



# Article The Role of Gene Elongation in the Evolution of Histidine Biosynthetic Genes

### Sara Del Duca, Sofia Chioccioli, Alberto Vassallo<sup>D</sup>, Lara Mitia Castronovo and Renato Fani \*

Department of Biology, University of Florence, 50019 Sesto Fiorentino, Italy; sara.delduca@unifi.it (S.D.D.); sofia.chioccioli@unifi.it (S.C.); alberto.vassallo@unifi.it (A.V.); laramitia.castronovo@unifi.it (L.M.C.)

\* Correspondence: renato.fani@unifi.it

Received: 7 April 2020; Accepted: 11 May 2020; Published: 13 May 2020



**Abstract:** Gene elongation is a molecular mechanism consisting of an in-tandem duplication of a gene and divergence and fusion of the two copies, resulting in a gene constituted by two divergent paralogous modules. The aim of this work was to evaluate the importance of gene elongation in the evolution of histidine biosynthetic genes and to propose a possible evolutionary model for some of them. Concerning the genes *hisA* and *hisF*, which code for two homologous  $(\beta/\alpha)_8$ -barrels, it has been proposed that the two extant genes could be the result of a cascade of gene elongation/domain shuffling events starting from an ancestor gene coding for just one  $(\beta/\alpha)$  module. A gene elongation event has also been proposed for the evolution of *hisB* and *hisD*; structural analyses revealed the possibility of an early elongation event, resulting in the repetition of modules. Furthermore, it is quite possible that the gene elongations responsible for the evolution of the four proteins occurred before the earliest phylogenetic divergence. In conclusion, gene elongation events seem to have played a crucial role in the evolution of the histidine biosynthetic pathway, and they may have shaped the structures of many genes during the first steps of their evolution.

Keywords: gene elongation; histidine biosynthesis; patchwork hypothesis

### 1. Introduction

In the study of molecular evolution, because there are many missing links between the past and the present, the investigation of extant structures is fundamental to attempts to infer the ancestral situation [1]. The ever-increasing availability of complete genome nucleotide sequences from different organisms provides a huge dataset for the study of the structure and organization of genes and genomes; this makes possible a better understanding of the molecular mechanisms responsible for the evolution of metabolic pathways [2]. It is generally accepted that ancestral protein-coding genes were probably short sequences, encoding little polypeptides presumably corresponding to functional and/or structural domains. The evolution of genes, comprising the increase of their size and complexity, is the result of different molecular mechanisms including (i) gene duplication, (ii) gene fusion, and (iii) gene elongation [3].

### 1.1. Gene Duplication and Gene Fusion Events

The duplication of a gene can provide benefits to the organism simply because extra amounts of protein or RNA products are provided; however, one of the most important outcomes of gene duplication is the emergence of novel functions [4]. The duplication of genes and their subsequent functional divergence can lead to the formation of paralogous genes families that are evolutionarily related but have different functions [5]. In fact, after a gene duplicates, one of the two copies becomes dispensable and can undergo (several types of) mutational events. Over time, the mutation-containing gene can become a (pseudo)gene that can either be deleted from the genome or acquire a novel function

with respect to the parental gene [4]. Hence, gene duplication is one of the principal mechanisms involved in protein evolution [6].

In addition to gene duplication, one of the major forces driving gene evolution is the fusion of independent cistrons leading to bi- or multifunctional proteins [3]. Gene fusions allow the physical association of different protein domains (catalytic or regulatory) and can occur between genes located near to each other or far apart on the DNA molecule(s). Fusions frequently involve genes coding for proteins that function in a concerted manner, such as successive enzymes in metabolic pathways, enzymes and their regulatory domains, or DNA- and ligand-binding domains in prokaryotic transcriptional regulators [7].

#### 1.2. The Gene Elongation Mechanism

Gene elongation consists of an in-tandem gene duplication, which produces two (or more) copies of the same gene, followed by the loss of the intergenic region and the mutation of the stop codon of the first copy into a sense codon, resulting in the elongation, by fusion, of the initial gene and its copy. Thus, gene elongation is actually the combination of gene duplication and gene fusion. The newly formed gene is constituted by two paralogous modules, which might independently undergo different mutations and further duplications [3] that, over time, can cover up the traces of these early events. As shown in Figure 1, once the two copies have originated, in principle they might follow two different evolutionary pathways in which they can either immediately fuse and then diverge or, vice versa, diverge and then fuse. The final outcome of these two alternative pathways is the same: the formation of a gene constituted by two divergent paralogous modules. Gene elongation produces two or more copies of the same protein fold within a single polypeptide chain, often generating a direct repeat of domains and thus a pseudo-symmetrical structure with internal symmetry axes [1]. Many present-day proteins show internal repeats of amino acid sequences, which often correspond to functional or structural domains; since gene elongation has occurred in so many cases, this event must be considered an evolutionary advantage. The biological significance of proteins with repetitive structures might include (i) the improvement of the protein function by increasing the number of active sites; (ii) the acquisition of an additional function by modifying a redundant segment [3], thus obtaining a bifunctional enzyme; and/or (iii) the stabilization of a protein structure, thus increasing the enzyme's catalytic activity. Lots of cases of genes with internal sequence repetitions are reported in the literature. For example, in the bacterial ferredoxin, the second half of the amino acid sequence is an almost exact duplicate of the first [8]. Tang et al. [9] observed that the pepsin family of proteases have an intramolecular two-fold symmetry axis that relates two topologically similar domains, and proposed a mechanism for its evolution by gene elongation. The *carB* gene of *Escherichia coli*, which encodes a subunit of carbamoyl-phosphate synthetase, was proposed to be formed by the duplication of an ancestral gene, since its amino acid sequence shows a highly significant homology between the amino- and carboxyl-terminal halves of the protein [10]. Rubin et al. [11] found that the two halves of Gram-negative bacterial tetracycline efflux pumps share a process of tandem gene duplication and divergence. Gupta and Singh [12] suggested a model for the evolution of the heat-shock protein 70 (Hsp70) of Archaea and Bacteria based on gene duplication. Moreover, domain fusions have occurred in the evolution of the ATP binding cassette (ABC) superfamily [13]. However, the most documented case of gene elongation involves the histidine biosynthesis genes *hisA* and *hisF*; they are paralogous and may have originated from the duplication (and the subsequent divergence and fusion of the two resulting copies) of an ancestral gene, half the size of the current genes [14]. To the best of our knowledge, this is the only reported case of a "universal" gene elongation event; indeed, *hisA* and hisF genes share the same internal organization in all histidine-synthesizing organisms, suggesting the antiquity of this elongation event and that it occurred (long) before the appearance of the Last Universal Common Ancestor (LUCA) [15]. Thus, gene elongation might have shaped the structures of many genes during the first steps of molecular and cellular evolution.



**Figure 1.** The gene elongation mechanism. The two possible evolutionary routes are depicted. Genes are represented with arrows, and the encoded proteins with rounded rectangles. Representation adapted from Fani and Fondi [3].

### 1.3. Histidine Biosynthesis and Its Evolution

One of the most studied metabolic pathways, which shows a plethora of gene structures and organizations, is histidine biosynthesis, an ancient metabolic pathway present in Bacteria, Archaea, lower eukaryotes, and plants [16] that converts 5-phosphoribosyl-1-pyrophosphate (PRPP) to L-histidine through 10 enzymatic reactions (Figure 2). L-histidine biosynthesis plays an important role in cellular metabolism, being interconnected with both the de novo synthesis of purines and to nitrogen metabolism, which is why it is defined as a "metabolic cross-roads" [15]. It is one of the best characterized pathways from different biochemical, genetic, and evolutionary viewpoints. Its study is interesting because of (i) the presence of several quite uncommon reactions for a biosynthetic pathway, (ii) the existence of connections with other metabolic pathways, (iii) the structural features of several of the biosynthetic enzymes [17], and (iv) the different organization of *his* genes in different (micro)organisms. Genetic studies of this pathway have contributed, over the years, to the formulation of key concepts of molecular biology, like the operon hypothesis [18], and to the understanding of the mechanisms underlying the regulation of biosynthetic pathways, such as feedback inhibition, energy charge, and the setting of basal biosynthetic enzyme levels [19].

P

de novo synthesis of purines





соон L-histidine HisD

NADH + H<sup>+</sup> NAD<sup>+</sup>

NH:

L-histidinal

нс́ — о

This pathway has been extensively studied mainly in the enterobacteria *E. coli* and *Salmonella enterica*, for which a lot of information is available, including pathway structure, gene expression, and growing volumes of sequence data [15,19]. In these two bacteria the route is identical, consisting of nine intermediates

and eight enzymes encoded, respectively, by eight genes arranged in a compact operon (*hisGDC[NB]HAF[IE]*) (Figure 3): three of these (i.e., *hisNB*, *hisD*, and *hisIE*) code for bifunctional enzymes, while another one is heterodimeric, being composed of the *hisH* and *hisF* gene products [17].



Figure 3. Schematic representation of Escherichia coli his operon.

Chemical and biological evidence suggests that histidine was already present on Earth during the long period of chemical abiotic synthesis of organic compounds. It is known that the imidazole group of the lateral chain of His is present in the active sites of many enzymes, playing an important role in metabolism. Indeed, the dipeptide His-His was one of the first simple peptides to show the ability to form peptide bonds under plausible prebiotic conditions. It is reasonable to speculate that small peptides containing histidine might have been involved in the prebiotic formation of other peptides and nucleic acid molecules [20,21]. Since histidine is the only amino acid of which biosynthesis starts from a nucleotide rather than small metabolic intermediates, it has also been suggested that histidine may be the molecular vestige of a catalytic ribonucleotide from an earlier biochemical era in which RNA played a major role in catalysis [22,23]. If primitive enzymes required histidine, then the exhaustion of the prebiotic supply of histidine imposed a selective pressure favoring those organisms capable of synthesizing this molecule. The need to produce histidine suggests that this biosynthetic route is ancient and might have been assembled long before the appearance of LUCA [3,21]. Comparison of the *his* genes available in databases shows that after the divergence from LUCA, the structure, organization, and order of these genes underwent substantial rearrangements in the three cell lineages (Archaea, Bacteria, and Eukarya) [17]; indeed, a wide variety of different clustering strategies of his genes have been documented [3]. Although it is not possible from analysis of the available data to infer the organization and localization on the genome of the *his* genes in the last common ancestor, the same analysis may help to understand the primitive structures of some of them and their evolution [17]. Nowadays, this reconstruction of evolutionary dynamics is facilitated by the increasing number of complete genome sequences available. Analyses of the structure of his genes revealed that at least three different molecular mechanisms played an important role in shaping this pathway [2], i.e., gene duplication, gene fusion, and gene elongation, making this route a very good model for understanding the molecular mechanisms driving the shaping of metabolic pathways [15].

Since it is supposed that in many cases, functionally related genes arise from the duplication of an ancestral gene and subsequent evolutionary divergence [6], in the pre-genomic era, Horowitz [24] proposed that all of genes belonging to an operon might have been originated from an ancestral gene via duplication and divergence; as a corollary, all the histidine biosynthetic genes could have evolved via the duplication of a single ancestor gene. However, analysis of the nucleotide sequence of *his* genes in *E. coli* and *S. enterica* has not revealed any extensive sequence homology between these different genes [14]. In spite of this, an event of gene duplication led to the formation of two *his* genes, *hisA* and *hisF*, starting from a common ancestral gene. Gene fusion appears to have been one of the most important mechanisms of gene evolution in the histidine biosynthesis [25]; indeed, at least seven of the ten biosynthetic genes (i.e., *hisD*, *hisB*, *hisN*, *hisI*, *hisE*, *hisF*, and *hisH*) underwent single or multiple fusion events in different prokaryotic and eukaryotic phylogenetic lineages [15]. Finally, regarding the gene elongation event, the most documented example concerns the evolution of *hisA* and *hisF* genes, encoding two ( $\beta/\alpha$ )<sub>8</sub>-barrel (TIM-barrel) proteins [15].

The aim of this work was to evaluate the importance of gene elongation in the evolution of histidine biosynthetic genes and, if possible, to refine the model depicted for the evolutionary history of some of them and/or to trace the evolutionary trajectories of other ones.

# 2. Materials and Methods

### 2.1. Three-Dimensional Structure Prediction

All the 3D protein structures reported in the present work were predicted using Phyre2 software [26], and were visualized and modified with UCSF Chimera [27]. Files of the predicted structures are provided in PDB format in Supplementary Material S1.

# 2.2. PDB Search for Protein Structures

For HisA/HisF, HisB (IGPD), and HisD (HDH), available in the Protein Data Bank (PDB) archive [28], the secondary and 3D structures were investigated. The available structures (as of April 2020) are reported below.

HisA PDB IDs:

- 1QO2, 2CFF, 2W79—Thermotoga maritima;
- 1VZW, 2VEP, 2X30, 5DN1—Streptomyces coelicolor;
- 2AGK—Saccharomyces cerevisiae;
- 2Y85, 2Y88, 2Y89, 3ZS4—Mycobacterium tuberculosis;
- 4AXK—Corynebacterium efficiens;
- 4GJ1—Campylobacter jejuni;
- 4TX9, 4U28—Streptomyces sviceus;
- 4W9T, 4X9S—Streptomyces sp.;
- 4WD0—Paenarthrobacter aurescens;
- 4X2R—Actinomyces urogenitalis;
- 5AB3, 5ABT, 5G4W, 5G5I, 5G1T, 5G1Y, 5G4E, 5G2H, 5G2I, 5G2W, 5AC6, 5AC7, 5AC8, 5AHF, 5A5W, 5AHE, 5AHI, 5L9F—*Salmonella enterica*;
- 5L6U—Salmonella heidelberg.

# HisF PDB IDs:

- 1GPW, 1THF, 1VH7, 2A0N, 2WJZ, 3ZR4, 4EWN, 4FX7, 5TQL, 6VDG—Thermotoga maritima;
- 1H5Y—Pyrobaculum aerophilum;
- 1JVN, 1OX4, 1OX5, 1OX6—Saccharomyces cerevisiae;
- 1KA9—Thermus thermophilus.

# IGPD PDB IDs:

- 1RHY—Filobasidiella (Cryptococcus) neoformans;
- 2AE8—Staphylococcus aureus;
- 2F1D, 4MU0, 4MU1, 4MU3, 4MU4, 4QNJ, 4QNK, 5ELW, 5EKW, 5EL9, 6EZJ—Arabidopsis thaliana;
- 4GQU, 4LOM, 4LPF, 5XDS, 5ZQN—Mycobacterium tuberculosis;
- 5DNL, 5DNX—Pyrococcus furiosus;
- 6EZM—Saccharomyces cerevisiae;
- 6FWH—Acanthamoeba castellanii.

# HDH PDB IDs:

- 1K75, 1KAE, 1KAH, 1KAR—Escherichia coli;
- 4G07, 4G09—Brucella suis;
- 4GIC—Methylococcus capsulatus;
- 5VLB, 5VLC, 5VLD—Medicago truncatula;
- 6AN0—Elizabethkingia anophelis.

#### 2.3. Amino Acid Sequence Alignment

Amino acid sequences from HisA/HisF, IGPD, and HDH, available in the PDB, were downloaded from UniProt [29], aligned using BioEdit [30] through the ClustalW tool [31], and the conservation of the secondary structure organization was evaluated.

A total of 81 IGPD amino acid sequences from Archaea, 359 from Bacteria, and 95 from Eukarya were downloaded from UniProt to give a total of 535 IGPD sequences, chosen from every principal taxonomic group. Sequences were aligned using BioEdit through the ClustalW tool (Supplementary Material S2), and the conservation of the (D/N)XHHXXE motif was investigated.

#### 3. Results and Discussion

#### 3.1. hisA and hisF Gene Elongation

The hisA gene codes for a N'-[(5'-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamideribonucleotide (5'-ProFAR) isomerase catalyzing the fourth step of the histidine biosynthetic pathway, and the hisF gene encodes a cyclase catalyzing the fifth step (Figure 2) [19]. HisF is part of a heterodimeric complex, an imidazole glycerol phosphate (IGP) synthase composed of the cyclase subunit HisF and the glutamine amidotransferase (GAT) subunit HisH. The HisF: HisH association is a stable 1:1 dimeric complex [17], which is a branch-point enzyme, since its two products IGP and AICAR (5-aminoimidazole-4-carboxamide ribonucleotide) are used in histidine biosynthesis and in the de novo synthesis of purines, respectively [32]. HisA, like TrpF, catalyzes an Amadori rearrangement, i.e., the irreversible isomerization of an aminoaldose to an aminoketose [33]. The same chemical rearrangement is responsible for the imidazole ring closure by HisF [34]. HisA and HisF share the ability to bind PRFAR (N9-[(59-phosphoribulosyl)-formimino]-5-aminoimidazole-4-carboxamide-ribonucleotide): HisA catalyzes the formation of PRFAR, and HisF catalyzes an ammonia-dependent reaction in which PRFAR is converted to IGP and AICAR. Thus, they act like phosphoribosyl anthranilate isomerase (TrpF) and indole-3-glycerol phosphate synthase (TrpC), which catalyze successive reactions in the tryptophan biosynthetic pathway [35]. Both TrpF and TrpC are homologs to the HisF/HisA pair; although a prebiotic synthesis of tryptophan has been reported, it is likely that this amino acid has occurred later in evolution. Therefore, the HisF/HisA pair is more ancient than the TrpF and TrpC proteins, antiquity that is in agreement with the key catalytic role played by histidine in many extant enzymes [34].

The comparative analysis of the amino acid sequences of *hisA* and *hisF* gene products in different organisms showed a high degree of sequence similarity (ranging from 32% to 81%, considering similar amino acids), suggesting that the two genes are paralogous and that they originated from a common ancestor gene via duplication and subsequent evolutionary divergence [14]. HisA and HisF can be therefore considered examples of retrograde evolution of enzymes in a biosynthetic pathway; according to this hypothesis, when the substrate for an enzyme in a biosynthetic pathway is depleted, a new enzyme can evolve to supply the substrate via duplication of the original coding gene followed by divergent evolution [24,36]. A more elaborate analysis highlighted an internal duplication in both hisA and hisF genes. As proposed by Fani et al. [14], hisA and hisF may have originated from the duplication of a smaller ancestral gene, half the size of the current genes, followed by evolutionary divergence and fusion of the two derivatives. They proposed a "two-step model" for the evolution of *hisA* and *hisF* genes; according to this hypothesis, *hisA* originated from the duplication of a *hisA* 1 module, and *hisF* then originated from the duplication of *hisA*. The biological significance of this gene elongation event became clear when the structures of the HisA and HisF proteins were determined. Crystallographic studies showed that these two enzymes are structurally homologous  $(\beta/\alpha)_8$ -barrels (Figure 4) [37], with a two-fold repeat pattern; the first and second  $(\beta/\alpha)_4$ -half barrels of both enzymes are related by a two-fold axis of symmetry. These data led to the proposal of a model for the evolution of the HisA and HisF ( $\beta/\alpha$ )<sub>8</sub>-barrels via two successive gene duplication events: a first duplication of a single gene encoding a  $(\beta / \alpha)_4$ -half barrel, subsequent fusion and modest divergent evolution, followed by a second gene duplication and diversification leading to the extant enzymes HisA and

HisF [32]. The hypothesis that the *hisF* gene originated from *hisA* finds support in the organization and structure of these genes in different organisms; however, the possibility that *hisF* gave rise to *hisA* cannot be ruled out. Since the structure of *hisA* is the same in all the microorganisms in which this gene has been identified, it is likely that the duplication event leading to the entire *hisA* gene occurred before the earliest phylogenetic divergence [25]. Concerning the gene elongation event leading to the gene encoding the entire barrel, it is still not possible on the basis of the available data to discern between two different scenarios, i.e., the fusion event between the two copies could have occurred either before or after they underwent an evolutionary divergence.



**Figure 4.** Three-dimensional structure of *E. coli* HisA (**A**,**B**) and HisF (**C**,**D**). (**A**,**C**) top view of the barrels; (**B**,**D**) side view.  $\alpha$ -helices and  $\beta$ -strands are labeled.

It is known that the  $(\beta/\alpha)$ -barrel fold is a versatile structure for different enzymatic functions; indeed, 10% of known enzymes possess this fold [38], and catalyze very different reactions. The first protein that was discovered to have an eight-stranded  $(\beta/\alpha)$ -barrel domain was triose phosphate isomerase, hence the name "TIM-barrel" [39]. The barrel structure is composed of eight concatenated  $(\beta$ -strand)-loop- $(\alpha$ -helix) units. The  $\beta$ -strands delimit the internal lumen of the barrel, whereas the  $\alpha$ -helices form the external surface [15]. Although the fold is known as a barrel, the  $(\beta/\alpha)$  structure is usually not circular in cross section, but rather elliptical [39]. The active sites are located at the C-terminal ends of the barrels, with the active site functional groups found in the loops that connect the  $\beta$ -sheets with the  $\alpha$ -helices [38]. A large number of  $(\beta/\alpha)_8$ -barrel proteins (including HisA and HisF) contain phosphate-binding motifs formed by the C-terminal end of the seventh  $\beta$ -strand and the N-terminal end of the eighth  $\alpha$ -helix in the barrel [35]. Considering that the  $(\beta/\alpha)_8$ -barrel fold is the most common active site scaffold, and that this fold may also represent an ideal scaffold for the creation of new enzyme activities, since the chemistry catalyzed by the protein can be altered without loss of specificity [40], much attention has been devoted to understanding the phylogeny of proteins sharing this fold [35]. During the first years of the 2000s, lots of experiments were performed to better understand the evolution of this class of proteins. Jürgens et al. [33], using random mutagenesis and selection, generated several HisA variants that catalyze the same reaction of TrpF both in vivo and in vitro, suggesting that HisA and TrpF may have evolved from one common ancestral enzyme of broader substrate specificity. This is in full agreement with the patchwork hypothesis of the evolution of metabolic pathways, independently proposed by Yčas [41] and Jensen [42], in 1974 and 1976, respectively. Noteworthy, this occurrence is observed in the actinomycetes Streptomyces coelicolor and Mycobacterium tuberculosis, where a single enzyme takes part in both histidine and tryptophan biosynthesis. Indeed, *trpF* is not present in their genomes, and the corresponding enzymatic activity is encoded by priA, the ortholog gene of hisA [43]. Gerlt and Babbitt [35] studied the evolution of IGP synthase. The N- and C-terminal  $(\beta/\alpha)_4$ -half barrels of HisF were separately expressed and purified. Each one assumed a stable and soluble homodimeric structure, but neither was catalytically active; in contrast, when expressed together, a functional heterodimer was formed, suggesting that the ancestral half-barrel gave a functional enzyme by homodimerization. Fani et al. [15] investigated the possibility of an even older gene elongation event involving  $(\beta/\alpha)$ -mers smaller than the  $(\beta/\alpha)_4$ units of the ancestral half-barrel precursor. We cannot exclude the possibility that an ancestral gene encoding a single ( $\beta/\alpha$ )-module might represent one of the "starter types", i.e., one of the few most ancestral genes not derived from the duplication of a pre-existing one [44]. According to this idea, that module might have been able to aggregate in a homo-octamer to form a still unstable and thus not efficient complete TIM barrel, which, according to the patchwork hypothesis [42], might have been embedded with a broad specificity. Subsequently, a "cascade" of three gene elongation events would have given rise to the complete ancestor of the extant TIM-barrel coding genes (Figure 5). These studies highlighted that the elongation event leading to the ancestor of *hisA* and *hisF* genes resulted in the covalent fusion of two half-barrels; once assembled, the complete gene underwent gene duplication, leading to the ancestor of *hisA*, *hisF*, and maybe all the TIM-barrel-protein-coding genes. This is the most parsimonious scenario to explain the origin and evolution of *hisA* and *hisF* genes. However, since, according to Woese [45], the early cells might have been embedded with multiple short informational molecules, it could be also possible that different copies of the  $(\beta/\alpha)$  coding starter type might exist in the same or different DNA molecules in the progenote. Hence, the possibility that the ancestor  $(\beta/\alpha)_2$  coding gene might have been the result of a domain shuffling event cannot be excluded a priori [46]. Once the two different ( $\beta/\alpha$ ) coding genes fused, the resulting gene might, in turn, have undergone a gene elongation event or one or more additional domain shuffling event(s) (Figure 5). Indeed, the degree of sequence similarity shared by the single  $(\beta/\alpha)$  modules is not sufficient to support the three gene elongations model. However, the degree of sequence similarity between the two halves of the HisA and HisF proteins strongly suggests that the final molecular rearrangement leading to the extant *hisA* and *hisF* genes should have been the elongation of a gene half the size of the extant ones. Interestingly, HisA is the only one maintaining an almost perfect subdivision in two modules that are half the size of the entire gene and that share a high degree of sequence similarity [15].

In the present work, for those organisms for which HisA and/or HisF (or their relative orthologs, as PriA for *Streptomyces* spp. and *M. tuberculosis*, and His6 and His7 for *Saccharomyces cerevisiae*) 3D structures were available in the PDB (see Materials and Methods), the secondary structure was investigated, highlighting the conservation of the  $\alpha$ -helix and  $\beta$ -strand repetitions among members of different taxonomic groups (Figure 6).

All these findings support the idea that the reaction specificities of current  $(\beta \alpha)_8$ -barrel enzymes evolved from an ancestral enzyme with low substrate specificity, which appeared before LUCA. Subsequently, this enzyme must have undergone several evolutionary divergence events, leading to

the high variety of TIM-barrel proteins currently present. An important corollary of this idea is that  $(\beta \alpha)_8$ -barrel enzymes might be engineered to acquire new activities [19].



Figure 5. Evolutionary model proposed to explain the origin and evolution of *hisA* and *hisF* genes.

		β1	<u> </u>	α1 β2	>
	210 220 230	240 250	260 270	280 290	300
CONTRAL High A unegonitalia			.         .		•
Q8FNZ7 HisA C. efficiens	м	TFTILPAVDVVNGQAVRLDQ	GEAGTEKSY	G-TPLESALRWQEQGAEWLHFV	DL
Q9PM74 HisA C. jejuni		MTQIIPALDLIDGEVVRLVK	GDYEQKKVY	KYNPLKKFKEYEKAGAKELHLV	DL
AlR562 HisA P. aurescens	MTTSAQS	VLELLPAVDIVDGQAVRLLQ	GEAGSETSY	G-TPLEAALNWQNDGAEWVHMV	DL
P40545 His6 S. cerevisiae P16250 PriA S. coelicolor	MS	MTKFIGCIDLHNGEVKQIVG KLELLPAVDVRDGOAVRLVH	GTLTSKKED GESGTETSY	VPKTNFVSQHPSSYYAKLYKDR G-SPLEAALAWORSGAEWLHLV	
B4T9N8 HisA S. heidelberg		MIIPALDLIDGTVVRLHQ	GDYARQRDY	GNDPLPRLQDYAAQGAGVLHLV	DL
B5I4P8 PriA S. sviceus P10372 HisA S. tvphimurium	MPSKLS	RLELLPAVDVRDGQAVRLVH MIIPALDLIDGTVVRLHO	GESGTETSY GDYARORDY	G-SPLEAALSWQRAGAEWLHLV GNDPLPRLODYAAOGAGVLHLV	DL DL
B4V386 PriA Streptomyces sp.	MPVN	KLELLPAVDVRDGQAVRLVH	GVSGSETSY	G-SPLEAALAWQASGAEWLHLV	DL
Q9XUC7 HISA T. maritima Q8ZY16 HisF P. aerophilum	MDV	ALRIIPCLDIDGKAGVVVK	GVNFQGIREV	G-DPVELVERLIEEGFTLIHVV	DI
P33734 His7 S. cerevisiae	VIENFLKQQSPPIPNYSAEEKELLMNDYSNYGL	TRRIIACLDVRTNDQGDLVVTK	GDQYDVREKSDGKGVRNL	G-KPVQLAQKYYQQGADEVTFL	NI
Q7SIB9 HisF T. thermophilus	MSL	AKRIVPCLDVHAGRVVK	GVNFVNLRDA	G-DPVEAARAYDEAGADELVFL	DI
	(7) (83)	α3	BA	α <b>4</b>	
				u+	
	310 320 330	340 350	360 370 •   • • • •   • • • •   • • • •   •	380 390	400
COW7K4 HisA A. urogenitalis	DAAFGRGS-NAPLLERIVGEVGIKVELSGGIR-	DDASLTRALKAG	AARVNLGTAALEDPQ	WTARVIAEHG-E	KI
Q8FNZ/ HISA C. efficiens Q9PM74 HisA C. jejuni	TGAKDPSKRQFALIEKLAKEVSVNLQVGGGIR-	SKEEVKALLDCG	WRVVIGSMAIKDAT	LCLEILKEFGSE	AI
P9WMM5 PriA M. tuberculosis	DAAFGRGS-NHELLAEVVGKLDVQVELSGGIR-	DDESLAAALATG	CARVNVGTAALENPQ	WCARVIGEHG-D	QV
P40545 His6 S. cerevisiae	QGCHVIK-LGPNNDDAAREALQESPQFLQVGGGIN-	DTNCLEWLKW	ASKVIVTSWLFTKEG	HFQLKRLERLTELCGKD	RI
P16250 PriA S. coelicolor B4T9N8 Hish S. beidelberg	DAAFGTGD-NRALIAEVAQAMDIKVELSGGIR- TGAKDPAKROIPLIKTLVAGVNVPVOVGGGVR-	DDDTLAAALATG	CTRVNLGTAALETPE	WVAKVIAEHG-D	KI AT.
B5I4P8 PriA S. sviceus	DAAFGTGD-NRELVRQVTEAMDIKVELSGGIR-	DDASLAAALATG	CTRVNLGTAALESPE	WVAKVIAEHG-D	RI
P10372 HisA S. typhimurium B4V386 PriA Streptomyces sp.	TGAKDPAKRQIPLIKTLVAGVNVPVQVGGGVR- DAAFGTGD-NRALVAEITGAMDIKVELSGGIR-	TEEDVAALLKAG	VARVVIGSTAVKSPD CTRVNLGTAALETPE	VVKGWFERFGAQ	AL RI
Q9X0C7 HisA T. maritima	SNAIENSGENLPVLEKLS-EFAEHIQIGGGIR-	SLDYAEKLRKLG	YRRQIVSSKVLEDPS	FLKSLREID-V	EP
282116 Hisf P. aerophilum P33734 His7 S. cerevisiae	TAAPEGRATFIDSVKRVAEAVSIPVLVGGGVR- TSFRDCPLKDTPMLEVLKQAAKTVFVPLTVGGGIKD	SLEDATTLFRAG IVDVDGTKIPALEVASLYFRSG	ADKVSVNTAAVRNPQ ADKVSIGTDAVYAAEKYY	ELGNRGDGTSPIETISKAYGAQ	AV
Q9X0C6 HisF T. maritima	TASVEKRKTMLELVEKVAEQIDIPFTVGGGIH-	DFETASELILRG	ADKVSINTAAVENPS	LITQIAQTFGSQ	AV
Q'SIBSTRISE 1. CHEIMOPHIIUS	SATHEERAILEDVVARVAERVFIFETVGGGVR-	STEDWKKTTPP26	MDRVSVNSAAVRRPL	DI KELADIR GAQ	
	β5	α5	β6	α6	
	β5 410 420 430	α5 440 450	β6 460 470	α6 480 490	50(
COWYK4 [Hish A. urogenitalis	β5 410 420 430 	440 450 	460 470 	α6 480 490 	50( .  AP
COW7K4 EisA A. urogenitalis Q8FNZ7 EisA C. efficiens Q9FN74 EisA C. jejuni	β5 410 420 430 	440 450	β6 460 470 LNEACCRYVYDVTDVTKDG LDSQCCSRFVVTDVSKDG 9SNKGLKHLCTDJSKDG	α6  480 490  TITGPNTELLRQVAARTS,  TITGPNVDLLRQVAAATS,  MYGGVNYRLKULHELFPN	50 .  AP AP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FN74 HisA C. jejuni F9WM45]FriA M. tuberculosis	β5 410 420 430 	440 450 LAARGWIKE GG-D LWQTLDR WEIT KONGWSD GG-D LWEVLER HRLIRGRGWETD GG-D LWDVLER	460 470 LINEAGCREYVVTUVTKDG LINEAGCSEFVVTUVTKDG LINEAGCSEFVVTUVTKDG LINEAGCSEFVVTUTTKDG	480 490 	500 •   AP AP IC AP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNT4 HisA C. jejuni P9MAM5 FriA M. tuberculosis A18562 HisA P. aurescens P40545 His6 S. cerevisiae	β5 410 420 430 	440 450 	460 470 LINEACCRYVVTUVTKO LINEACCRYVVTUVTKO SINKCLKILLCTU SIKO LIDSECSRFVVTU ITKO LEDACCRYVVTUVTKO ELRVYTNETL HAADVES		500 .  AP IC AP KP KP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NRM5 F1Ri M. tuberculosis A1R562 HisA P. aurescens P40545 His6 S. cerevisiae P16250 FriA S. coelicolor	β5 410 420 430 	440 450 	460 470 Δ100 Δ100 Δ100 Δ100 Δ100 Δ100 Δ100 Δ1		500 .   AP IC AP KP LK RP
COW7K4 HisA A. urogenitalis Q8FN27 HisA C. efficiens Q9FN74 HisA C. jejuni P9NMM5 FirIA M. tuberculosis A1R562 HisA P. aurescens P40545 His6 S. cervisiae P16250 PriA S. socilcolor B4T9N8 HisA S. heidelberg B514P8 FriA S. sviccus	β5        410      420      430        AVGLDVRGTT      1      1      1        AVDIAVRLENG      E      2      2        VLALDY ILKED      AVGLDVQTIDG      E      2        AVGLDVGTT      VULSCRTKTOPG      R      AVGLDVRGTT      VULSCRTKTOPG        VLALDVRIDE      AVGLDVRGTT      T      T      T        VLALDVRIDE      AVGLDVRGTT      T      T	440 450 	460 470 LNEAGCREYVYDUYTKDG LDS-QCSRFVYDUSKDG USNGCLKHILCTDISKDG LEDAGCARYVYDUYTKDG LEDAGCARYVYDUYTKDG LIKNGCARYVYDIATADVEG LIKNGCARYVYDIATADVEG LIKNGCARYVYDIATADVEG LIKNGCARYVYDIATADVEG	480 490 TITGPNVELLRQVAART	500 .  AP IC AP KP LK RP IA RP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NMM5 FirlA M. tuberculosis A18562 HisA P. aurescens P40545 His6 S. cerevisiae P16250 PriA S. socilcolor B4T9N8 HisA S. heidelberg B514P8 FriA S. styphimurium B4V386 PriA Strotowces SD.	β5        410      420      430        AVGLDVRGTT-           AVDIAVRLENG-      E          VLALDTILKED-            AVGLDVGTTIDG-             VVDLSCRKTQDG-   <	440 450 	460 470 LNEACCRYVVIDVTKDG LDSQCCSFVVIDVTKDG LDSQCCSFVVIDVTKDG LDSECSFVVIDVTKDG ELRKVTNEFLIHAADVEG LINECCARYVVIDVTKDG LINECCARYVVIDVIDVTKDG VIEVGLKIVLCTDISRDG LINECCARYVVIDVIDVISRDG VIEVGLKIVLCTDISRDG	480 490 TITGPNNELLRQVAART	500 .   AP IC AP IC AP IC IR RP IA RP IA RP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NMM5 FriA M. tuberculosis A18562 HisA F. aurescens P40545 His6 S. cerevisiae P16250 FriA S. secilcolor B4T9N8 HisA S. heidelberg B51478 FriA S. svjceus P10372 HisA S. typhimurium B4386(FriA Streptomyces sp. Q4007(HisA T. maritima	BS        410      420      430        AVGLDVRGTT           AVDIAVRLENG           AVGLDVRGTT           AVGLDVRGTT           AVGLDVRGTT           VDLSCRKTQDG           AVGLDVRGTT           VLALDVRIDERG           VLALDVRIDERG           VLALDVRIDERG           VSLDTRGGR	440 450 	460 470 LNEACCRYVVTDVTKDG LDSQCCSFVVTDVTKDG LDSQCCSFVVTDVTKDG LDSGCCSFVVTDVTKDG ELRVTNEFLHAADVEG LNECCAFVVTDITKDG LDSECCAFVVTDITKDG LDFECCAFVVTDISRDG LDFECCAFVVTDISRDG LDFECCAFVVTDISRDG LDFECCAFVVTDISRDG LDFECCAFVVTDISRDG	480 490 TITGPNNELLRQVAART	500 .   AP IC AP KP LK RP IA RP IA RP VK
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. efficiens Q9FNZ4 HisA C. jejuni P9WMM5 PriA M. tuberculosis A18562 HisA F. aurescens P40545 HisA F. aurescens P16250 FriA S. scielicolor B4T9N8 HisA S. heidelberg B51488 FriA S. typhimurium B4V386 FriA S. typhimurium B4V386 FriA S. typhimurium B4V386 FriA S. treptomyces sp. Q9XC7 HisA T. martina Q8Z16 HisF F. aerophilum P33734 His7 S. crevisiae	β5        410      420      430        AVGLDVRGTT	440 450 	460 470 LNEAGCREYVVTDVTKDG LSEAGCREYVVTDVTKDG LSEAGCREYVVTDVTKDG YSNKGLKHILCTDISKDG LSEAGCSRFVVTDITKDG ELRYVTDISKDG LDSEGCAEYVVTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG YLPVGLKIRVLCTDISKDG VESLGAGSILLTSTDRDG CERLGAGSILLTSTDRDG		500 .  AP AP IC AP KP IA RP IA RP IA RP IA RP IA RP IA
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9WMM5 PriA M. tuberculosis A18562 HisA P. aurescens P40545 HisA S. cerevisiae P16250 PriA S. coclicolor B4T9N8 HisA S. heidelberg B51488 PriA S. styphimurium B4V386 PriA S. typhimurium B4V386 PriA S. typhimurium G8ZV16 HisF T. aerophilum C7GFB0 HisF T. maritima	β5        410      420      430        AVGLDVRGTT	440 450 	460 470 LNEAGCREYVVTDVTKDG LSEAGCREYVVTDVTKDG LDSGCSEFVVTDVSKDG YSNKGLKHILCTDISKDG SNKGLKHILCTDISKDG ELRKYTNEFLHAADVEG LNEGCAEYVVTDIAKDG YLPVGLKIVLCTDISEDG LDKEGCAEYVVTDIAKDG YLPVGLKIVLCTDISEDG LDSEGCAEYVVTDIAKDG VEFLGAGEILLINCIDKDG VEFLGAGEILLINCIDKDG VEFLGAGEILLINCIDKDG		500 .  AP IC AP IC AP IC RP IA RP IA RP IA RP IA RP IA RP IA
COWYK4 HisA A. urogenitalis QSENZ7 HisA C. efficiens Q9EN74 HisA C. efficiens J9MAM5 FriA M. tuberculosis Al8562 HisA F. aurescens P40545 HisA S. avervisiae P16250 FriA S. coelicolor B4T9N8 HisA S. tyridelberg B514P8 FriA S. sviceus P10372 HisA S. tyridelberg B47086 FriA S. treptomyces sp. Q9X027 HisA T. maritima Q82X16 HisF T. aeruphilum P33734 HisF T. thermophilus	β5        410      420      430        AVGLDVRGT	440 450 	460 470 LNEAGCREYVVTDVTKDG LSEAGCREYVVTDVTKDG SNKGLKHILCTDISKDG YSNKGLKHILCTDISKDG SNKGLKHILCTDISKDG LDSEGCSRYVVTDISKDG LDSEGCARYVVTDISKDG YLPVGLKIVLCTDISKDG YLPVGLKIVLCTDISKDG YLPVGLKIVLCTDISKDG YLPVGLKIVLCTDISKDG VEFLGAGEILLINCIDKDG VEFLGAGEILLINCIDKDG GVELGAGEILLINCIDKDG		500 .   AP IC AP IC AP IC AP IC IC IC IC IC IC IC IC IC IC
COWYK4 HisA A. urogenitalis QSENZ7 HisA C. efficiens Q9EM74 HisA C. jejuni P9MAM5 FriA M. tuberculosis AltS62 HisA P. aurescens P40545 HisA S. acerisiae P16250 FriA S. coelicolor B4T9N8 HisA S. haidelberg B514P8 FriA S. sviceus P10372 HisA S. typhimurium B4V386 FriA Streptomyces sp. Q9X027 HisA T. martina Q82X16 HisF T. aerophilum P33734 HisF T. thermophilus	β5        410      420      430        AVGLDVRGTT      1      1        AVDILAVRLENG      E      2        AVGLDVRGTT      E      2        VNLADFILKED      E      2        AVGLDVRGTT      E      2        VNLSCRTCDBG      E      E        AVGLDVRGTT      E      2        VLALDVRIDERG      T      T        AVGLDVRGTT      VLALDVRIDERG      T        VNLADKINGE      Y      Y        VVALDAKINGE      Y      Y        VALDAKINGE      Y      Y        VALDAKINGE      Y      Y        YALDAKINGE      Y      Y        YALDAKINGE      Y      Y        YALDAKINGE      Y      Y	440 450 	460 470 LNEAGCREYVVTDVTKDG LSEAGCREYVVTDVTKDG YSNKGLKHILCTDISKDG SYNKGLKHILCTDISKDG LDSEGCSRFVVTDITKDG ELRYGTNEFLHAADVEG LINEGCARYVVTDISKDG YLPYGLKIKVLCTDISKDG YLPYGLKIKVLCTDISKDG YLPYGLKIKVLCTDISKDG YLPYGLKIKVLCTDISKDG YLPYGLKIKVLCTDISKDG VEELGAGEILLINCIDKDG VEELGAGEILLINCIDKDG GVELGAGEILLINCIDKDG	α6        480      490        TLTGPNTELLRQVARTS        TITGPNYLLRDVARTD        TMGGVNURLYKL HEITFPN        TLGGPLDLLAGVARTD        TLGGPNULLKNUCARTD        TLGGNUSLLRQVARTD        TLGGNUSLLRVCARTD        TLGGNUSLLRVCARTD        TLGGNUSLERVCARTPQ        TLGGNUSLERVCART        TLGGUSTELRVCART        TLGGUSTELRVCART	500 .   AP AP KP LK RP LA RP LA RP LA RP LA RP VK IP LP VP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9MMM5 FriA M. tuberculosis A18562 HisA P. aurescens P16250 FriA S. coelicolor B4T986 HisA S. toidelberg B51486 FriA S. sviceus P10372 HisA S. typhimurium B4V366 FriA S.treptomyces sp. Q9X0C6 HisF T. maritima Q3X16 HisF T. maritima Q7SIB9 HisF T. thermophilus	β5        410      420      430        AVGLDVGFGT      -      -        AVDLDVGFGT      -      -        AVGLDVGFGT      -      -        AVGLDVGFGT      -      -        AVGLDVQTDEBEG      -      -        AVGLDVQTDEBEG      -      -        VI_ALDVRIDEBEG      -      -        VVALDAKNNGE      -      -        VVALDAKWNGE      -      -        VALDAKNOG      -      -        Ø7      Ø8      -	440 450 LARROWTKE GG-D LWQTLDB HITKKNNGWSDG-D LWQTLDB HITKKNNGWSDG-D LWQTLBB YVVANNAQEASDKKLMEVLDE YVVANNAQEASDKKLMEVLDE LAGRGWTE GG-D LWQTLBR HIVVANNKQCHTUDLELNADTER 	460 470 400 470 400 400 400 400 400 400 40	480 490 	500 .   AP AP IC AP KP IA RP IA RP IA RP IA RP VK IP IP VP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NMM5 FriA M. tuberculosis A18562 HisA P. aurescens P40545 HisA S. aversisiae P16250 FriA S. coelicolor B4T9N8 HisA S. tophimrium P10372 HisA S. sviceus P10372 HisA S. sviceus P10372 HisA S. typhimrium Q82116 HisF P. aerophilum P33734 HisF T. marifina Q7SIB9 HisF T. thermophilus	β5        410      420      430        AVGLDVRGFT	440 450	AGO 40 40 40 40 40 40 40 40 40 40	udd      480    490      1110    1110	500 .   APP ICP ICP ICP ICP ICP ICP ICP I
COW7K4   HisA A. urogenitalis Q8FNZ7   HisA C. efficiens Q9FN74   HisA C. jejuni P9NMM5   FriA M. tuberculosis Alt562   HisA S. aurescens P16250   FriA S. coelicolor B4T9N6   HisA S. heidelberg B5T4P6   FriA S. sviceus P10372   HisA S. sviceus P10372   HisA S. typhimurium B4V386   FriA Streptomyces sp. Q9X0C7   HisA T. maritima Q3X16   HisF T. maritima Q7SIB9   HisF T. thermophilus COW7K4   HisA A. urogenitalis	β5        410      420      430        AVGLDVRGTT      430      430        AVGLDVRGTT      430      430        VLALDVRIENG      E      E        AVGLDVRGTT      630      E        VLALDVRIDERG      E      E        AVGLDVRGTT      7      100        VLALDVRIDERG      T      T        AVGLDVRGTT      7      100        VVAIDARWRGD      F      510      520      530        VVASGGISSLEDIAALARUVPGOUSAIVGKAL      SLEDIAALARUVPGOUSAIVGKAL      F	440      450       LAARGWTKEGG-DLWQTLDD       LARGWTKEGG-DLWQTLDE        WIXTANGWQLSSKILMEVLDE        VUVAVNAWQLSSKILMEVLDE        VUVAVNAWQLSSKILMEVLDE       LAGRGWTEGG-DLWEVLRA        KUVANSGUSSKSULGULVET       LKGRGWTESGG-DLYETLAR        -VAFKGWLSEGG-DLYETLAR        S40      550        YMG - NTLEQALAVAGGAAVQ	A60 470 LNEAGCREYVYTUVTKDG LSSQCSEFVYTUVSKDG LSSQCSEFVYTUVSKDG LSSQCSEFVTUVSKDG LEDAGCARYVYTUVTKDG LEDAGCARYVYTUTKDG LEDAGCARYVYTUTKDG LNEGCARYVYTUTKDG LNEGCARYVYTUTKDG LNEGCARYVYTUTIKNDG LNEGCARYVYTUTIKNDG SECALVYTUTIKNDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDG SECALVYTUTIKTSKDF SECALVYTUTKT SECALVYTUTTKT SECALVYTT SECALVYTT SECALVYTT SECALVYT	a6      480    490      TITGPNVELLRQVAARTD      TITGPNVELLRQVAARTD      TIGGPNLLLAGVARTD      TLQGPNLELLRQVAATD      TLQGPNLELLRNVCAATD      TLQGPNLELLRNVCAATD      TLQGPNLELLRNVCAATD      TLQGPNLELLRNVCAATD      TLQGPNLELLRNVCAATD      TLQGPNLELLRNVCAATD      TLGSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSU	500 .   AP AP IC
COW7K4   HisA A. urogenitalis Q8FNZ7   HisA C. efficiens Q9FNZ7   HisA C. jejuni P9NMM5   FriA M. tuberculosis AltS62   HisA P. aurescens P16250   FriA S. tuberculosis P16350   FriA S. coelicolor B4T9N8   HisA S. heidelberg B51478   FriA Streptomyces sp. Q9X07   HisA T. maritima Q8Z16   HisF P. aerophilum P33734   HisF T. mtritma Q7SIB9   HisF T. thermophilus COW7K4   HisA A. urogenitalis Q8ENZ6   HisA C. efficiens	β5        410      420      430        AVGLDVRØTT      -      -        VNLADVRIDT      -      -        VVLLSCRITC      -      -        VVLLSCRITC      -      -        VVLLSCRITC      -      -        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        VVLALDKYLDEHG      -      -        VVAIDAKINGE      -      -        VVAIDAKINGE      -      -        VIAIDAKINGE      -      -        VVAIDAKINGE      -      -        VVASGGIS      -      -        VVASGGIS      -      -        VOASGUN      -      -	440      450	460 470 LINEAGCREYVYDUYTKDG LINEAGCREYVYDUYTKDG USNGCIKIILLCTUISKDG USNGCARYVYDUYTKDG LIENACARYVYDUYTKDG LIENACARYVYDIISKDG LIENACARYVYDIISKDG LIENACARYVYDIISKDG LIENACARYVYDIISKDG LIENACARYVYDIISKDG LIENACARYVYDIISKDG USEGCARYVYDIISKDG USEGCARYVYDIISKDG USEGCARYVYDIISKDG CEALGAGEILLINSIDROG GVELGAGEILLINSIDROG GVELGAGEILLINSIDROG S60 	480    490      1110000000000000000000000000000000000	500 .   AP IC AP IC AP IC AP IC IC IC IC IC IC IC IC IC IC
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NMM5 FiRA M. tuberculosis AlR562 HisA P. aurescens P40545 His6 S. cerevisiae P16250 PriA S. boiceus P10372 HisA S. boiceus P10372 HisA S. boiceus P10372 HisA S. boiceus Q9X16 HisF P. aerophilum P33734 HisF P. aerophilum P33734 HisF T. maritima Q7SIB9 HisF T. thermophilus COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNA5 FiRA M. tuberculosis	β5        410      420      430        AVGLDVRGTT      420      430        AVGLDVRGTT      400 LAVGLDVRGTT      400 LAVGLDVRGTT        VVLLSCRTVRDBG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVLLSCRTVRDBG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVLLSCRTVRDBE      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVLLSCRTVRDBE      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVLLADARNORG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVLLADARNORG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVLLADARNORG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VVALDARNORG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VIALDARNORG      700 LAVGLDVRGTT      700 LAVGLDVRGTT        VIALDARNORG      700 LAVGLDVRGTGT      700 LAVGLDVRGTGT        VIALDARNORG      700 LAVGLDVRGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	440      450       LAARGWTKE GG-D LWQTLDD        WITKONGWYJSGG-D LWQTLDD        WITKONGWYJSGG-D LWQTLDD        WITKONGWYJSGG-D LWQTLDD       LAGRGWTKE GG-D LWQTLDD       LAGRGWTKE GG-D LWQTLDD       LKORGWTES GG-D LWQTLDD       LKORGWTSE GG-D LWQTLDD       LKORGWTSE GG-D LYFALER        VYVYNGGUSSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUSUS	460 470 LINEAGCREYVYDUYTKDG LINEAGCREYVYDUYTKDG USSGCSEFVYDUYTKDG SSNCGLKHILCTDISKDG LEDAGCARYVYDUYTKDG LEDAGCARYVYDUYTKDG LINEGCARYVYDIILHADVEG LINEGCARYVYDIILHADVEG LINEGCARYVYDIILHADVEG LINEGCARYVYDIILHAD LINEGCARYVYDIILAN LINEGCARYVYDIILHAD LINEGCARYVYDIILHAD LINEGCARYVYDIILHAD LINEGCARYVYDIILHAD LINEGCARYVYDIILHAD LINEGCARYVYDIILHAD LINEGCARYVYDIILHAD LINEGCARYVYDIILAN LINEGCARYVYDIILAN LINEGCARYVYDIILAN LINEGCARYVYDIILAN LINEGCARYVYDIILAN LINEGCARYVYDIILAN LINEGCARYVYDII	480    490      1110000000000000000000000000000000000	500 .   APP APP IC APP KPR IC APP
COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NMM5 FiRA M. tuberculosis A1R562 HisA P. aurescens P40545 HisA S. corrisiae P10550 PiA S. social B10372 HisA S. heidelberg B51478 FiA S. styphimurium B47386 FiA Streptomyces sp. Q9X0C7 HisA T. maritima Q8Z116 HisF P. aerophilum Q3X0C6 HisF T. maritima Q7SIB9 HisF T. thermophilus COW7K4 HisA A. urogenitalis Q8FNZ7 HisA C. efficiens Q9FNZ7 HisA C. efficiens Q9FNZ7 HisA C. jejuni P9NM55 FiA M. tuberculosis A1R562 HisA P. aurescens P40545 His6 S. corrections	β5        410      420      430        AV0LDVRGTT      -      -        VIALDPILDBEG      -      -        AV0LDVRGTT      -      -        VIALDDVRDEGG      -      -        VVALDDRIDEGG      -      -        VVALDDRINGE      -      -        VVALDDRINGE      -      -        VIALDVRIDEGG      -      -        VIALDVRIDEGG      -      -        VVALDARINGE      -      -        VIALDRANNGE      -      -        VIADARINGE      -      -        VIADARNNGE	440 450	460 470 LNEACCRYVYDVTKDG LSQCCSFVYDDYSKDG USSCCSFVYDDYSKDG LDSQCCSFVYDDYSKDG LDSGCSRVVTDYTKDG LDSGCSRVVTDYTKDG LDSGCSRVVTDIXKDG LDSGCARVVTDIXKDG LDSGCARVVTDIXKDG LDSGCARVVTDIXKDG LDSGCARVVTDIXKDG LDSGCARVVTDIXKDG SCOL LCSLGAGZILLTSIDRDG GVELGAGZILLTSIDRDG GVELGAGZILLTSIDRDG SCO SCO SCO SCO SCO SCO SCO SCO	480    490      111000000    1110000000000000000000000000000000000	500 .   AP AP IC AP KP IA RP VK IP IP VP
COW7K4 HisA A. urogenitalis QSFNZ7 HisA C. efficiens QSFNZ7 HisA C. jejuni PSMAM5 FriA M. tuberculosis AltS62 HisA P. aurescens P16250 FriA S. coelicolor B479N8 HisA S. haidelberg B51498 FriA S. sviceus P10372 HisA S. typhumurium B47386 FriA S. typhumurium B47386 FriA S. typhumurium QSZ16 HisF P. aerophilum P33734 His7 S. cerevisiae QSXC6(HisF T. maritima Q7SIB9 HisF T. thermophilus COW7K4 HisA A. urogenitalis QSENZ7 HisA C. efficiens QSENZ7 HisA C. jejuni P9WAM5 FriA M. tuberculosis AltS62 HisA P. aurescens P40545 His6 S. cerevisiae P16250 FriA S. celicolor	β5        410      420      430        AVGLDVRGT      1      1      1        AVDILDVRGT      1      1      1      1        AVDILDVRGT      1      1      1      1      1        AVGLDVRGT      1      1      1      1      1      1      1        AVGLDVQIDGT      2      2      1	440      450	460 470 LNEACCRYVVTDVTKDG LDSQCCSFVVTDVTKDG LDSQCCSFVVTDVTKDG LDSQCCSFVVTDVTKDG ELRKVTNEFLHAADVEG LDACCARYVVTDVTDVTKDG ELRKVTNEFLHAADVEG LNECCARYVVTDIXKDG LNECCARYVVTDIXKDG LNECCARYVVTDIXKDG LNECCARYVVTDIXKDG SECARYVVTDIXKD LNECCARYVVTDIXKDG SECARYVVTDIXKD SECARYVVTDIXKD SECARYVVTDIXKD SECARYVVTDIXKD SECARYVTD SECARYVTDIXKD SECARYVTDIXKD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SECARYVTD SE	480 490 	500 .   AP IC AP IC AP IC AP IC IC IC IC IC IC IC IC IC IC
COW7K4 HisA A. urogenitalis QSFNZ7 HisA C. efficiens QSFNZ7 HisA C. efficiens QSFNZ7 HisA C. jejuni PSMMM5 FrIA M. tuberculosis AltS62 HisA S. aurescens P16250 FrIA S. coelicolor B4T98 HisA S. toicelberg B51486 FrIA S. sviceus P10372 HisA S. typhimurium B4V366 FrIA S.treptomyces sp. QSMC5(HisF T. maritima QS216 HisF T. maritima QS216 HisF T. maritima QS159 HisF T. thermophilus COW7K4 HisA A. urogenitalis QSFNZ7 HisA C. jejuni PSMM5 FrIA M. tuberculosis AltS62 HisA P. atrescens P40545 HisA P. atrescens	β5        410      420      430        AVGLDVGCTT      -      -        AVDILDVGCTT      -      -        AVGLDVGCTT      -      -        AVGLDVGCTT      -      -        AVGLDVGCTT      -      -        AVGLDVGCTT      -      -        VLALDVILDEBG      -      -        AVGLDVRGTT      -      -        VVALDAKNNGE-      -      -        VVALDAKNOG-      -      -        VVALDAKNOG      -      -        VVADAKKUNO      -      -        VVADAKKUNO      -      -        VVASQGIS	440      450	460 470 LNEAGCREYVVTDVYTKDG LSSQCCSRYVVTDVYTKDG LSSQCCSRYVVTDVYTKDG LEDAGCREYVVTDVYTKDG LEDAGCARYVVTDTKDG LEDAGCARYVVTDTKDG LNEGCARYVVTDTKDG LNEGCARYVVTDTSKDG LNEGCARYVVTDTSKDG LNEGCARYVVTDTSKDG LNEGCARYVVTDTSKDG SCOL LNESCCARYVVTDTSKDG SCOL LNESCCARYVTDTSKDG SCOL	480 490 	500 .  AP IC AP IC AP IC AP IC AP IC IC IC IC IC IC IC IC IC IC
COW7K4  HisA A. urogenitalis Q8FNZ7  HisA C. efficiens Q9FN74  HisA C. jejuni P9NMM5  FPIA M. tuberculosis A18562  HisA S. aurescens P16250  FPIA S. coelicolor B4T9N8  HisA S. heidelberg B51498  PPIA S. sviceus P10372  HisA S. sviceus P10372  HisA S. typhimurium P33734  His7 F. aervisiae Q9X0C6  HisF T. maritima Q7SIB9  HisF T. atrima Q7SIB9  HisF T. thermophilus COW7K4  HisA A. urogenitalis Q8FN74  HisA C. jejuni P9NMM5  FPIA M. tuberculosis A18562  HisA S. theidelberg B16250  FPIA S. cervisiae P10572  HisA S. theidelberg B16250  FPIA S. sviceus P40545  HisA S. theidelberg B16250  FPIA S. sviceus P10572  HisA S. theidelberg B1479N8  HisA S. theidelberg B1479N8  HisA S. sviceus	β5        410      420      430        AVGLDVRGFT      AVGLDVRGFT      AVGLDVRGFT        AVGLDVRGTT      AVGLDVRGTT      B        VULLOVRTDEBG      R      AVGLDVRGTT        VULLOVRTDEBG      T      T        AVGLDVRGTT      VILLOVRTDEBG      T        VVLLOVRTDEBG      T      T        VVLLOVRTDEGG      T      T        VVLLOVRTSGEGG      T      T        VVLLOVRTSGEGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	440      450	460 470 LNEAGCRYVVTUVTKDG LSSQCSFVTUVTVKDG LSSQCSFVTUVTVKDG LEDAGCRYVVTUVTKDG LEDAGCARYVVTUVTKDG LEDAGCARYVVTUTKDG LNEGCARYVVTUTLANDG LNEGCARYVVTUTLANDG LNEGCARYVVTUTLANDG LNEGCARYVVTUTLANDG LNEGCARYVVTUTLANDG SCALGAGEILLTSIDNG GVELGAGEILLTSIDNG GVELGAGEILLTSIDNG S60 	480 490 TITGPNTELLRQVAART	500 .   APP IC AP
COW7K4   HisA A. urogenitalis Q8FNZ7   HisA C. efficiens Q9FN74   HisA C. jejuni P9NMM5   FriA M. tuberculosis Alt562   HisA S. aurescens P16250   FriA S. turescens P16250   FriA S. topolicolor B4T986   HisA S. heidelberg B514P8   FriA S. typoimurium B4V386   FriA Streptomyces sp. Q9XCC7   HisA T. maritima Q8Z156   HisF T. maritima Q7SIB9   HisF T. thermophilus COW7K4   HisA A. urogenitalis Q8FN74   HisA C. efficiens Q9FN74   HisA C. efficiens Q9FN74   HisA S. cerevisiae P16250   FriA M. tuberculosis Alt562   HisA S. acrevisiae P16250   FriA M. tuberculosis Alt562   HisA S. topolicolor B4T986   FriA S. sviceus P16372   HisA S. treptomyces sp. Q9X0C6   HisA T. maritima	β5        410      420      430        AVGLDVRGT      -      -        AVDIAVRLENG      E      -        AVGLDVRGT      -      -        VULADVILENG      E      -        AVGLDVRGTT      -      -        VULADVRIDGE      E      -        AVGLDVRGTT      -      -        VULADVRIDBERG      T      -        VVALDAKVNDGE      -      -        VVALDAKVNDGE      -      -        VVALDAKVNDGE      -      -        VVALDAKVNDGE      -      -        VVASGGISSLEDIAALARUVPGVDSIJVEKAU      -      F        VVASGGISSLEDIAALARUVPGVDSIJVEKAU      -      -        VVASGGISSLEDIAALARUVPGVDSIJVEKAU      -      -        VVASGGISSLEDIAALARUVPGVDSIJVEKAU      -      -        VVASGGISSLEDIAALARUVPGVDSIJVEKAU      -      -        VVASGGISSLEDIAALARUVPGVDSIVEKAU      -      -        VVASGGISSLEDIAALARUVPGVDSIVEKAU      -      -        VVASGGISSLEDIAALARUVPGVDSIVEKAU      -      -	440      450	460 470 LNEAGCREYVYDUYTKDG LSSQCCSFYVYDUYTKDG USNCLKIILLCTUISKDG LEDAGCARYVYDUYTKDG LEDAGCARYVYDUYTKDG LEDAGCARYVYDUIKKO LINKEGCARYVYDIIAKDG NIEVGLKIVLCTUISKDG LNEGCARYVYDIIKKO LNEGCARYVYDIIKKO LNEGCARYVYDIG LNEGCARYVYDIG UNFELGAGEILLINSTDRDG GVELGAGEILLINSTDRDG GVELGAGEILLINSTDRDG 560 	udd      480    490      111000000    1110000000000000000000000000000000000	500 .   AP AP IC AP KIK RP IA RP IA RP IA VP
COW7K4   HisA A. urogenitalis Q8FNZ7   HisA C. efficiens Q9FNZ7   HisA C. efficiens Q9FNZ7   HisA C. jejuni P9NMM5   FriA M. tuberculosis AltS62   HisA P. aurescens P16250   FriA S. topelicolor B4T908   HisA S. heidelberg B514P8   PriA S. sylceus P10372   HisA S. typhimurium B4V386   FriA Streptomyces sp. Q9X0C6   HisF T. maritima Q8Z16   HisF T. maritima Q7SIB9   HisF T. thermophilus COW7K4   HisA A. urogenitalis Q8FNZ7   HisA C. jejuni P9NM5   PriA M. tuberculosis AltS62   HisA P. aurescens P16250   FriA M. tuberculosis AltS62   HisA P. aurescens P16250   FriA S. certificiens P16378   HisA S. topelicolor B4T9N8   HisA S. topelicolor B4T9N8   HisA S. treptomyces sp. Q9X0C7   HisA T. maritima Q9X0C7   HisA S. treptomyces sp. Q9X0C7   HisA T. maritima Q9X0C7   HisA T. maritima Q9X0C7   HisA T. maritima Q9X0C7   HisA T. maritima	β5        410      420      430        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        VIALDPTLIKED      -      -        AVGLDVRGTT      -      -        VVLALDVRTTDE      -      -        VVLALDVRTTDE      -      -        VVLALDVRTDE      -      -        VVLALDVRTT      -      -        VVLALDVRTDE      -      -        VVLALDVRTDE      -      -        VVLALDVRUDE      -      -        VVLALDVRUDE      -      -        VVAIDAKINGE      -      -        VIALDVRUDE      -      -        VIAIDAKINGE      -      -        VIAIDAKINGE      -      -        VIAIDAKINGE      -      -        VIAIDAKINGE      -      -        VIASGGIS      <	440      450	460 470 LNEAGCRYVYUUVTKDG LSSQCSFYVTUVTKDG VSNKGLKIILLCTUISKDG LSSQCSFYVTUVSKG LEDAGCARYVYTUVTKDG LEDAGCARYVYTUISKDG LIKYCHARVVTDIALOG LIKYCARYVVTDIALG LIKYCARYVVTDIALG LIKYCARYVVTDIALG LIKYCARSTLINSTDRG GVELGAGETLLNSTDRG GVELGAGETLLNSTDRG S560 	480    490      1110000000000000000000000000000000000	500 APP APP IC AP
COW7K4 HisA A. urogenitalis Q8PNZ7 HisA C. efficiens Q9PN74 HisA C. jejuni P9NRM5 FiRA M. tuberculosis Alt562 HisA P. aurescens P16250 PriA M. tuberculosis P16350 PriA S. coelicolor B4T9N8 HisA S. heidelberg B51428 PriA S. sviceus P10372 HisA S. typhimurium B43734 HisF P. aerophilum P33734 HisF T. maritima Q82116 HisF T. maritima Q9X0C6 HisF T. maritima Q7SIB9 HisF T. thermophilus COW7K4 HisA A. urogenitalis Q8PN27 HisA C. efficiens Q9X0C6 HisF T. maritima P16352 HisF T. thermophilus COW7K4 HisA A. urogenitalis Q8PN27 HisA C. fejuni P3734 HisF S. coervisiae P16250 PriA M. tuberculosis Alt562 HisA F. aurescens P16454 HisA S. highelberg B51468 PriA S. sviceus P10372 HisA S. typhimurium P10372 HisA S. typhimurium P10372 HisA S. typhimurium P33734 HisF F. maritima Q82116 HisF T. maritima	β5        410      420      430        AVGLDVRGTT      -      -        AVDIAUVRLENG      -      -        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        AVGLDVRGTT      -      -        VIALDPILKED      -      -        AVGLDVRGTT      -      -        VIALDDVRDBEG      -      -        TJADDRINGE      -      -        VVALDDRINGE      -      -        VIALDVRIVERGE      -      -	440 450 	460 470 LNEAGCRYVYUVYUVYUVYUG LNEAGCRYVYUVYUVYUG SSNCGLKIILCUUSKOG LEDAGCARYVYUVYUVYUG LIDAGCARYVYUVYUTKOG LIDAGCARYVYUVUTKOG LINEGCARYVVUVUTLANDVEG LINEGCARYVVUTUSKOG LINEGCARYVVUTUSKOG LINEGCARYVVUTUSKOG LINEGCARYVVUTUSKOG LINEGCARYVVUTUSKOG SSOO LINEGCARYVVUTUSKOG SSOO	480    490      111000000    1110000000000000000000000000000000000	500 .   APP APP IC IC IC IC IC IC IC IC IC IC

**Figure 6.** Alignment of the amino acid sequences of HisA, HisF, and relative orthologs reported in the PDB. Secondary structure elements are reported in orange ( $\alpha$ -helices), blue ( $\beta$ -strands), and fuchsia (turns).  $\alpha$ -helices and  $\beta$ -strands are also schematized with rectangles and arrows, respectively; the elongation event is depicted with the two halves of the proteins colored in grey and green.

### 3.2. hisB

While a huge amount of information regarding their early evolution is available for *hisA* and *hisF*, for the histidine biosynthetic genes *hisB* and *hisD*, only a few data are available, and they mainly concern the crystal structures of their gene products. These analyses allowed the possibility of a gene elongation event having occurred in the evolution of these genes to be suggested. In both cases, an internal sequence repeat was observed, presumably deriving from an in-tandem gene duplication of a starting module followed by the fusion of the derived copies.

of histidine biosynthetic pathway, the  $Mn^{2+}$ -dependent dehydration of imidazole-glycerol-phosphate to imidazole-acetyl-phosphate (Figure 2) [17,47]. IGPDs from fungi [48,49], plants [50,51], Archaea, and some Bacteria are monofunctional [52]. Other bacterial genomes harbor bifunctional genes, in which IGPD coding gene is fused to the histidinol-phosphate phosphatase one, the penultimate enzyme of histidine biosynthesis [17]. Previous structures of IGPDs available in the PDB, belonging to all the three cell lineages, show that IGPD is a homo 24-mer with, correspondingly, 24 active sites, each formed by residues from three adjacent subunits [53], and that it presents two manganese ions bound at each of the catalytic centers [52]. This structure and other biochemical data reveal that, unlike other enzymes in which  $Mn^{2+}$  can be exchanged with  $Mg^{2+}$  or  $Zn^{2+}$  with little effect on activity, IGPD has a peculiar requirement for manganese [47]; indeed, in the absence of  $Mn^{2+}$ , plant and fungal IGPDs are stable but inactive trimers [52]. All the available three-dimensional structures belong to monofunctional IGPDs; thus, it is not possible to know whether the bifunctional ones are able to form the 24-mer too. IGPD is an interesting candidate for structural studies because of its unusual chemical reaction, its aggregation properties, the metal dependence, the lack of sequence similarity to proteins of known structure, and the potential as an herbicide target [52].

Sinha et al. [52] analyzed the crystal structure of IGPD from the fungus *Filobasidiella neoformans*. This analysis revealed a new kind of structure in which an unusual structural motif is duplicated into a single compact domain. The IGPD polypeptide is composed of four  $\alpha$ -helices ( $\alpha 1-\alpha 4$ ), located between two four-stranded  $\beta$ -sheets ( $\beta 1-\beta 2-\beta 4-\beta 3$  and  $\beta 5-\beta 6-\beta 8-\beta 7$ ) (Figure 7A,B). This configuration possesses an internal repeat, in which the first  $\beta$ -sheet and two  $\alpha$ -helices ( $\beta 1-\beta 2-\beta 3-\alpha 1-\beta 4-\alpha 2$ ) have an identical topology to the second  $\beta$ -sheet and two  $\alpha$ -helices ( $\beta 5-\beta 6-\beta 7-\alpha 3-\beta 8-\alpha 4$ ). They observed that the structural repeat matches an internal sequence repeat; in particular, the amino acid sequences of the two halves of *F. neoformans* IGPD are 19% identical. Each half-domain includes a specific motif, Asx-Xaa-His-His-Xaa-Xaa-Glu [(D/N)XHHXXE], that is highly conserved in IGPD sequences. A subsequent experiment performed by Glynn et al. [54] showed that the structure of the monomer of *Arabidopsis thaliana* IGPD closely resembles that of the *F. neoformans* enzyme, with the two halves of the molecule related by a two-fold pseudo-symmetrical axis.

In Sinha et al. [52], a multiple-sequence alignment of IGPD from 56 organisms was performed. In the present work, 81 IGPD sequences from Archaea, 359 from Bacteria, and 95 from Eukarya were aligned, for a total of 535 IGPD sequences. The repetition of the (D/N)XHHXXE motif between each half-domain of IGPD was strongly conserved among all the analyzed sequences, and this motif also proved to be highly conserved between the IGPD of the different organisms. In particular, the HHXXE motif was present in all the analyzed sequences. On the other hand, the D/N residues appeared to be more variable; indeed, they were totally conserved in Eukarya, while in Bacteria and Archaea they were less conserved. Moreover, for those organisms with IGPD structures reported in the PDB (see Materials and Methods), the secondary structure was investigated, confirming the presence of the internal structural repetition (Figure 8). As proposed by Sinha et al. [52], and on the basis of these additional results, the repetition of sequence and structural elements strongly suggests a gene elongation event in the evolution of IGPD. The presence of this highly conserved motif in both the half-domains of all the analyzed IGPD sequences belonging to the three cell lineages suggests that the gene elongation event leading to the extant IGPDs very likely took place very early during evolution, before the divergence from LUCA.

As reported by Sinha et al. [52] and Glynn et al. [54], according to a topology search in the PDB, no other examples of proteins with folds like the IGPD have emerged. However, several proteins have subdomains with topologies identical to the IGPD half-domain; the presence of the IGPD half-domain motif in other proteins supports the hypothesis that it constitutes an initial folding unit, which evolved to produce the extant IGPD via gene elongation [52]. Thus, it is possible to hypothesize that the IGPD half-domain might represent a starter type that subsequently underwent different fates, namely a gene elongation event which led to the formation of the current complete IGPD or a fusion event with other genes, leading to the evolution of proteins possessing subdomains with the IGPD half-domain topology (Figure 7C).



**Figure 7.** (**A**,**B**) Three-dimensional structure of IGPD domain of *E. coli* HisB.  $\alpha$ -helices and  $\beta$ -strands are labeled. (**A**) Top view of IGPD; (**B**) side view. (**C**) Evolutionary model proposed to explain the origin and evolution of IGPD and proteins with subdomains with topologies identical to the IGPD half-domain.

A0A1P7ZJC8 A. castellanii 023346 A. thaliana P34047 A. thaliana P00022 C. neoformans P9WALS M. tuberculosis P586800 P. furiosus P54373 S. aureus P06633 S. cerevisiae	ID      20      30      40      50      60      70      80      90      100        MELLS      SSPAQLLRPKLSSRALLPPRNSIXAS      HPPPRFLVMNSQSQRR91SCASPFQINOFPATTARPIESAKIGKVMRUTTERNVSV        MELSSAALLAPKKIGFTDLIPPRNITVSSPSSSLPRFLMMESGSQRR91SCASA      SSSMAQLAPKKIGFTDLIPPRNITVSSPSSLPRFLMMESGSQRR91SCASA      MEXAMINITARTINETNUSV        MELSSAALLAPKKIGFTDLIPPRNITVSSPSSLPRFLMMESGSQRR91SCASA      MEXAMINITARTINETNUSV      MEXAMINITARTINETNUSV        MELSSAALLAPKKIGFTDLIPPRNITVSSPSSLPRFLMMESGSQRR91SCASA      MEXAMINITARTINETNUSV      MEXAMINITARTINETNUSV        MELSSAALLAPKKIGFTDLIPPRNITVSSPSSLPRFLMMESGSQRR91SCASA      MEXAMINITARTINETNUSV      MEXAMINITARTINETNUSV        MELSSAALLAPKKIGFTDLIPPRNITVSSPSSLPRFLMMESGSQRR91SCASA      MEXAMINITARTINETNUSV      MEXAMINITARTINETNUSV        MELSSAALLAPKKIGFTDLIPPRNITVSSPSSLPRFLMMESGSQRR91CCASA      MEXAMINITARTINETNUSV      MEXAMINITARTINETNUSV
A0A1P7ZJC8 A. castellanii 023346 A. thaliana P34047 A. thaliana P90C021C. neoformans P9WRL9 M. tuberculosis P58880 P. furiosus P64373 S. aureus P06633 S. cerevisiae	B3      a1      B4      a2        110      120      130      140      150      160      170      180      190      200        RLSLDG      TGYBDNK      TGTPLDBAGALANGRFDLYLKACADLAYDDHRYBDOATLAGAALUGALEBEKGEKREGG      180      190      200        KINLDG      TGTPLDBAGALANGRFDLYLKACADLAYDDHRYBDOATLAGAALUGALEBEKGEKREGG      TGTPLDBAGALANGRFDLYLKACADLAYDDHRYBDOATLAGAALUGALEBEKGEKREGG        KINLDG      TGTPLDBAGALANGLANGLANGLAYDVENDAGUTANDDHRYBDOATLAGAALUGALEBEKGEKREGG      TGTPLDBAGALANGLANGLANGLANGLAYDVENDAGUTANDHRYBDAALAGAALUGAEKGANLGEBEKGEKREGG        KINLDG      TGYQUND      TGYPPYDBALTANGUGALUGAUGUCAUDDHRYBDOALLAGATALUGALGEKKGEKREGG        ELIGKK      TGQUAND      TGYPPYDBALTARAFYI (AGMAGUTANDYEL TAHTIGAGUALGEKKEEKREKREGG        SIGDQ      TGYGUAUMLATARYI (AGMAGUTANTYDUALAGUALGEKKEEKREKREGG      TGYGUAUMLATARYI (AGMAGUTANTYDUALAGUALGEKKEEKREKREGG        SIGDQ      SPSHIN      TGYGUAUMLATAFYI (GAUGUKAUTANTYDUALAGUALGEKKEEKREKREGG      SIGDQ      SPSHIN      TGYGUAUMLATAFYI (GAUGUKAUTANTYDUALAGUAUGUKAUKEKREGGE        SIGDQ      SPSHIN      TGYGUAUMLATTYTIGGUCAUMLAGUKAUGUKAUGUKAUGUKAUGUKAUGUKAUGUKAUGUK
AOAIP7ZJC8   A. castellanii O23346   A. thaliana P34047   A. thaliana P90022 [ C. neoformans P9WRL9   M. tuberculosis P58880   P. furiosus P64373   S. aureus P06633   S. cerevisiae	B5    B6    B7    C3    B8    C4      AYAP LDB SLARAVVD SSRPEAVUD    230    240    250    260    270    280    290    300      AYAP LDB SLARAVVD SSRPEAVUD
A0A1P72JC8   A. castellanii 023346   A. thaliana P34047   A. thaliana P0C022   C. neoformans P9WRL9   M. tuberculosis P56880   P. furiosus P64373   S. aureus P06533   S. cerevisiae	310 

**Figure 8.** Alignment of the amino acid sequences of IGPDs reported in the PDB. Secondary structure elements are reported in orange ( $\alpha$ -helices), blue ( $\beta$ -strands), and fuchsia (turns). The (D/N)XHHXXE motifs are underlined.  $\alpha$ -helices and  $\beta$ -strands are also schematized with rectangles and arrows, respectively; the elongation event is depicted with the two halves of the proteins colored in grey and green.

#### 3.3. hisD

The *hisD* gene encodes the enzyme L-histidinol dehydrogenase (HDH), which has a bifunctional activity catalyzing the last two steps in histidine biosynthetic pathway (Figure 2). This enzyme converts L-histidinol (HOL) to L-histidine through a L-histidinaldehyde (HAL) intermediate. The bifunctional enzymatic activity of the *hisD* gene product seems to be a universal property [17], and the sequence of HisD has been well conserved during evolution from bacteria to fungi and plants [25,55]. HDH is a homodimeric enzyme containing two metal binding sites per enzyme dimer, which normally bind  $Zn^{2+}$ , but which may also bind  $Mn^{2+}$  and still retain enzyme function [56].

Barbosa et al. [18], through the analysis of the crystal structure of *E. coli* HisD, observed a possible early event of gene elongation. The HisD monomer consists of a globule and a tail. The globule is made of two larger domains (1 and 2) and the extending L-shaped tail of two smaller domains (3 and 4) (Figure 9A). Domain 1 and Domain 2 display an  $\alpha/\beta/\alpha$  topology and they both contain a core with an incomplete Rossmann fold (with only five strand–helix hairpins, instead of six) (Figure 9B,C). When the two domains are superposed, their cores appear to be very similar, with only 11% amino acid sequence identity, but with a similarity value of about 41%. The resemblance of the three-dimensional structure of the two cores, the amino acid sequence similarity, and their tandem occurrence with only two residues between the end of the first and the start of the second suggested a gene duplication event followed by a fusion, which must have occurred early in the evolution of this enzyme; subsequently, the two moieties diverged and only one of the two domains retained the ability to bind NAD<sup>+</sup>, whereas the other evolved to bind  $Zn^{2+}$  and the substrate [18]. Indeed, the low amino acid sequence identity could be the result of an accumulation of mutations during time, suggesting that this gene elongation event presumably took place very early during evolution. In the present work, a possible evolutionary model for the evolution of HisD is proposed: starting from a gene encoding one of the two HDH cores, this could have undergone a gene elongation event. The resulting gene composed of the two cores then fused with two other regions, which formed the N-terminal and the C-terminal parts of the protein (Figure 9D).

To the best of our knowledge, only three publications on crystal structures of HDH enzymes are available: *E. coli* HDH (*Ec*HDH) [18], *Brucella suis* HDH (*Bs*HDH) [57], and *Medicago truncatula* HDH (*Mt*HDH) [58]. As reported by Ruszkowski and Dauter [58], the sequences of plant HDHs are highly similar, and significant similarities have also been highlighted between the plant and bacteria kingdoms. Furthermore, structural analyses have shown that *Mt*HDH is similar to the two aforementioned bacterial HDH.

For those organisms with histidinol dehydrogenase structures available in the PDB (see Materials and Methods), we investigated the secondary and the three-dimensional structures. This analysis highlighted the conservation of the domain structures of the HisD monomer among different phylogenetic lineages (Figure 10). It is possible that, also in this case, the gene elongation event involving the core, as well as the fusion of genes encoding the other moieties, occurred before the appearance of LUCA.



**Figure 9.** (**A**) Side view of the three-dimensional structure of the *E. coli* HisD monomer. The different domains are highlighted with different colors: Domain 1 in blue, Domain 2 in red, Domain 3 in green, and Domain 4 in violet. (**B**) Top view of *E. coli* HisD monomer: the elements of the two HDH cores are colored in blue (Domain 1) and red (Domain 2).  $\alpha$ -helices and  $\beta$ -strands of the cores are labeled. (**C**) Topology diagram. Secondary structure elements are colored in yellow ( $\alpha$ -helices), orange (3<sub>10</sub> helices), and light blue ( $\beta$ -strands). The four domains are circled with the corresponding colors used in (**A**). The cores of the two Domains are highlighted with the secondary structure element outlines colored in blue (in Domain 1) and red (in Domain 2). Diagram adapted from Barbosa et al. [18] and Ruszkowski and Dauter [58]. (**D**) Evolutionary model proposed to explain the origin and evolution of HDH.



**Figure 10.** Alignment of the amino acid sequences of HDHs reported in the PDB. Secondary structure elements are reported in orange ( $\alpha$ -helices), blue ( $\beta$ -strands), and fuchsia (turns).  $\alpha$ -helices and  $\beta$ -strands of the two cores of the proteins are also schematized with rectangles and arrows, respectively; the elongation event is depicted with the two cores of the proteins colored in grey and green.

### 4. Conclusions

As described above, the histidine biosynthetic pathway represents an excellent model for the study of molecular mechanisms that occurred during the early evolution of biosynthetic pathways. Indeed, it is known that at least three gene rearrangements shaped this route during time, i.e., gene duplication, fusion, and elongation. Concerning the third mechanism, the most documented case of gene elongation involves *hisA* and *hisF*; many studies have been performed in the last 20 years with the aim to better understand the evolutionary processes that led to the formation of the extant genes. The hypothesis for their evolution, starting from a gene half the length of the current genes that underwent a tandem event of "duplication and fusion", is confirmed by the high sequence similarity between the two halves of the proteins, and by structural and biochemical studies. In this work, we suggest an implementation of the model, in the sense that the two extant genes could be the result of a cascade of gene elongations or gene elongation/domain shuffling events starting from an ancestor gene coding for just one ( $\beta/\alpha$ ) domain. In our opinion, this ancestral mini gene might represent one of the starter types, according to the definition proposed by Lazcano and Miller [44].

Similar evolutionary dynamics seem to also have occurred in the case of both *hisB* and *hisD*. Although a smaller number of studies have been conducted on these genes, structural analyses revealed the possibility of early elongation events of ancestor genes resulting in the repetition of modules, which might have been followed by the fusion of additional genes coding for other moieties (in the case of *hisD*).

Furthermore, since the structures of the four proteins are highly conserved in the prokaryotic and eukaryotic histidine-synthesizing (micro)organisms, it is quite possible that these gene elongation events occurred very early in the evolution of *hisA*/*hisF*, *hisB*, and *hisD* and predated the appearance of

LUCA (Figure 11), which is in agreement with the idea that histidine biosynthesis is a very ancient metabolic route.



**Figure 11.** Schematic representation of the occurrence of the gene elongation events in the evolution of *his* genes throughout phylogeny.

In conclusion, gene elongation events seem to have played a crucial role in the evolution of the histidine biosynthetic pathway. It is not clear whether other histidine biosynthetic genes might be the result of other gene elongation events; indeed, the analysis of the amino acid sequence of the other enzymes involved in histidine biosynthesis from archaeal and/or bacterial strains did not reveal a degree of sequence similarity sufficient to suggest the presence of internal repetitions. This was in agreement with the analysis of the bibliographic data, which, to the best of our knowledge, did not report the presence of such repetitions on the basis of the crystal structures analyses. However, by assuming that histidine biosynthesis is a very ancient pathway, it could be possible that very old gene elongation events (that can be disclosed by analyzing the structure or the degree of sequence similarity) might be obscured by the long divergence time and/or by the lower structural constraints in other His proteins. Even though this issue is beyond the scope of this work, it is also possible that other *his* genes could be the result of duplication events of genes embedded with a broad range of substrates, i.e., also involved in other metabolic pathways, and might have been recruited into the histidine biosynthetic route, in accordance with the patchwork hypothesis [42].

**Supplementary Materials:** The following are available online at http://www.mdpi.com/2076-2607/8/5/732/s1, Supplementary Material S1: files of the predicted three-dimensional structures of *E. coli* HisA, HisF, HisB (IGPD) and HisD (HDH) in PDB format. Supplementary Material S2: alignment of the 535 IGPD sequences. Accession numbers of all the considered organisms are reported.

Author Contributions: Conceptualization, S.D.D. and R.F.; Formal analysis, S.D.D.; Investigation, S.D.D. and S.C.; Methodology, S.D.D. and R.F.; Project administration, R.F.; Supervision, R.F.; Validation, S.D.D. and A.V.; Visualization, S.D.D. and L.M.C.; Writing—original draft, S.D.D. and R.F.; Writing—review & editing, S.D.D., S.C., A.V., L.M.C. and R.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

### References

- 1. McLachlan, A.D. Gene duplication and the origin of repetitive protein structures. *Cold Spring Harb. Symp. Quant. Biol.* **1987**, *52*, 411–420. [CrossRef] [PubMed]
- 2. Brilli, M.; Fani, R. The origin and evolution of eucaryal HIS7 genes: From metabolon to bifunctional proteins? *Gene* **2004**, *339*, 149–160. [CrossRef] [PubMed]
- 3. Fani, R.; Fondi, M. Origin and evolution of metabolic pathways. *Phys. Life Rev.* 2009, *6*, 23–52. [CrossRef] [PubMed]

- 4. Zhang, J. Evolution by gene duplication: An update. Trends Ecol. Evol. 2003, 18, 292–298. [CrossRef]
- 5. Hughes, A.L. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. B* **1994**, 256, 119–124.
- 6. Ohno, S. Evolution by Gene Duplication; Springer Science & Business Media: New York, NY, USA, 1970.
- 7. Yanai, I.; Wolf, Y.I.; Koonin, E.V. Evolution of gene fusions: Horizontal transfer versus independent events. *Genome Biol.* **2002**, *3*, research0024-1.
- 8. Adman, E.T.; Sieker, L.C.; Jensen, L.H. Structure of a bacterial ferredoxin. J. Biol. Chem. 1973, 248, 3987–3996.
- 9. Tang, J.; James, M.N.G.; Hsu, I.N.; Jenkins, J.A.; Blundell, T.L. Structural evidence for gene duplication in the evolution of the acid proteases. *Nature* **1978**, *271*, 618–621. [CrossRef]
- 10. Nyunoya, H.; Lusty, C.J. The *carB* gene of *Escherichia coli*: A duplicated gene coding for the large subunit of carbamoyl-phosphate synthetase. *PNAS* **1983**, *80*, 4629–4633. [CrossRef]
- Rubin, R.A.; Levy, S.B.; Heinrikson, R.L.; Kézdy, F.J. Gene duplication in the evolution of the two complementing domains of Gram-negative bacterial tetracycline efflux proteins. *Gene* 1990, *87*, 7–13. [CrossRef]
- 12. Gupta, R.S.; Singh, B. Cloning of the HSP70 gene from *Halobacterium marismortui*: Relatedness of archaebacterial HSP70 to its eubacterial homologs and a model for the evolution of the HSP70 gene. *J. Bacteriol.* **1992**, *174*, 4594–4605. [CrossRef]
- Reizer, J.; Saier, M.H. Modular multidomain phosphoryl transfer proteins of bacteria. *Curr. Opin. Struc. Biol.* 1997, 7, 407–415. [CrossRef]
- 14. 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**, *38*, 489–495. [CrossRef] [PubMed]
- 15. Fani, R.; Brilli, M.; Fondi, M.; Lió, P. The role of gene fusions in the evolution of metabolic pathways: The histidine biosynthesis case. *BMC Evol. Biol.* **2007**, *7*, S4. [CrossRef] [PubMed]
- 16. Kulis-Horn, R.K.; Persicke, M.; Kalinowski, J. Histidine biosynthesis, its regulation and biotechnological application in *Corynebacterium glutamicum*. *Microb. Biotechnol.* **2014**, *7*, 5–25. [CrossRef] [PubMed]
- 17. Alifano, P.; Fani, R.; Liò, P.; Lazcano, A.; Bazzicalupo, M.; Carlomagno, M.S.; Bruni, C.B. Histidine biosynthetic pathway and genes: Structure, regulation, and evolution. *Microbiol. Rev.* **1996**, *60*, 44–69. [CrossRef]
- Barbosa, J.A.; Sivaraman, J.; Li, Y.; Larocque, R.; Matte, A.; Schrag, J.D.; Cygler, M. Mechanism of action and NAD<sup>+</sup>-binding mode revealed by the crystal structure of L-histidinol dehydrogenase. *PNAS* 2002, *99*, 1859–1864. [CrossRef]
- 19. Winkler, M.E.; Ramos-Montañez, S.M.I.R.L.A. Biosynthesis of histidine. EcoSal Plus 2009, 3. [CrossRef]
- 20. Shen, C.; Lazcano, A.; Oró, J. The enhancement activities of histidyl-histidine in some prebiotic reactions. *J. Mol. Evol.* **1990**, *31*, 445–452. [CrossRef]
- 21. Shen, C.; Mills, T.; Oró, J. Prebiotic synthesis of histidyl-histidine. J. Mol. Evol. 1990, 31, 175–179. [CrossRef]
- 22. White, H.B. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **1976**, *7*, 101–104. [CrossRef] [PubMed]
- 23. Fondi, M.; Emiliani, G.; Liò, P.; Gribaldo, S.; Fani, R. The evolution of histidine biosynthesis in archaea: Insights into the *his* genes structure and organization in LUCA. *J. Mol. Evol.* **2009**, *69*, 512. [CrossRef] [PubMed]
- 24. Horowitz, N.H. The Evolution of Biochemical Syntheses—Retrospect and Prospect. In *Evolving Genes and Proteins;* Academic Press: Cambridge, MA, USA, 1965; pp. 15–23.
- 25. Fani, R.; Liò, P.; Lazcano, A. Molecular evolution of the histidine biosynthetic pathway. J. Mol. Evol. 1995, 41, 760–774. [CrossRef] [PubMed]
- 26. Kelley, L.A.; Mezulis, S.; Yates, C.M.; Wass, M.N.; Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **2015**, *10*, 845. [CrossRef]
- 27. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612. [CrossRef] [PubMed]
- 28. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef]
- 29. UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019, 47, D506–D515. [CrossRef]

- 30. Hall, T.A. BioEdit: A User-friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.
- Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTALW: Improving the sensitivity of progressive multiple sequence through weighing, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994, 22, 4673–4680. [CrossRef]
- 32. Lang, D.; Thoma, R.; Henn-Sax, M.; Sterner, R.; Wilmanns, M. Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science* **2000**, *289*, 1546–1550. [CrossRef]
- Jürgens, C.; Strom, A.; Wegener, D.; Hettwer, S.; Wilmanns, M.; Sterner, R. Directed evolution of a (βα)<sub>8</sub>-barrel enzyme to catalyze related reactions in two different metabolic pathways. *PNAS* 2000, *97*, 9925–9930. [CrossRef]
- 34. Vázquez-Salazar, A.; Becerra, A.; Lazcano, A. Evolutionary convergence in the biosyntheses of the imidazole moieties of histidine and purines. *PLoS ONE* **2018**, *13*, e0196349. [CrossRef]
- 35. Gerlt, J.A.; Babbitt, P.C. Barrels in pieces? *Nat. Struct. Biol.* 2001, *8*, 5–7. [CrossRef]
- 36. Horowitz, N.H. On the evolution of biochemical syntheses. *PNAS* **1945**, *31*, 153. [CrossRef]
- Thoma, R.; Obmolova, G.; Lang, D.A.; Schwander, M.; Jenö, P.; Sterner, R.; Wilmanns, M. Efficient expression, purification and crystallisation of two hyperthermostable enzymes of histidine biosynthesis. *FEBS Lett.* 1999, 454, 1–6. [CrossRef]
- 38. Gerlt, J.A. New wine from old barrels. Nat. Struct. Biol. 2000, 7, 171–173. [CrossRef] [PubMed]
- Reardon, D.; Farber, G.K. The structure and evolution of α/β barrel proteins. *FASEB J.* **1995**, *9*, 497–503. [CrossRef] [PubMed]
- 40. Petsko, G.A. Design by necessity. *Nature* 2000, 403, 606–607. [CrossRef] [PubMed]
- 41. Yčas, M. On earlier states of the biochemical system. J. Theor. Biol. 1974, 44, 145–160. [CrossRef]
- 42. Jensen, R.A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **1976**, *30*, 409–425. [CrossRef]
- 43. Barona-Gómez, F.; Hodgson, D.A. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* **2003**, *4*, 296–300. [CrossRef] [PubMed]
- 44. Lazcano, A.; Miller, S.L. How long did it take for life to begin and evolve to cyanobacteria? *J. Mol. Evol.* **1994**, 39, 546–554. [CrossRef] [PubMed]
- 45. Woese, C.R. The universal ancestor. Proc. Natl. Acad. Sci. USA 1998, 95, 6854–6859. [CrossRef] [PubMed]
- 46. Li, W.H. Evolution by Gene Duplication and Domain Shuffling. In *Molecular Evolution*; Sinauer Associates Inc.: Sunderland, MA, USA, 1997.
- 47. Bisson, C.; Britton, K.L.; Sedelnikova, S.E.; Rodgers, H.F.; Eadsforth, T.C.; Viner, R.C.; Hawkes, T.R.; Baker, P.J.; Rice, D.W. Crystal Structures Reveal that the Reaction Mechanism of Imidazoleglycerol-Phosphate Dehydratase Is Controlled by Switching Mn(II) Coordination. *Structure* **2015**, *23*, 1236–1245. [CrossRef]
- Parker, A.R.; Moore, T.D.; Edman, J.C.; Schwab, J.M.; Davisson, V.J. Cloning, sequence analysis and expression of the gene encoding imidazole glycerol phosphate dehydratase in *Cryptococcus neoformans*. *Gene* 1994, 145, 135–138. [CrossRef]
- 49. Hawkes, T.R.; Thomas, P.G.; Edwards, L.S.; Rayner, S.J.; Wilkinson, K.W.; Rice, D.W. Purification and characterization of the imidazoleglycerol-phosphate dehydratase of *Saccharomyces cerevisiae* from recombinant *Escherichia coli*. *Biochem. J.* **1995**, *306*, 385–397. [CrossRef]
- Mano, J.; Hatano, M.; Koizumi, S.; Tada, S.; Hashimoto, M.; Scheidegger, A. Purification and properties of a monofunctional imidazoleglycerol-phosphate dehydratase from wheat. *Plant Physiol.* 1993, 103, 733–739. [CrossRef]
- Tada, S.; Volrath, S.; Guyer, D.; Scheidegger, A.; Ryals, J.; Ohta, D.; Ward, E. Isolation and characterization of cDNAs encoding imidazoleglycerolphosphate dehydratase from *Arabidopsis thaliana*. *Plant Physiol.* **1994**, 105, 579–583. [CrossRef]
- 52. Sinha, S.C.; Chaudhuri, B.N.; Burgner, J.W.; Yakovleva, G.; Davisson, V.J.; Smith, J.L. Crystal structure of imidazole glycerol-phosphate dehydratase: Duplication of an unusual fold. *J. Biol. Chem.* **2004**, 279, 15491–15498. [CrossRef]
- Agapova, Y.K.; Timofeev, V.I.; Komolov, A.S. Molecular Dynamics Study of Triazole Derivative Binding to the Active Site of Imidazole Glycerol Phosphate Dehydratase from *Mycobacterium tuberculosis*. *Crystallogr. Rep.* 2019, 64, 608–610. [CrossRef]

- 54. Glynn, S.E.; Baker, P.J.; Sedelnikova, S.E.; Davies, C.L.; Eadsforth, T.C.; Levy, C.W.; Rodgers, H.F.; Blackburn, G.M.; Hawkes, T.R.; Viner, R.; et al. Structure and mechanism of imidazoleglycerol-phosphate dehydratase. *Structure* **2005**, *13*, 1809–1817. [CrossRef] [PubMed]
- 55. Nagai, A.; Ward, E.; Beck, J.; Tada, S.; Chang, J.Y.; Scheidegger, A.; Ryals, J. Structural and functional conservation of histidinol dehydrogenase between plants and microbes. *PNAS* **1991**, *88*, 4133–4137. [CrossRef] [PubMed]
- 56. Grubmeyer, C.; Skiadopoulos, M.; Senior, A.E. L-histidinol dehydrogenase, a Zn<sup>2+</sup>-metalloenzyme. *Arch. Biochem. Biophys.* **1989**, 272, 311–317. [CrossRef]
- 57. D'ambrosio, K.; Lopez, M.; Dathan, N.A.; Ouahrani-Bettache, S.; Köhler, S.; Ascione, G.; Monti, S.M.; Winum, J.Y.; De Simone, G. Structural basis for the rational design of new anti-Brucella agents: The crystal structure of the C366S mutant of L-histidinol dehydrogenase from *Brucella suis*. *Biochimie* 2014, 97, 114–120. [CrossRef]
- Ruszkowski, M.; Dauter, Z. Structures of *Medicago truncatula* L-histidinol dehydrogenase show rearrangements required for NAD<sup>+</sup> binding and the cofactor positioned to accept a hydride. *Sci. Rep.* 2017, 7, 1–13. [CrossRef] [PubMed]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).