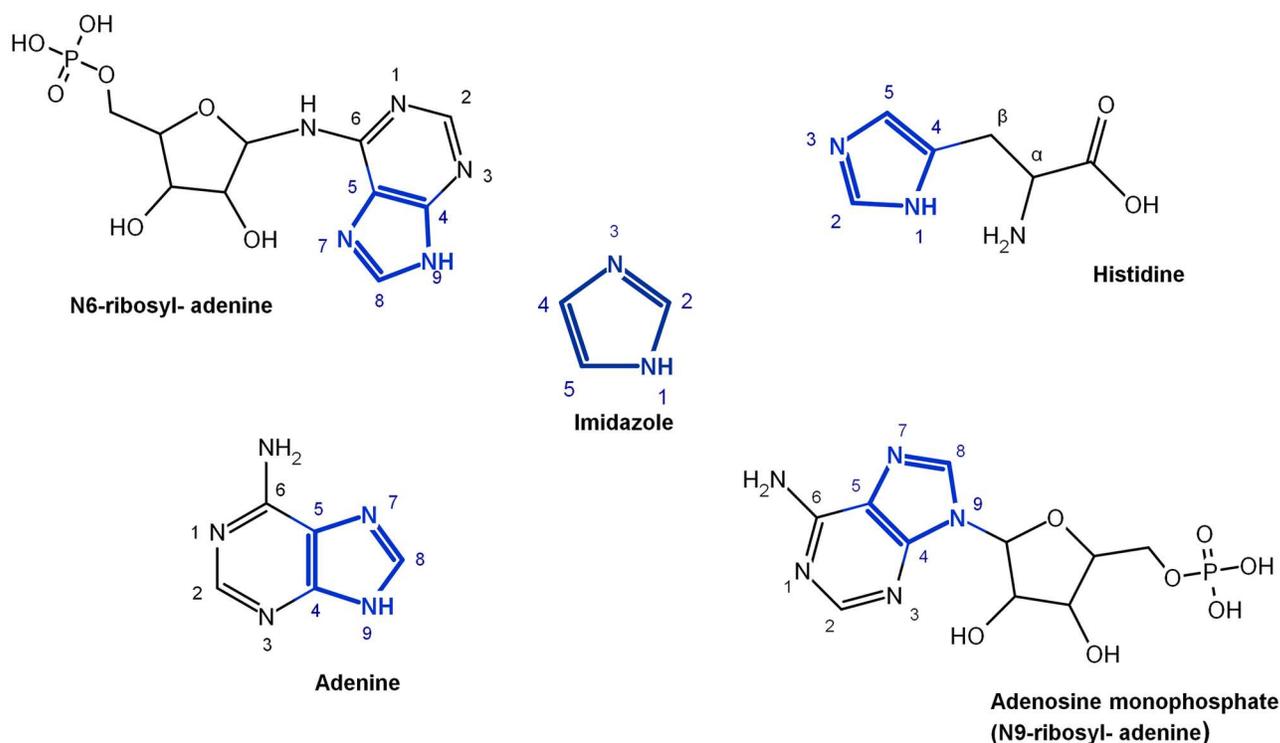


Fig 1. Structure of imidazole and imidazole-derivatives with atom numbering. The unsubstituted imidazole group is shown in the center. Other important imidazole molecules are shown: histidine (2-amino-3-(1H-imidazol-4-yl)), adenine (9H-purin-6-amine), adenosine monophosphate (N9-riboseyl-adenine), and N6-riboseyladenine (which synthesis has been reported in prebiotic chemistry experiments).

<https://doi.org/10.1371/journal.pone.0196349.g001>

protons during a chemical reaction. This property, together with the stable complexes that it can form with different metallic cations such as Cu^{2+} , Co^{2+} , Zn^{2+} , and Mn^{2+} [1,5,6], plays a key role in our understanding of the catalytic propensity of histidine.

Abiotic synthesis of adenine and histidine

Adenine was the first imidazole-bearing molecule synthesized under abiotic conditions from an aqueous solution of ammonium cyanide (NH_4CN) [7]. Since then, many other purines have also been synthesized in prebiotic simulations, including guanine and xanthine [8], hypoxanthine [9], 8-hydroxymethyladenine [10], and 2,6-diaminopurine [11]. Outstandingly, the imidazole group, as well as its 2-methyl derivative were obtained by Oró et al. [12] using glyoxal, aldehydes and ammonia as precursors.

The first attempt to synthesize histidine under possible prebiotic conditions was reported by Shen et al. [13,14]. In the proposed reaction, D-erythrose and formamidine are condensed, leading to imidazole-4-acetaldehyde, which then forms histidine via a Strecker synthesis with a relatively high yield of 3.5% based on the ratio of His/erythrose (Fig 2). As argued below, a detailed examination of this reaction suggests that the conditions of synthesis are prebiotically unrealistic, and that it is unlikely it took place in the primitive Earth.

Even though the *a priori* selection of the D-enantiomer of erythrose is irrelevant since the stereocenters do not partake in the synthesis, and the reaction would work if other 4-carbon aldoses were used as precursors, D-erythrose and D-threose are minor, unstable products of the formose reaction [15], and therefore could not have been available in the concentration required for the reaction reported by Shen et al. [13,14] to take place. In addition, formamidine is a labile compound that is rapidly hydrolyzed into formic acid and ammonia [16], which also makes it very difficult to achieve the concentration required for the proposed reaction (0.3 M). As concluded elsewhere, all this data strongly hinders the prebiotic significance of this synthesis [17].

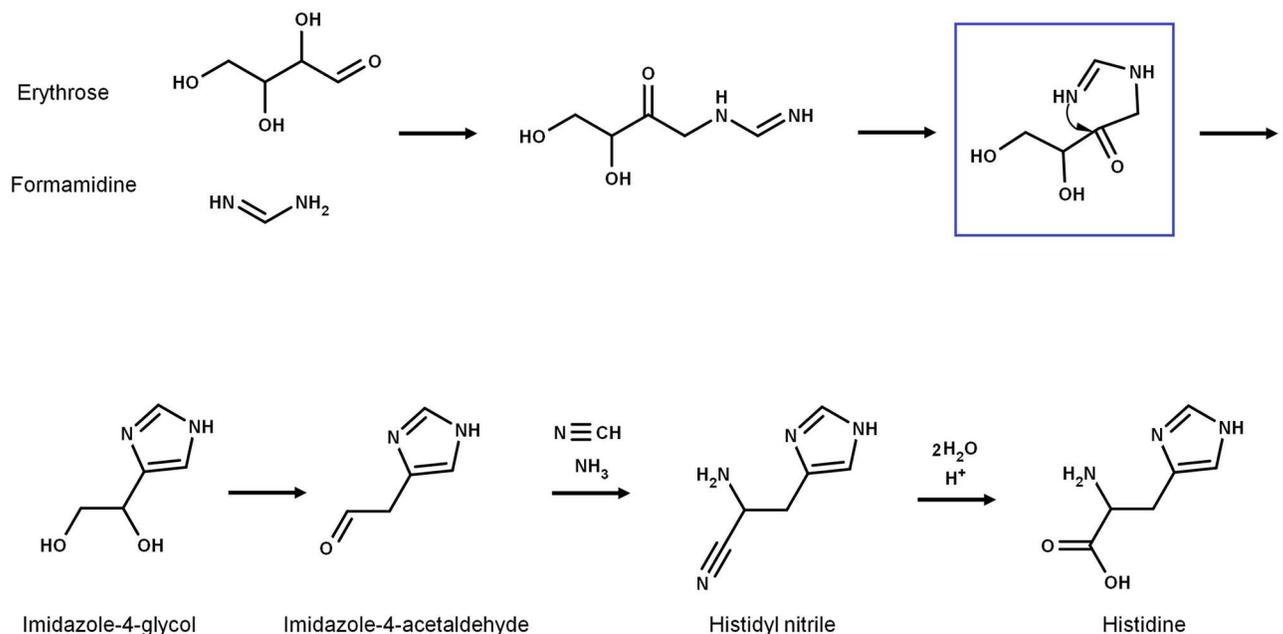


Fig 2. Abiotic synthesis of histidine. The synthesis was proposed by [13,14] and occurs via Strecker synthesis. As shown in the upper part of the scheme, an Amadori rearrangement gives place to the histidine imidazole moiety [cf. ref 17].

<https://doi.org/10.1371/journal.pone.0196349.g002>

Nevertheless, it is quite interesting that in the reaction mechanism proposed by Shen et al. [13,14] the formation of the imidazole group is the product of an Amadori rearrangement (Fig 2), which is also the mechanism involved in the synthesis of the imidazole moiety of imidazole glycerol phosphate (IGP) by the IGP synthase (HisHF) enzyme during histidine biosynthesis. However, as argued elsewhere [17], all the evidence suggests that there is no direct evolutionary connection between the abiotic synthesis reported by Shen et al. [13,14] and the extant histidine anabolism.

As of today, no imidazole-bearing amino acid has been produced in Miller-Urey-type experiments or in other type of laboratory simulations, and the search of histidine and its degradation products in carbonaceous chondrites has also yielded negative results. These results suggest that histidine as such may have not been present in the prebiotic environment [17–19] but, as argued here, may be in fact a product of very early biological evolution.

Role of imidazolides in the evolution of catalysis

Although it has been shown that clays, metallic cations, and relatively simple compounds like proline and other amino acids can affect the rates of chemical reactions under possible primordial conditions, the emergence of biological catalysis remains one of the key questions in modern biology.

The discovery of the catalytic properties of RNA [20,21] provided support for the RNA world proposal, where it is assumed that RNA molecules played a key role in both heredity and catalysis during the very early stages of biological evolution.

Although how the RNA world evolved from the prebiotic environment remains an open issue, there have been advances in our understanding of the role that imidazole and imidazole-derivatives could have played during this evolutionary stage. For instance, it has been demonstrated that 2-aminoimidazole can act as an activating group for non-enzymatic RNA copying by a mechanism involving an imidazolium-bridged dinucleotide intermediate at pH = 7 [22]. Unsubstituted imidazole may have also played a role during the RNA world, as suggested by the demonstration that RNA cleavage can be catalyzed by imidazole in buffer [23]. The unsubstituted imidazole group can also act as a cofactor for the hepatitis delta virus (HDV) self-cleaving ribozyme [24]. The reaction mechanism of this ribozyme involves a nucleophilic attack of the 2'-OH on the phosphorus in the phosphodiester backbone, where the catalytic cytosine accepts the proton from the attacking 2'-hydroxyl group. As shown by Perrotta et al. [24], the incubation with unsubstituted imidazole buffer (200 mM, pH 7.4) restores the catalytic activity of a mutated HDV ribozyme. This shows that imidazole by itself can act as a co-catalyst in reactions involving RNA molecules, which could be of key significance in an RNA world scenario, and of considerable significance in our understanding of the early evolution of biological catalysis.

The catalytic activity of His-containing di- and tri-peptides has also been investigated. Histidyl-histidine (His-His) catalyzes the dephosphorylation of deoxyribonucleoside monophosphate, the hydrolysis of oligo (A)₁₂, and the oligomerization of 2',3'-cAMP under cyclic wet-dry reaction conditions [25]. More recent studies have shown that seryl-histidine (Ser-His) catalyzes the oligomerization of trimers of imidazole-activated nucleotides [26,27], and catalyze peptide bond formation between two amino acids, an activity also found in seryl-histidyl-glycine (Ser-His-Gly) [28,29].

Evolution of enzyme catalytic activity from an RNA world

Although there have been alternative proposals to explain the origin of ribosome-mediated protein synthesis based on the simultaneous appearance of both RNA and proteins [30,31], it

can also be argued that the ribosome first appeared as an evolutionary outcome of the interaction of prebiotically synthesized amino acids with catalytic RNA molecules. How enzymes first evolved remains an open question, but it is reasonable to assume that their active sites must be one of their most conserved portions. However, if histidine was absent in the primitive Earth, how can its ample distribution in the active sites of enzymes be explained?

In what may be the first attempt to explain the origin of histidine biosynthesis, over 40 years ago Harold White III correlated the start of the histidine biosynthetic pathway, which involves a condensation reaction between ATP and the sugar phosphoribosyl pyrophosphate (PRPP), with the possibility of an ancient metabolism mediated by nucleic acid enzymes [32]. In his evolutionary biochemistry approach, White proposed that histidine was, in fact, the molecular vestige of an ancient catalytic nucleotide that was part of the RNA world [32,33]. According to White, this possibility is reinforced by a) the well-known fact that the histidine anabolic pathway is connected with the *de novo* synthesis of purines by the usage of PRPP and the 5-aminoimidazole-4-carboxamide ribonucleotide (AICAR) intermediate; b) His is the only known amino acid with a ribonucleotide-starting biosynthesis; and c) histidine is the only imidazole-bearing amino acid.

While it is true that histidine is the only biological amino acid with an imidazole side chain, the other two issues merit a reexamination. As underlined by White [32], during the first step of the histidine biosynthesis, PRPP, whose synthesis in prebiotic conditions has been recently reported by Akouche et al. [34], is condensed with a molecule of ATP to form N⁵-phosphoribosyl-ATP (PRATP) (Fig 3). Quite surprisingly, however, the imidazole component of the ATP does not partake in the biosynthesis of the imidazole group of histidine. Instead, the histidine's imidazole moiety is biosynthesized *de novo* by a mechanism completely different from that of purines.

As shown on Fig 3, histidine- and purine biosyntheses are indeed connected by the imidazole intermediate AICAR, which is formed independently in the two pathways. In the purine route it is formed from succinyl-AICAR (SAICAR), and from N⁷-((5'-phosphoribuloyl) formimino)-5-aminoimidazole-4-carboxamide-ribonucleotide (PRFAR) in the histidine pathway, but it does not play any further role in the subsequent steps of the histidine anabolism. Instead, it is recycled directly into the purine pathway, where it is used as a substrate by the bifunctional enzyme AICAR transfromylase/inosine monophosphate (IMP) cyclo-hydrolase (PurH) [EC: 2.1.2.3/EC: 3.5.4.10], that catalyzes the final two steps in the route and synthesizes IMP (Fig 3). We suggest that the incorporation of AICAR from the histidine biosynthetic route into the *de novo* synthesis of purines can be best understood as an intracellular salvage pathway (Fig 3), in which the imidazole compound is used to supply the requirements of purine synthesis.

Biosynthesis of the imidazole group in histidine and purines

In the histidine biosynthetic pathway, the enzyme imidazole glycerol phosphate synthase (IGPS) [EC: 2.4.2.-/EC:4.1.3.-] catalyzes the conversion of PRFAR and glutamine to imidazole glycerol phosphate (IGP), AICAR, and glutamate [35]. As mentioned above, AICAR is discarded. In other words, in what appears to be a rather convoluted, non-parsimonious route, the histidine's imidazole moiety is synthesized *de novo* as part of IGP, while AICAR is recycled into purine synthesis.

The bacterial IGPS enzyme is a heterodimer formed by the *hisH* and *hisF* gene products (Fig 4A), where *hisH* encodes a glutamine amidotransferase (GATase type I) subunit [EC: 2.4.2.-], and *hisF* encodes a cyclase subunit [EC:4.1.3.-] that adopts the (β/α)₈-barrel fold [36,37] (Fig 4B). Binding of PRFAR to the HisF active site stimulates the glutaminase activity

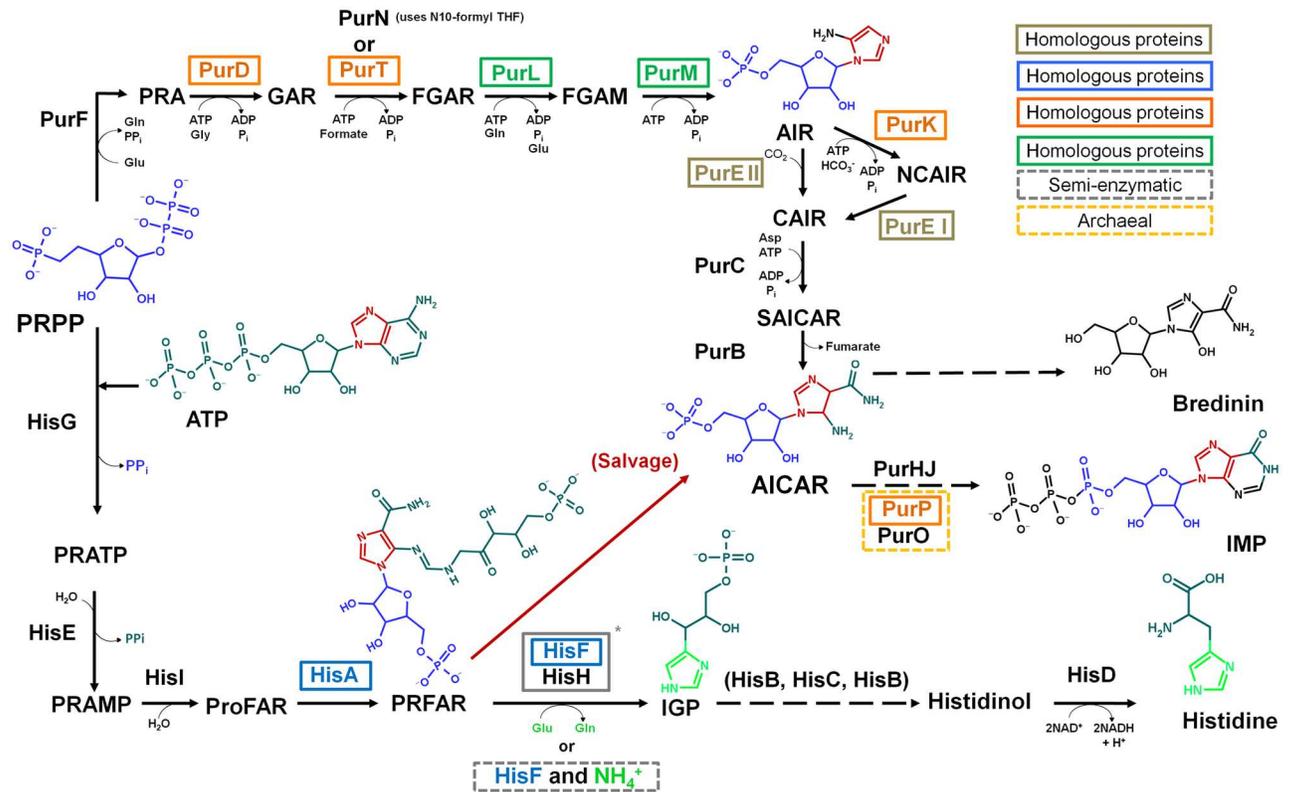


Fig 3. Pathways for the biosynthesis of histidine (bottom) and purines (top). Both routes start with PRPP and are connected by AICAR. The enzymes for each step are depicted. The enzyme name's color and the boxes indicate homologous proteins within pathways as well as the semi-enzymatic reaction. It is evident how the imidazole group of the ATP molecule along with the pentose of PRPP become part of AICAR, and goes into the purine synthesis, while the imidazole in IGP forms independently. The red arrow indicates the salvage pathway in which AICAR is recycled into the de novo synthesis of purines. The dashed arrows indicate two or more steps in the route, and for simplicity the cofactors and coenzymes used in these steps have been omitted. *The complex formed by the HisF and HisH proteins is named imidazole glycerol phosphate synthase (IGPS). The full names of the most relevant metabolites and enzymes are given in the text.

<https://doi.org/10.1371/journal.pone.0196349.g003>

in the HisH enzyme, revealing an allosteric regulation [38]. The synthesis of IGP can also take place *in vitro* in the absence of the HisH protein under a high concentration of NH₄⁺ [39]. This phenomenon was also shown to occur *in vivo* in the gammaproteobacteria *Klebsiella pneumoniae*, where the *hisH* gene was mutated but the reaction was restored in high concentrations of NH₄⁺ [40] (Fig 5). As argued by Lazcano and Miller [41], the fact that the pathway can take place in the absence of HisH suggests that a simpler pathway with lesser number of enzymes may have existed in the distant past. This provides an example of the semi-enzymatic origin of metabolic pathways, in which a primitive biosynthetic route could have exploited a favorable chemical environment as well as available enzymes (Fig 5) [41].

As summarized in Fig 3, the ATP-dependent enzyme phosphoribosylformylglycinamide cyclo-ligase [EC:6.3.3.1] (PurM) synthesizes 1-(5'-Phosphoribosyl)-5-aminoimidazole (AIR) from 1-(5'-phosphoribosyl)-N-formylglycinamide (FGAM). This enzyme is present in all purine biosynthetic pathway variations found in the three domains of life. Structurally, each PurM monomer is divided into two domains, an A- or N-terminal domain, and a B- or C-terminal domain (Fig 4c) [42]. The active enzyme is a homodimer formed by the interaction of two four-stranded β -sheets (one from each subunit) forming an eight-stranded β -barrel-like structure (Fig 4C) [42]. AIR is then enzymatically processed into SAICAR, which is the

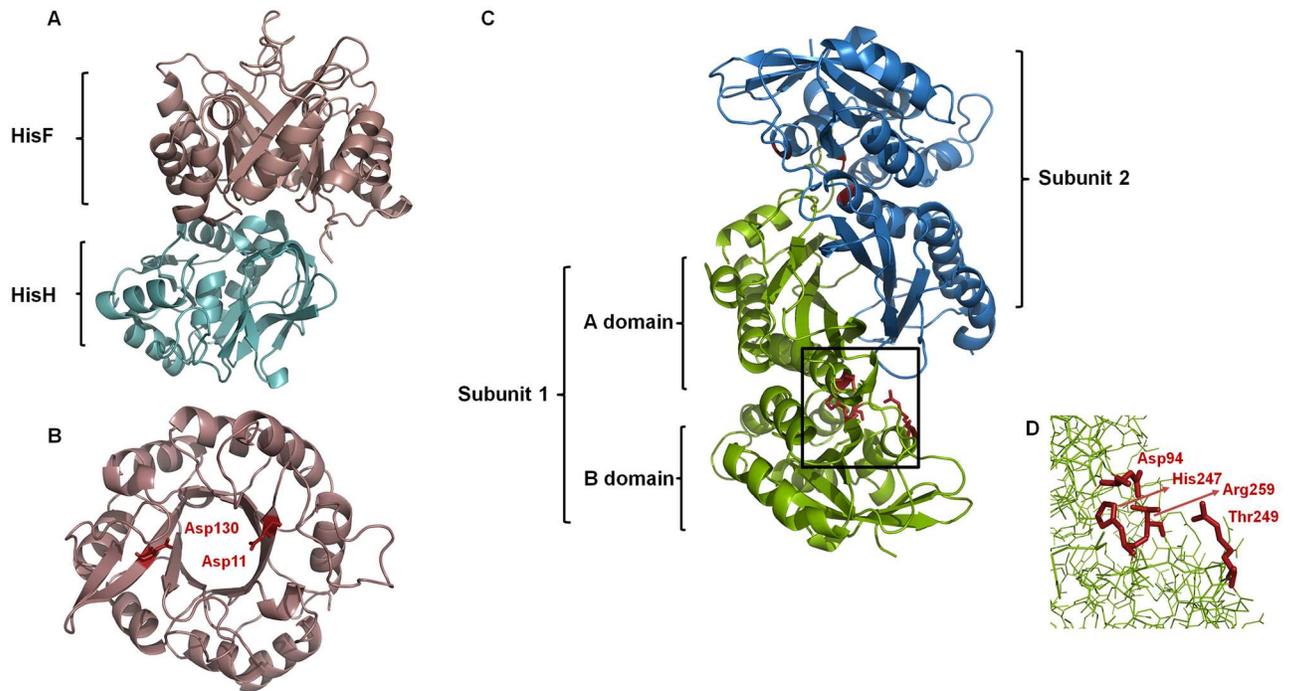


Fig 4. Crystallographic structures of HisF and PurM. (A) Structure of the imidazole glycerol phosphate synthase (IGPS) of *Thermotoga maritima* (PDB:1GPW). (B) Structure of the cyclase (HisF) subunit of *Thermotoga maritima* (PDB: 1THF). Catalytic residues (Asp11 and Asp130) are indicated. (C) Structure of the phosphoribosylformylglycinamide cyclo-ligase (PurM) of *Escherichia coli* (PDB: 1CLI). (D) Proposed active site for PurM, which comprises the residues Asp94, His247, Thr249, and Arg259.

<https://doi.org/10.1371/journal.pone.0196349.g004>

substrate of the enzyme adenylosuccinate lyase [EC:4.3.2.2] (PurB), that catalyzes the synthesis of AICAR [43].

In summary, HisF and PurM are not homologues but they catalyze the biosyntheses of the imidazole motif in the histidine and purine pathways (Fig 3) and do so by completely different

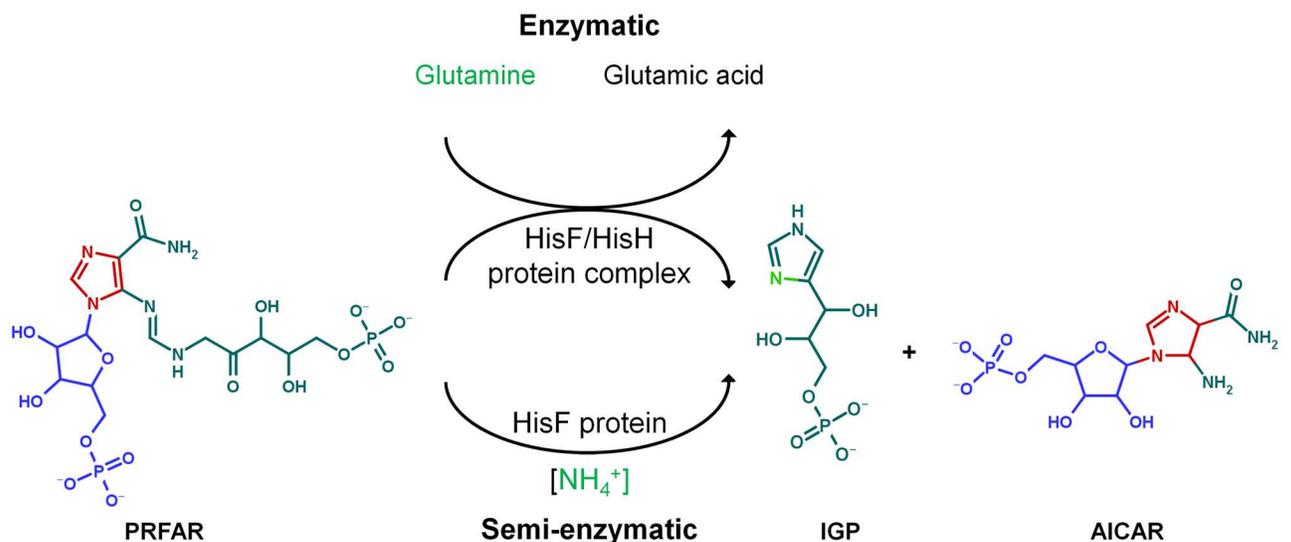


Fig 5. Enzymatic and semi-enzymatic synthesis of imidazole glycerol phosphate (IGP). The enzymatic branch needs both HisF and HisH proteins. HisH transfers nitrogen from glutamine, forming glutamate. In contrast, the semi-enzymatic branch only requires the HisF protein, since the high concentration of NH_4^+ in the medium can replace the transferase activity of HisH. In both cases, IGP and AICAR are formed (modified from ref. [41]).

<https://doi.org/10.1371/journal.pone.0196349.g005>

mechanisms. The presence of two different independent syntheses of imidazole intermediates in highly conserved, widely distributed routes is rather intriguing and becomes an interesting question from an evolutionary perspective. How can the existence of such biochemically redundant enzymatic complexes in highly connected, early pathways be understood? Here we analyze the active sites of both enzymes and the reactions they catalyze, and propose an explanation for their polyphyletic origin.

Materials and methods

Sequence analysis

The sequences of bacterial HisF (*Escherichia coli* K-12 MG1655: b2025), and PurM (*Escherichia coli* K-12 MG1655: b2499) proteins were retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [44]. Pairwise sequence alignment was performed using the Needleman-Wunsch algorithm as implemented in the EMBOSS Needle program of the European Molecular Biology Open Software Suite [45].

Structural analysis

Visualization and analysis of protein structures were performed using the PyMol package [46]. Structure comparisons were performed using the FATCAT algorithm [47] included in the protein comparison tool of the Protein Data Bank (PDB) web server. DaliLite v3 was used to search for structural related proteins [48].

Structure-based dendrogram construction

The available crystal structures of the HisF, PurM, and their paralogues were retrieved from the PDB. A total of 38 structures available as of January 2018 (S1 Table) were analyzed following a previously described procedure [49], only structures with no bounded ligands and no reported mutations were considered in this work. To avoid redundancy, for each protein only the structures with the best resolution were selected if two or more of them were available for the same species. Pairwise structural comparisons were performed using the Secondary Structure Matching (SSM) program of the European Protein Data Bank (PDBe) webserver [50]. The structural alignment score (SAS) for each protein-protein comparison was calculated following the formula: $(\text{RMSD} \times 1000) / \text{number of aligned residues}$ [51]. Geometric distance values were calculated using the SAS and then transformed into evolutionary distances using the FITCH program of the PHYLIP package [52]. Finally, dendrograms were constructed using the DRAWTREE program included in PHYLIP. Dendrograms were visualized and edited using FigTree version 3.2 [53].

Results

Imidazole-biosynthetic enzymes HisF and PurM: A case of evolutionary convergence

The *E. coli* PurM (354 aa) and HisF (258 aa) proteins share only 12.2% of overall sequence identity (21.4% similarity) (S1 Fig). When the crystallographic structures of the two proteins are superimposed, the root-mean-square deviation (RMSD) of the α -carbons is 8.27 Å (Fig 6A). This very high RMSD value reflects of the very different domain architectures of the two enzymes. As noted above, HisF has the $(\beta/\alpha)_8$ -barrel fold and belongs to the ribulose-phosphate binding barrel superfamily, while PurM adopts a completely different fold and is part of the PurM superfamily. Considering these structural differences, it can be readily concluded that

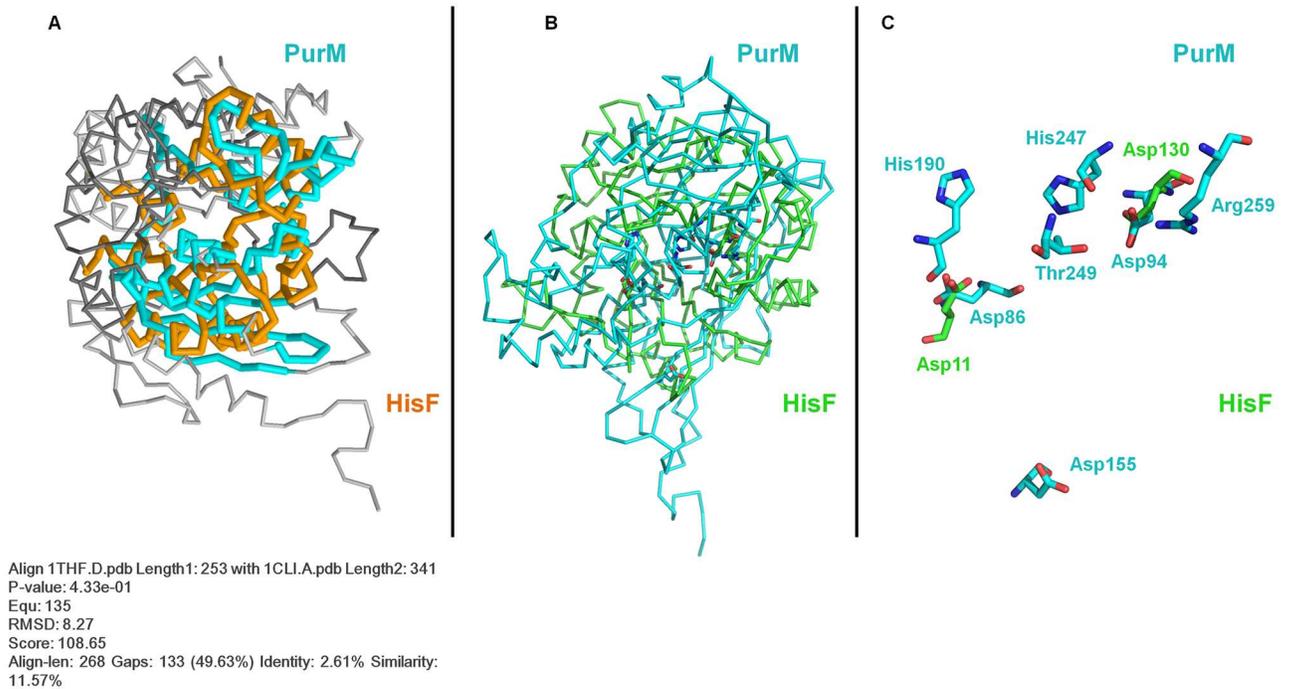


Fig 6. Structural analysis of the HisF and PurM proteins. (A) Structural alignment of HisF from *Thermotoga maritima* (PDB: 1THF) and PurM from *Escherichia coli* (PDB: 1CLI) using the FATCAT algorithm (B) Structural alignment of the active sites of HisF and PurM using the *align* command of PyMol. (C) Catalytic residues that comprise the active sites of HisF and PurM isolated from the alignment in B.

<https://doi.org/10.1371/journal.pone.0196349.g006>

these are different enzymes with different evolutionary histories that share no common ancestor. The evolutionary relationship between HisF and its structural homologues is depicted in the structural phylogeny shown in Fig 7. The branches more closely related to HisF are those corresponding to the enzyme N⁻[(5'-phosphoribosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase [EC:5.3.1.16] (HisA). It has been shown that HisF and HisA share a similar internal organization that can be explained as two paralogous modules half the size of the extant sequences [54]. An ancestral module duplicated and gave rise to the ancestral *hisA* gene which, in turn, underwent another duplication and fused, giving rise to the *hisF* gene [35,54,55]. This shared evolutionary history explains the closeness between the branches.

Interestingly, HisA catalyzes an Amadori rearrangement that isomerizes the aminoaldose motif in ProFAR into the aminoketose PRFAR. The same chemical rearrangement is responsible for the imidazole ring closure by HisF, and also occurs in the reaction catalyzed by the tryptophan biosynthetic enzyme phosphoribosylanthranilate isomerase [EC:5.3.1.24] (TrpF), which transforms N-(5-phospho-beta-D-ribose) anthranilate (PRAn) into 1-(2-carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate (CdRP) [56,57]

Once synthesized, CdRP is then processed by the enzyme indole-3-glycerol phosphate synthase [EC:4.1.1.48] (TrpC), which catalyzes the formation of 1-C-(indol-3-yl) glycerol 3-phosphate (rCdRP), CO₂, and H₂O. Both TrpF and TrpC are homologues to the HisF/HisA pair, and their evolutionary relationship is shown in Fig 7. Although a prebiotic synthesis of tryptophan has been reported [58], it is likely that this amino acid is a latecomer in evolution [59]. This implies that, in fact, the HisF/HisA pair is more ancient than the TrpF and TrpC proteins. This assumed antiquity is concordant with the key catalytic role played by histidine in many extant enzymes (see above).

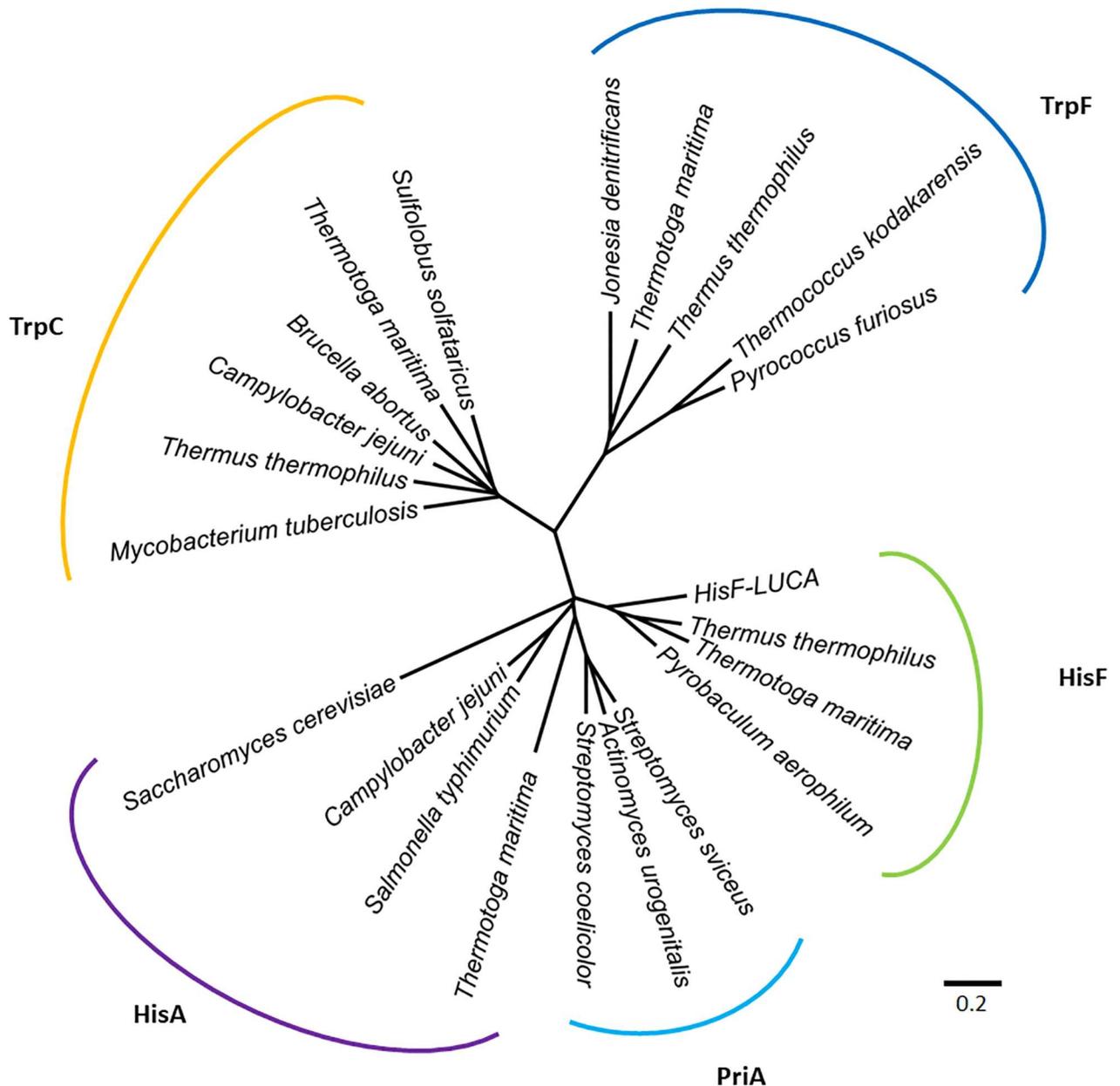


Fig 7. Structural phylogeny of HisF and its homologues. The unrooted tree shows the evolutionary relationships between the different HisF protein paralogs. Of particular interest is the position of the PriA branch, which suggests that this protein arose after diversification of HisA and TrpF. The tree was constructed from comparison of tertiary structures using the Fitch-Margoliash algorithm included in the PHYLIP package [52].

<https://doi.org/10.1371/journal.pone.0196349.g007>

PriA is a bisubstrate enzyme capable of catalyzing the isomeration, also via an Amadori rearrangement, of both ProFAR and PRAN into the corresponding aminoketoses. It was first described in *Streptomyces coelicolor* and *Mycobacterium tuberculosis* [60] and, as can be seen in the tree, its branch is projected from the HisA branch, suggesting that PriA evolved from HisA (Fig 7). These results confirm the antiquity of the HisA/HisF pair respect to PriA, and support previous conclusions of a close evolutionary relationship between PriA and HisA [61].

On the other hand, PurM is part of the PurM superfamily, and it is homologous to phosphoribosylformylglycinamide synthase [EC: 6.3.5.3] (PurL), the enzyme which catalyzes the previous step in the *de novo* synthesis of purines. Both enzymes share structural domains and catalyze biochemical equivalent processes using a very similar ATP-dependent reaction mechanism [43]. The available evidence suggest that PurL is the outcome of a gene duplication followed by a fusion event [62], confirming the role of gene duplication events in the early evolution of metabolic pathways [63].

The search for other structural homologues of PurM identified several proteins of the PurM superfamily, such as the Ni-Fe hydrogenase maturation protein (HypE), selenophosphate synthetase [EC: 2.7.9.3] (SelD), and thiamin phosphate kinase [EC: 2.7.4.16] (ThiL). The evolutionary relationship between all these enzymes is depicted in Fig 8.

Thus, the HisA/HisF pair are homologous proteins that catalyze similar reactions using the same protein scaffold. The same occurs in the PurL/PurM pair. However, the HisA/HisF and the PurL/PurM pairs are not homologous. In fact, it is well known that analogous catalytic functions can be present in two or more unrelated protein folds. In other words, similar or even identical catalytic functions can appear independently more than once during the evolutionary process, i.e. they are the outcome of convergent evolution [64,65]. As argued here, this is the case of the HisA/HisF and PurL/PurM pairs.

Active sites and reaction mechanisms of HisF and PurM

There are several types of convergent evolution, including enzymes with similar protein folds that catalyze the identical or similar reactions, as well as convergent evolution of different protein folds catalyzing chemically equivalent reactions on similar or identical substrates, i.e. functional convergent evolution [66].

One of the best-known examples of evolutionary functional convergence at a biochemical level is the case of the subtilisin-like and chymotrypsin-like families of serine proteases. These enzymes adopt different three-dimensional structures, but their active sites evolved independently and resulted in the same catalytic amino acids (the Ser-His-Asp triad), as well as a very similar catalytic mechanism for the hydrolysis of the peptide bond [67,68]. As described below, we have searched for this kind of convergence in the HisF and PurM enzymes.

Biochemical studies of HisF of *Thermotoga maritima* have shown that the essential residues for catalytic activity are Asp11 and Asp130 (Fig 4B). It has been proposed that the carboxylate groups of both aspartate residues are involved in general acid/base catalysis [69]. On the other hand, based on sequence conservation between several PurM enzymes, residues Asp94, His247, Thr249, Arg259 (Fig 4D) have been identified as key components of the active site for the PurM of *Escherichia coli*, where these residues may be involved in general acid and base catalysts, as well as in the coordination of metal ions [42]. As shown in Fig 6C, the use of the *align* command of PyMol to superimpose the active sites of HisF and PurM demonstrates that no spatial correspondence exists between them (Fig 6C). Our comparison reveals that the similarities between the two enzymes are limited to the structural correspondence between the Asp130 of HisF and the Asp94 of PurM and, to a certain extent, to the proximity between the Asp11 of HisF and Asp86 of PurM. However, the rest of the active site residues of PurM have no structural counterpart in the HisF structure. This implies that the architecture of the active sites of HisF and PurM is completely different, and thus it can be safely concluded that their biochemical activities are not due to the convergence of their catalytic sites.

Moreover, analysis of the reaction mechanisms proposed for each of the imidazole-synthesizing enzymes [42,69] demonstrates their profound differences, and allows to describe the specific details of the cyclization of the imidazole motif in each case. As noted above, the HisF

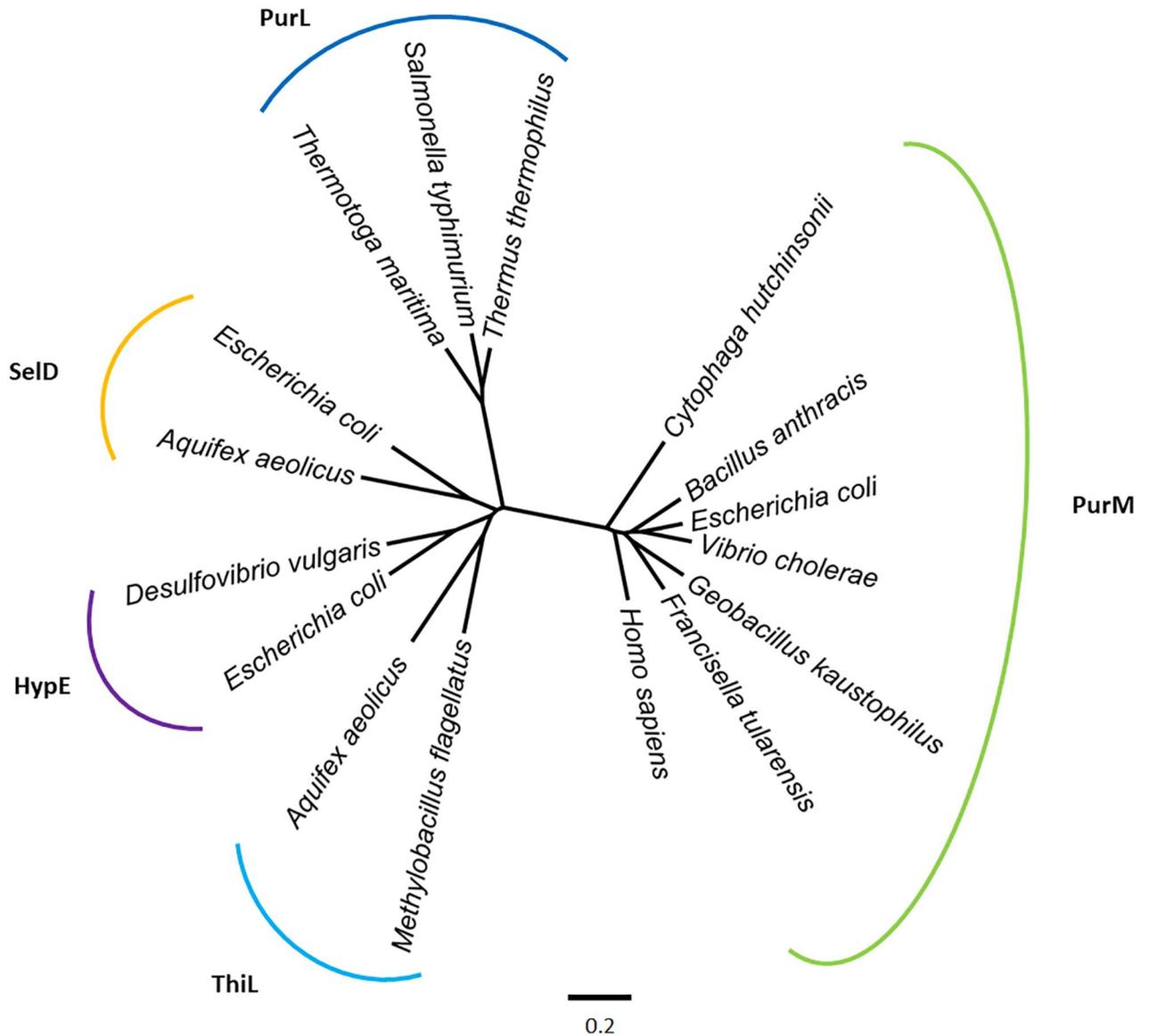


Fig 8. Structural phylogeny of PurM and its homologues. The unrooted tree shows the relationships among the different paralogs of the protein PurM. The tree was constructed from comparison of tertiary structures using the Fitch-Margoliash algorithm included in the PHYLIP package [52].

<https://doi.org/10.1371/journal.pone.0196349.g008>

enzyme catalyzes an Amadori rearrangement that is responsible for the formation of the imidazole ring, where the cyclization occurs after the hydrolysis of AICAR from the substrate PRFAR. This exposes a carbonyl group that can then react with the amino group of the same substrate (Fig 9). On the other hand, PurM uses one ATP molecule to activate a carbonyl group present in the FGAM structure, generating an iminophosphate intermediate. Immediately after this activation, an amino group makes a nucleophilic attack towards the previously-activated carbonyl group, leading to the cyclization of the imidazole ring. As depicted in Fig 9, the imidazole cyclization in both enzymes occurs by means of a nucleophilic attack of a nitrogen towards a carbonyl group. However, the overall mechanisms are completely different, with HisF catalyzing an Amadori rearrangement, and PurM catalyzing a strictly ATP-

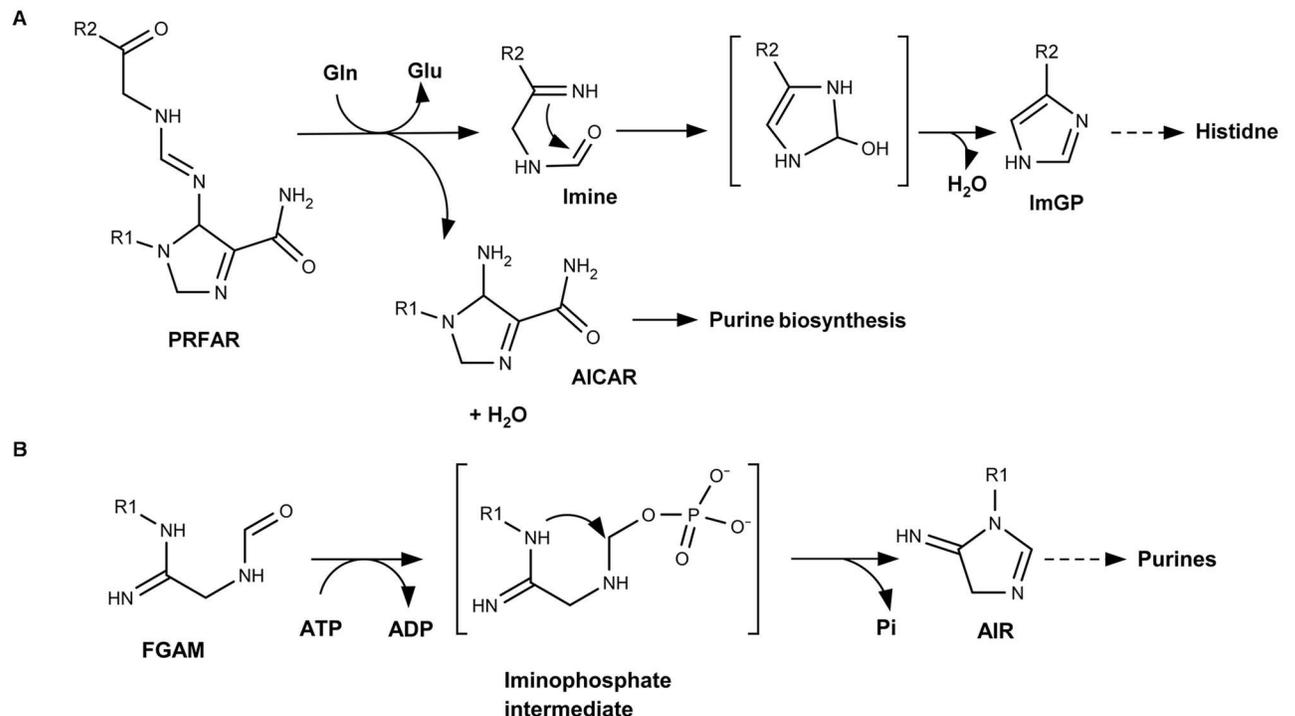


Fig 9. Proposed reaction schemes for HisF and PurM. The reaction schemes are based on [42, 69]. The nucleophilic attack of amino to the carbonyl group is depicted. R1: ribose phosphate; R2: triose phosphate.

<https://doi.org/10.1371/journal.pone.0196349.g009>

dependent reaction. In other words, the two enzymes are not homologous and the biosyntheses of the two imidazole groups proceed via different mechanisms.

Discussion

Evolutionary convergence in biochemistry

An example of biochemical redundancy is found in the biosyntheses of lysine, where three different anabolic pathways have been reported, i.e., the diaminopimelic acid pathway (DAP), the α -aminoadipic acid route (AAA) [70,71], and the recently-discovered TK0280-TK0283 variant found in *Thermococcus kodakarensis* [72].

Evolutionary functional convergence has also been described in the families of kinases that catalyze the phosphorylation of sugars, specifically in the hexokinase, ribokinase and galactokinase families, where two types of convergence can be recognized [66]. In the first case, a common trait originated independently several times in the same protein fold, such as the specificity to phosphorylate glucose, which clearly appeared more than once within the hexokinase family. In the second case, different protein structures convergently evolve to the same catalytic function, as occurred with the use of fructose as a substrate, which appeared independently in both the ribokinase and the hexokinase enzyme families [66]. The case of the imidazole-synthesizing HisF and PurM enzymes discussed here not only represents another example of this second kind of convergent evolution, but also underlines the role of redundancy in the biochemical syntheses of the imidazole group. As summarized below, this is best understood in terms of the extraordinary biological significance of this chemical moiety.

Evolution of two different imidazole-synthesizing enzymatic systems

As described above, histidine biosynthesis starts with ATP. This ribonucleotide is condensed with PRPP to form PRATP (Fig 3), which is then enzymatically processed to form AICAR. As underlined above, rather surprisingly the imidazole moiety in AICAR (which is derived from ATP) is not used to form histidine. Instead, a new imidazole moiety is biosynthesized and becomes part of histidine. The obvious question is why the available imidazole moiety already present in AICAR is not used for such task. As explained below, we suggest that the answer is found in the differences between the enhanced catalytic properties of the imidazole moiety of histidine as compared to that of purines.

Adenine has a well-known ability to coordinate a series of metal ions, including Cu^{2+} and Ag^+ [73]. It has been shown that artificially constructed adenine-containing matrices can coordinate such metals and exhibit catalytic activities, including hydrolysis of ester substrates such as *p*-nitrophenyl phosphate (*p*NPP), bis-(*p*-nitrophenyl) phosphate (bNPP) and 2-hydroxypropyl-*p*-nitrophenyl phosphate (hNPP), as well as the hydrolysis of 2'-3'-cAMP [73]. Metal binding is equally significant for the catalytic capabilities of RNA, together with the highly reactive ribose 2'-hydroxyl group and the tertiary structure that these molecules adopt. Although it has been demonstrated that the catalytic activity of the ribosome is dependent on an adenine base [74,75], the reaction mechanism involves a proton abstraction by the N3 of the adenine with no direct participation of the imidazole ring, which, as of today, has not been shown to participate in any of the known catalytic activities of RNA.

These results are explained by the fact that in both N6 ribosyl adenine and histidine the imidazole rings remain unbound, which leaves them free to react [76,77]. Nevertheless, this could change if the covalent bond that joins the ribose with the base is changed from position N9 to the position N6. In fact, as demonstrated by Maurel and Ninio [76], N6 ribosyl adenine can accelerate the hydrolysis of *p*-nitrophenyl acetate (PNPA) in solution at pH 7.7, a reaction that neither adenosine nor AMP can catalyze. The rate at which N6 ribosyl adenine accelerates the reaction is comparable to that observed when histidine is used as catalyst.

In sharp contrast with histidine, the β -glycosidic bond between the N9 of the imidazole ring in purine nucleotides and the C1' of the ribose blocks the otherwise acidic N9 atom, completely inhibiting the catalytic capabilities of its imidazole moiety. In other words, if a molecule such as N6-ribosyl adenine would be incorporated into RNA, we would expect that its imidazole ring would exhibit catalytic activities. This possibility is amenable to experimental analysis.

It is worth noting that in sharp contrast with pyrimidines, in the *de novo* synthesis of purines this β -N9-glycosidic bond is formed in the very first step of the anabolic pathway, and along the entire anabolic process the purine biosynthetic intermediates are ribotide derivatives, constraining the catalytic capabilities of the imidazole moiety from the start of the route.

We surmise that these structural changes in the imidazole moieties account for the major differences in the catalytic properties between histidine and adenine, and were probably the critical issue in the emergence of histidine biosynthesis, as well as an important selection pressure for the evolution of two non-homologous routes for the synthesis of imidazole group.

Evolution of histidine and purine biosynthesis

The phylogenetic distribution of the enzymes involved in histidine and purine *de novo* biosyntheses shows that both pathways are highly conserved through the three domains of life, suggesting that they originated and were well established before the diversification of the Last Universal Common Ancestor (LUCA) into major kingdoms [78]. Taking into consideration the functional convergence of the enzymes that catalyze imidazole syntheses, the two pathways

must have evolved independently during very early stages of protein evolution, using the same pentose derivative as precursor, and later on became connected through AICAR, which as underlined above can be considered as a mere intracellular salvage step. Given the key role of purines in RNA properties, including catalysis, it is reasonable to conclude that purine biosynthesis appeared before histidine anabolism.

Sequence analysis of the enzymes that catalyze the different steps in the pathways provide information of their evolution. Considering the number of paralogous genes that both HisF and PurM have within their respective routes (Fig 3), it is reasonable to hypothesize that the present enzymatic activities of these two proteins evolved from ancestral enzymes with less specificity for substrates. For instance, in addition to the PurL/PurM pair, the archeal *purD*, *purT*, *purK* and *purP* genes are also homologous [43]. This suggests the existence of at least two ancestral enzymes with relaxed substrate specificity, that could catalyze all the corresponding activities of the modern paralogous enzymes. This possibility is certainly consistent with the idea of a patchwork assembly of the routes [79]. In histidine biosynthesis, the experimental evidence indicating that the synthesis of IGP can take place, both *in vivo* and *in vitro*, in the absence of the HisH protein [40], as well as the likely existence of an ancestral enzyme with both HisA and HisF activities, suggest that primitive histidine biosynthesis was simpler in the early stages of its evolution, during which spontaneous transformations coexisted with enzymes with low substrate specificity catalyzing similar chemical reactions [41]. As summarized in Fig 3, a simpler histidine anabolism can thus be envisioned.

AICAR has an evolutionary adaptive value by its own

The manifold roles of AICAR in biochemical process go beyond the proposed salvage pathway connecting the histidine and the *de novo* synthesis of purines. For example, it has been shown that AICAR accumulates in *purH* mutants and inhibits the conversion of AIR to the 4-amino-5-hydroxymethyl-2-methylpyrimidine (HMP) moiety of thiamine [80], and that can also decrease total coenzyme A pools inhibiting the activity of the enzyme pantoate-alanine ligase in *Salmonella enterica* [81]. Perhaps more significantly, both AICAR and its triphosphorylated form (5-amino 4-imidazole carboxamide riboside 5'-triphosphate, also known as ZTP) can act by themselves as alarmones in C-1-folate deficiency [82]. AICAR accumulation also leads to histidine auxotrophy and even growth arrest in yeast [83]. AICAR is also the precursor for the antibiotic bredinin (5-hydroxyl-1-beta-D-ribofuranosyl-1H-imidazole-4-carboxamide), which is part of the naturally occurring family of nucleotide antibiotics first isolated from *Eupenicillium brefeldianum* M2166 [84]. This underlines the significance of AICAR, which during evolution has acquired other functions beyond that of an intermediate in purine biosynthesis (Fig 10). This may be seen as a case of exaptation, in which evolution selected AICAR into other roles, with that of alarmone being among the earliest. The fact that both AICAR and ZTP can bind to functional RNA molecules [85], and that 4-amino-5-imidazole carboxamide (AICA) can be synthesized under prebiotic conditions [86,87] may also speak of the antiquity of this molecule.

Other imidazole syntheses in the cell

In addition to the histidine and purines pathways there are other imidazole-synthesizing routes present in the extant cells, which also underline the key role of imidazole and imidazole-derivatives. All the biological and abiotic forming mechanisms of imidazolides represent a case of redundancy for the synthesis of this chemical group and highlight its catalytic significance.

the most abundant amino acids found in prebiotic experiments [91,92] as well as in carbonaceous chondrites [93] a possible model prebiotic experiment using the Ala-Ser-Gly in order to obtain an imidazolide has recently been proposed [17].

Conclusions

In contrast to other cases of convergence in biosynthetic pathways such as those reported for polyphyletic caffeine biosyntheses [65], in the different routes for imidazole formation discussed here the final products are not the outcome of enzyme recruitment via a patchwork mechanism, but rather examples of deterministic chemical processes that lead via different processes to the same functional product.

Like all other biological traits, the properties of biochemical pathways have been shaped by their evolutionary history. History, however, is not a mere sequence of chance events. Natural selection can overcome contingent effects to an extent, and chemical and biochemical constraints may have had a greater significance than is normally recognized in shaping metabolic traits. This appears to be the case of the manifold ways in which imidazole moieties are formed in extant cells.

As reviewed here, analysis of metabolic processes and biochemical properties demonstrate that several independent mechanisms of imidazole biosyntheses have evolved. The significance of this chemical group for living organisms is evident as their metabolism became redundant in respect to their syntheses, although the specific reaction mechanisms of each pathway are clearly different. In the purine and histidine biosynthesis, HisF and PurM have different sequences and unrelated three-dimensional structures; however, the two enzymes catalyze analogous reactions that lead to the formation of the imidazole group, representing a case of convergence at a biochemical level. This speaks for the catalytical and structural significance of imidazole and its derivatives.

The two-different enzyme-dependent syntheses of histidine and purines, catalyzed by PurM and HisF, together with the autocatalytic formation of creatinine (from creatine-phosphate) and MIO (from the Ala-Ser-Gly tripeptide) are unique examples of biochemical redundancy that underline for the significance of imidazole derivatives in biological processes. As argued here, the polyphyletic origin of PurM and the HisF is best understood by the enhanced catalytic activity of the histidine imidazole group compared to that of purines, where the β -glycosidic bond between the ribose and the imidazole ring of purines is established from the very start of IMP biosynthesis and results in the inhibition of purine catalytic activities.

The formation of the imidazole motifs in purines and histidine proceeds by means of a nucleophilic attack of a nitrogen atom in an amino group to a carbon atom in the substrate. However, as detailed here, the corresponding imidazole-synthesizing enzymes have different amino acid sequences and different three-dimensional folds, but catalyze chemically equivalent reactions. Thus, in spite of chemical restrictions, considerable degrees of freedom in the sequence space were available during the early stages of metabolic evolution.

Supporting information

S1 Fig. Pairwise alignment of bacterial (*Escherichia coli*) HisF and PurM sequences. The corresponding HisF and PurM sequences of *E. coli* were aligned using the Needleman-Wunsch algorithm as implemented in the EMBOSS Needle program of the European Molecular Biology Open Software Suite [45]. The alignment shows 58.5% of gaps with only 12.2% of sequence identity and 21.4% of sequence similarity.

(TIF)

S1 Table. Crystallographic structures used for the analysis. Complete list of the crystallographic structures and their PDB codes used to construct the dendograms. (DOCX)

Acknowledgments

Alberto Vázquez-Salazar is a doctoral student from Programa de Doctorado en Ciencias Biomédicas, Universidad Nacional Autónoma de México (UNAM) and received fellowship 404507 from CONACYT. Financial support from PAPIIT-UNAM (IN223916) is gratefully acknowledged. We are grateful to Ada Yonath, Marie-Christine Maurel, and Renato Fani for many conversations over the years on the origin of histidine biosynthesis, and thank Ricardo Hernandez-Morales for help with the manuscript.

Author Contributions

Conceptualization: Alberto Vázquez-Salazar, Arturo Becerra, Antonio Lazcano.

Investigation: Alberto Vázquez-Salazar, Arturo Becerra, Antonio Lazcano.

Writing – review & editing: Alberto Vázquez-Salazar, Arturo Becerra, Antonio Lazcano.

References

1. Sundberg RJ, Martin RB. Interactions of histidine and other imidazole derivatives with transition metal ions in chemical and biological system. *Chem Rev.* 1974; 74, 4:471–517.
2. Edgcomb SP, Murphy KP. Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins.* 2002; 1; 49(1):1–6. <https://doi.org/10.1002/prot.10177> PMID: 12211010
3. Grimsley GR, Scholtz JM, Pace CN. A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sci.* 2009; 18(1):247–251. <https://doi.org/10.1002/pro.19> PMID: 19177368
4. Holliday GL, Fischer JD, Mitchell JB, Thornton JM. Characterizing the complexity of enzymes on the basis of their mechanisms and structures with a bio-computational analysis. *FEBS J.* 2011; 278(20):3835–45. <https://doi.org/10.1111/j.1742-4658.2011.08190.x> PMID: 21605342
5. Maley LE, Mellor DP. Stability of some metal complexes of histidine. *Nature.* 1950; 165: 4194: 453.
6. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. Metal ions in biological catalysis: from enzyme databases to general principles. *J Biol Inorg Chem.* 2008; 13(8):1205–18. <https://doi.org/10.1007/s00775-008-0404-5> PMID: 18604568
7. Oró J. Synthesis of adenine from ammonium cyanide. *Biochem Biophys Res Commun.* 1960; 2:407–412.
8. Oró J. Synthesis of organic compounds by electric discharges. *Nature.* 1963; 197:862–67.
9. Lowe CU, Ress A, Markham FRS. Synthesis of complex organic compounds from simple precursors: Formation of amino-acids, amino-acid polymer, fatty acids and purines from ammonium cyanide. *Nature.* 1963; 19:219–222.
10. Schwartz AW, Bakker CG. Was adenine the first purine? *Science.* 1989; 245:1102–1104. PMID: 11538344
11. Borquez E, Cleaves HJ, Lazcano A, Miller SL. An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers. *Orig Life Evol Biosph.* 2005; 35(2):79–90. PMID: 16010991
12. Oró J, Basile B, Cortes S, Shen C, Yamrom T. The prebiotic synthesis and catalytic role of imidazoles and other condensing agents. *Orig Life.* 1984; 14(1–4):237–42. PMID: 6462668
13. Shen C, Yang L, Miller SL, Oró J. Prebiotic syntheses of imidazole-4-acetaldehyde and histidine. *Orig Life.* 1987 17: 295–305.
14. Shen C, Yang L, Miller SL, Oró J. Prebiotic synthesis of histidine. *J Mol Evol.* 1990; 31:167–174. PMID: 11536478
15. Decker P, Schweer H, Pohlmann R. Bioids. X. Identification of formose sugars, presumable prebiotic metabolites using capillary gas chromatography/gas chromatography-mass spectrometry of *n*-butoxime trifluoroacetates on OV-225. *Journal of Chromatography.* 1982; 244:281–291.

16. Bada JL, Chalmers JH, Cleaves HJ. Is formamide a geochemically plausible prebiotic solvent? *Phys Chem Chem Phys*. 2016; 18:20085–90. <https://doi.org/10.1039/c6cp03290g> PMID: 27253848
17. Vázquez-Salazar A, Tan G, Stockton A, Fani R, Becerra A, Lazcano A. Can an imidazole be formed from an alanyl-seryl-glycine tripeptide under possible prebiotic conditions? *Orig Life Evol Biosph*. 2017; 47:345–354. <https://doi.org/10.1007/s11084-016-9525-y> PMID: 27771860
18. Álvarez-Carreño C, Becerra A, Lazcano A. Norvaline and norleucine may have been more abundant protein components during early stages of cell evolution. *Orig Life Evol Biosph*. 2013; 43:363–75. <https://doi.org/10.1007/s11084-013-9344-3> PMID: 24013929
19. Raggi L, Bada JL, Lazcano A. On the lack of evolutionary continuity between prebiotic peptides and extant enzymes. *Phys Chem Chem Phys*. 2016; 18:20028–20032. <https://doi.org/10.1039/c6cp00793g> PMID: 27121024
20. Kruger K, Grabowski PJ, Zaugg AJ, Sands J, Gottschling DE, Cech TR. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*. 1982; 31:147–157. PMID: 6297745
21. Guerrier-Takada C, Gardiner K, Marsh T, Pace T, Altman S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*. 1983; 35:849–857. PMID: 6197186
22. Walton T, Szostak JW. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J Am Chem Soc*. 2016; 138(36):11996–20022. <https://doi.org/10.1021/jacs.6b07977> PMID: 27552367
23. Breslow R. Bifunctional acid-base catalysis by imidazole groups in enzyme mimics. *J Mol Cat*. 1994; 91(2): 161–174.
24. Perrotta AT, Shih I, Been MD. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science*. 1999; 286(5437):123–6. PMID: 10506560
25. Shen C, Lazcano A, Oró J. The enhancement activities of histidyl-histidine in some prebiotic reactions. *J Mol Evol*. 1990; 31:445–52. PMID: 11540924
26. Wieczorek R, Dörr M, Chotera A, Luisi PL, Monnard PA. Formation of RNA phosphodiester bond by histidine-containing dipeptides. *Chembiochem*. 2013; 14(2):217–23. <https://doi.org/10.1002/cbic.201200643> PMID: 23255284
27. Wieczorek R, Adamala K, Gasperi T, Polticelli F, Stano P. Small and random peptides: an unexplored reservoir of potentially functional primitive organocatalysts. The case of seryl-histidine. *Life (Basel)*. 2017; 7(2). pii: E19.
28. Gorlero M, Wieczorek R, Adamala K, Giorgi A, Schininà ME, Stano P, et al. Ser-His catalyses the formation of peptides and PNAs. *FEBS Lett*. 2009; 583:153–6. <https://doi.org/10.1016/j.febslet.2008.11.052> PMID: 19071124
29. Adamala K, Anella F, Wieczorek R, Stano P, Chiarabelli C, Luisi PL. Open questions in origin of life: experimental studies on the origin of nucleic acids and proteins with specific and functional sequences by a chemical synthetic biology approach. *Comput Struct Biotechnol J*. 2014; 9:e201402004. <https://doi.org/10.5936/csbj.201402004> PMID: 24757502
30. Petrov AS, Gulen B, Norris AM, Kovacs NA, Bernier CR, Lanier KA, et al. History of the ribosome and the origin of translation. *Proc Natl Acad Sci U S A*. 2015; 112(50):15396–401. <https://doi.org/10.1073/pnas.1509761112> PMID: 26621738
31. Kovacs NA, Petrov AS, Lanier KA, Williams LD. Frozen in time: the history of proteins. *Mol Biol Evol*. 2017; 34(5):1252–1260. <https://doi.org/10.1093/molbev/msx086> PMID: 28201543
32. White HB. Coenzymes as fossils of an earlier metabolic state. *J Mol Evol*. 1976; 7:101–104. PMID: 1263263
33. White HB. Evolution of coenzymes and the origin of pyridine nucleotides. In: Everse J, Anderson B, You KS, editors. *The pyridine nucleotide coenzymes*. New York: Academic Press; 1982. pp. 1–17.
34. Akouche M, Jaber M, Maurel MC, Lambert JF, Georgelin T. Phosphoribosyl pyrophosphate: a molecular vestige of the origin of life on minerals. *Angew Chem Int Ed Engl*. 2017; 56(27):7920–7923. <https://doi.org/10.1002/anie.201702633> PMID: 28558156
35. Alifano P, Fani R, Liu P, Lazcano A, Bazzicalupo M, Carlomagno MS, et al. Histidine biosynthetic pathway and genes: structure, regulation, and evolution. *Microbiol Rev*. 1996; 60:44–69. PMID: 8852895
36. Chaudhuri BN, Lange SC, Myers RS, Chittur SV, Davisson VJ, Smith JL. Crystal structure of imidazole glycerol phosphate synthase: a tunnel through a (beta/alpha)₈ barrel joins two active sites. *Structure*. 2001; (10):987–97. PMID: 11591353
37. Chaudhuri BN, Lange SC, Myers RS, Davisson VJ, Smith JL. Toward understanding the mechanism of the complex cyclization reaction catalyzed by imidazole glycerol phosphate synthase: crystal structures of a ternary complex and the free enzyme. *Biochemistry*. 2003; 42(23):7003–12. <https://doi.org/10.1021/bi034320h> PMID: 12795595

38. Rivalta I, Sultan MM, Lee NS, Manley GA, Loria JP, Batista VS. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc Natl Acad Sci U S A.*, 2012; 109 (22): E1428–E1436. <https://doi.org/10.1073/pnas.1120536109> PMID: 22586084
39. Smith DW, Ames BN. Intermediates in the early steps of histidine biosynthesis. *J Biol Chem.* 1964; 239:1848–1855. PMID: 14213364
40. Rieder G, Merrick MJ, Castorph H, Kleiner D. Function of *hisF* and *hisH* gene products in histidine biosynthesis. *J Biol Chem.* 1994; 269 (20):14386–14390. PMID: 8182043
41. Lazcano A, Miller SL. On the origin of metabolic pathways. *J Mol Evol.* 1999; 49(4): 424–31 PMID: 10486000
42. Li C, Kappock TJ, Stubbe J, Weaver TM, Ealick SE. X-ray crystal structure of aminoimidazole ribonucleotide synthetase (PurM), from the *Escherichia coli* purine biosynthetic pathway at 2.5 Å resolution. *Structure.* 1999; 7:1155–1166. PMID: 10508786
43. Zhang Y, Morar M, Ealick SE. Structural biology of the purine biosynthetic pathway. *Cell Mol Life Sci.* 2008; 65 (23):3699–724. <https://doi.org/10.1007/s00018-008-8295-8> PMID: 18712276
44. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. 2011; *Nucleic Acids Res.* 40(D1): D109–D114.
45. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. 2015; 43(W1):W580–W584.
46. The PyMOL Molecular Graphics System, Version 1.7 Schrödinger, LLC.
47. Prlic A, Bliven S, Rose PW, Bluhm WF, Bizon C, Godzik A, et al. Pre-calculated Protein Structure Alignments at the RCSB PDB website Bioinformatics. 2010; 26(23): 2983–2985. <https://doi.org/10.1093/bioinformatics/btq572> PMID: 20937596
48. Holm L, Rosenström P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* 2010; 38(Web Server issue): W545–W549. <https://doi.org/10.1093/nar/gkq366> PMID: 20457744
49. Jácome R, Becerra A, Ponce de León S, Lazcano A. Structural analysis of monomeric RNA-dependent polymerases: evolutionary and therapeutic implications. *PLoS One.* 2015; 10(9):e0139001. <https://doi.org/10.1371/journal.pone.0139001> PMID: 26397100
50. Krissinel E, Henrick K. Secondary-structure matching (PDBFold), a new tool for fast protein structure alignment in three dimensions. *Acta Cryst.* 2004; D60, 2256–2268.
51. Subbiah S, Laurents DV, Levitt M. Structural similarity of DNA-binding domains of bacteriophage repressors and the globin core. *Curr Biol.* 1993; 3:141–148. PMID: 15335781
52. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle
53. FigTree v1.4.2. <http://tree.bio.ed.ac.uk/software/figtree/>
54. Fani R, Liò P, Chiarelli I, Bazzicalupo M. The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the *hisA* and *hisF* genes. *J Mol Evol.* 1994; 5:489–95.
55. Lang D, Thoma R, Henn-Sax M, Sterner R, Wilmanns M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science.* 2000; 289(5484):1546–1550. PMID: 10968789
56. Henn-Sax M, Höcker B, Wilmanns M, Sterner R. Divergent evolution of (beta/alpha)₈-barrel enzymes. *Biol Chem.* 2001; 382(9):1315–1320. <https://doi.org/10.1515/BC.2001.163> PMID: 11688714
57. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, Sterner R. Two (beta/alpha)₈-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry.* 2002; 41(40):12032–12042. PMID: 12356303
58. Friedmann N, Haverland WJ, Miller SL. Prebiotic synthesis of the aromatic and other amino acids. In: Buvet R, Ponnamperna C, editors. *Chemical evolution and the origin of life: Proceedings of the international conference on the origin of life.* Amsterdam: North-Holland Publishing Company; 1971. pp 123–135.
59. Jukes TH. *Molecules and evolution.* New York: Columbia University Press; 1966.
60. Barona-Gómez F, Hodgson DA. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* 2003; 4(3):296–300. <https://doi.org/10.1038/sj.embor.embor771> PMID: 12634849
61. Due AV, Kuper J, Geerlof A, von Kries JP, Wilmanns M. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc Natl Acad Sci U S A.* 2011; 108(9):3554–3559. <https://doi.org/10.1073/pnas.1015996108> PMID: 21321225
62. Suzuki S, Yanai H, Kanagawa M, Tamura S, Watanabe Y, Fuse K, et al. Structure of N-formylglycinamide ribonucleotide amidotransferase II (PurL) from *Thermus thermophilus* HB8. *Acta Crystallogr Sect*

- F Struct Biol Cryst Commun. 2012; 68(Pt 1):14–9. <https://doi.org/10.1107/S1744309111048184> PMID: 22232163
63. Becerra A, Lazcano A. The role of gene duplication in the evolution of purine nucleotide salvage pathways. *Orig. Life Evol. Biosph.* 1998; 28:539–553. PMID: 9742728
 64. Todd AE, Orengo CA, Thornton JM. Evolution of protein function, from a structural perspective. *Curr Opin Chem Biol.* 1999 3:548–556. PMID: 10508675
 65. Huang R, O'Donnell AJ, Barboline JJ, Barkman TJ. Convergent evolution of caffeine in plants by co-option of exapted ancestral enzymes. *Proc Natl Acad Sci U S A.* 2016; 113(38):10613–10618. <https://doi.org/10.1073/pnas.1602575113> PMID: 27638206
 66. Bork P, Sander C, Valencia A. Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases. *Protein Sci.* 1993; 2: 31–40. <https://doi.org/10.1002/pro.5560020104> PMID: 8382990
 67. Reardon D, Farber GK. The structure and evolution of alpha/beta barrel proteins. *FASEB J.* 1995; 7: 497–503.
 68. Frey PA, Hegeman AD. *Enzymatic reaction mechanisms.* New York: Oxford University Press; 2007.
 69. Beismann-Driemeyer S, Sterner R. Imidazole glycerol phosphate synthase from *Thermotoga maritima*. Quaternary structure, steady-state kinetics, and reaction mechanism of the holoenzyme complex. *J Biol Chem.* 2001; 276(23): 20387–20396. <https://doi.org/10.1074/jbc.M102012200> PMID: 11264293
 70. Kobashi N, Nishiyama M, Tanokura M. Aspartate kinase-independent lysine synthesis in an extremely thermophilic bacterium, *Thermus thermophilus*: lysine is synthesized via alpha-aminoadipic acid not via diaminopimelic acid. *J Bacteriol.* 1999; 181(6):1713–1718. PMID: 10074061
 71. Velasco AM, Leguina JI, Lazcano A. Molecular evolution of the lysine biosynthetic pathways. *J Mol Evol.* 2002; 55(4):445–459. <https://doi.org/10.1007/s00239-002-2340-2> PMID: 12355264
 72. Yoshida A, Tomita T, Atomi H, Kuzuyama T, Nishiyama M. Lysine biosynthesis of *Thermococcus kodakarensis* with the capacity to function as an ornithine biosynthetic system. *J Biol Chem.* 2016; 291(41): 21630–21643. <https://doi.org/10.1074/jbc.M116.743021> PMID: 27566549
 73. Verma S, Mishra AK, Kumar J. The many facets of adenine: coordination, crystal patterns, and catalysis. 2010; *Acc Chem Res.* 43(1): 79–91. <https://doi.org/10.1021/ar9001334> PMID: 19719100
 74. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA. The structural basis of ribosome activity in peptide bond synthesis. *Science.* 2000; 289(5481): 920–930. PMID: 10937990
 75. Yonath A. Ribosomal crystallography: peptide bond formation, chaperone assistance and antibiotics activity. *Mol Cells.* 2005; 20(1): 1–16. PMID: 16258236
 76. Maurel MC, Ninio J. Catalysis by a prebiotic nucleotide analog of histidine. *Biochimie.* 1987; 69:551–553. PMID: 2445388
 77. Décout JL, Maurel MC. N6-substituted adenine derivatives and RNA primitive catalysts. *Orig Life Evol Biosph.* 1993; 23: 299–306. PMID: 7509474
 78. Delaye L, Becerra A, Lazcano A. The last common ancestor: what's in a name? *Orig Life Evol Biosph.* 2005; 35(6): 537–554. <https://doi.org/10.1007/s11084-005-5760-3> PMID: 16254691
 79. Jensen RA. Enzyme recruitment in evolution of new function. *Ann. Rev. Microbiol.* 1976; 30: 409–425.
 80. Allen S, Zilles JL, Downs DM. Metabolic flux in both the purine mononucleotide and histidine biosynthetic pathways can influence synthesis of the hydroxymethyl pyrimidine moiety of thiamine in *Salmonella enterica*. *J Bacteriol.* 2002; 184(22): 6130–6137. <https://doi.org/10.1128/JB.184.22.6130-6137.2002> PMID: 12399482
 81. Bazarro JV, Downs DM. Amino-4-imidazolecarboxamide ribotide directly inhibits coenzyme A biosynthesis in *Salmonella enterica*. *J Bacteriol.* 2014; 196(4): 772–779. <https://doi.org/10.1128/JB.01087-13> PMID: 24296672
 82. Bochner BR, Ames BN. ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. *Cell.* 1982; 29(3): 929–937. PMID: 6185232
 83. Daignan-Fornier B, Pinson B. 5-Aminoimidazole-4-carboxamide-1-beta-D-ribofuranosyl 5'-monophosphate (AICAR), a highly conserved purine intermediate with multiple effects. *Metabolites.* 2012; 2(2): 292–302. <https://doi.org/10.3390/metabo2020292> PMID: 24957512
 84. Suhadolnik RJ. Naturally occurring nucleoside and nucleotide antibiotics. *Prog Nucleic Acid Res Mol Biol.* 1979; 22: 193–291. PMID: 230535
 85. Jones CP, Ferré-D'Amaré AR. Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Nat Struct Mol Biol.* 2015; (9): 679–685. <https://doi.org/10.1038/nsmb.3073> PMID: 26280533
 86. Oró J, Kimball AP. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch Biochem Biophys.* 1961; 94: 217–227. PMID: 13731263

87. Joyce GF. RNA evolution and the origins of life. *Nature*. 1989; 338: 217–224. <https://doi.org/10.1038/338217a0> PMID: 2466202
88. Borsook H, Dubnoff JW. The hydrolysis of phosphocreatine and the origin of urinary creatinine. *J Biol Chem*. 1947; 168(2): 493–510. PMID: 20238607
89. Cooke HA, Chistianson CV, Bruner SD. Structure and chemistry of 4-methylideneimidazole-5-one containing enzymes. *Curr Opin Chem Biol*. 2009; 4: 460–468.
90. Rétey J. Discovery and role of methylidene imidazolone, a highly electrophilic prosthetic group. *Biochim Biophys Acta*. 2003; 1647: 179–184. PMID: 12686130
91. Miller SL. A production of amino acids under possible primitive earth conditions. *Science*. 1953; 117: 528–529. PMID: 13056598
92. Cleaves HJ, Chalmers JH, Lazcano A, Miller SL, Bada JL. A reassessment of prebiotic organic synthesis in neutral planetary atmospheres. *Orig Life Evol Biosph*. 2008; 38(2): 105–115. <https://doi.org/10.1007/s11084-007-9120-3> PMID: 18204914
93. Burton AS, Stern JC, Elsila JE, Glavin DP, Dworkin JP. Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites. *Chem Soc Rev*. 2012; 41: 5459–5472. <https://doi.org/10.1039/c2cs35109a> PMID: 22706603
94. MarvinSketch v16.1.4 ChemAxon. <https://chemaxon.com/products/marvin-js>