Evolution of Substrate Specificity in a Recipient's Enzyme Following Horizontal Gene Transfer

Lianet Noda-García,‡,1 Aldo R. Camacho-Zarco,§,1,2 Sofía Medina-Ruíz,¶,1,2 Paul Gaytán,2 Mauricio Carrillo-Tripp,3 Vilmos Fülöp,4 and Francisco Barona-Gómez*,1

[1]Evolution of Metabolic Diversity Laboratory, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), Cinvestav-IPN, Irapuato, México

[2]Instituto de Biotecnología, Universidad Nacional Autónoma de México (UNAM), Cuernavaca, México

[3]Biomolecular Diversity Laboratory, Laboratorio Nacional de Genómica para la Biodiversidad (Langebio), Cinvestav-IPN, Irapuato, México

[4]School of Life Sciences, University of Warwick, Coventry, United Kingdom

‡Present address: Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel

§Present address: Department for NMR-based Structural Biology, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

¶Present address: Department of Molecular & Cell Biology, University of California, Berkeley, CA

***Corresponding author:** E-mail: fbarona@langebio.cinvestav.mx.

Associate editor: Jianzhi Zhang

Abstract

Despite the prominent role of horizontal gene transfer (HGT) in shaping bacterial metabolism, little is known about the impact of HGT on the evolution of enzyme function. Specifically, what is the influence of a recently acquired gene on the function of an existing gene? For example, certain members of the genus *Corynebacterium* have horizontally acquired a whole L-tryptophan biosynthetic operon, whereas in certain closely related actinobacteria, for example, *Mycobacterium*, the *trpF* gene is missing. In *Mycobacterium*, the function of the *trpF* gene is performed by a dual-substrate $(\beta\alpha)_8$ phosphoribosyl isomerase (*priA* gene) also involved in L-histidine (*hisA* gene) biosynthesis. We investigated the effect of a HGT-acquired TrpF enzyme upon PriA's substrate specificity in *Corynebacterium* through comparative genomics and phylogenetic reconstructions. After comprehensive in vivo and enzyme kinetic analyses of selected PriA homologs, a novel $(\beta\alpha)_8$ isomerase subfamily with a specialized function in L-histidine biosynthesis, termed subHisA, was confirmed. X-ray crystallography was used to reveal active-site mutations in subHisA important for narrowing of substrate specificity, which when mutated to the naturally occurring amino acid in PriA led to gain of function. Moreover, in silico molecular dynamic analyses demonstrated that the narrowing of substrate specificity of subHisA is concomitant with loss of ancestral protein conformational states. Our results show the importance of HGT in shaping enzyme evolution and metabolism.

Key words: **enzyme evolution, horizontal gene transfer, amino acid metabolism.**

Introduction

The core view of enzyme evolution is that gene duplication of multispecific enzymes, followed by narrowing of substrate specificity, is the primary mechanism by which novel enzyme families have evolved (Ohno 1970; Jensen 1976; Piatigorsky 2007). Gene duplication as a driving force in adaptation seems to be more frequent in eukaryotes than in prokaryotes (Dittmar and Liberles 2010). In prokaryotes, horizontal gene transfer (HGT) has been proposed as the primary mechanism for the expansion of extant protein families (Lerat et al. 2005; Pal et al. 2005; Treangen and Rocha 2011). Despite these observations, studies investigating the impact of HGT upon the relationship between the horizontally acquired enzymes and the assembly of prokaryotic metabolic pathways are scarce. The few available examples are limited to in silico evolutionary analyses that remain uninvestigated experimentally but suggest that unique evolutionary mechanisms may operate when HGT takes place (Pal et al. 2005; Klassen 2009).

We investigated the effect of HGT upon enzyme evolution using as model L-tryptophan and L-histidine biosynthesis within the ancestral *Actinobacteria* phylum. Two late-diverging actinobacteria, *Corynebacterium glutamicum* and *Corynebacterium diphtheriae*, have acquired by HGT a *w*hole-*p*athway *t*ryptophan *o*peron (WPTO). Previous comprehensive phylogenetic and gene organization analyses of this WPTO demonstrated that this metabolic pathway was acquired en bloc from a member of *Gammaproteobacteria* (Xie et al. 2003, 2004). In this WPTO the *trpF* gene, encoding a *N'*-(5'-phosphoribosyl)anthranilate (PRA) isomerase is fused with the pathways' downstream gene *trpC* (Indole-3-glycerol-phosphate synthase; fig. 1), a distinctive feature of *Gammaproteobacteria*. Moreover, the acquisition of the WPTO was hypothesized to prompt loss of the original

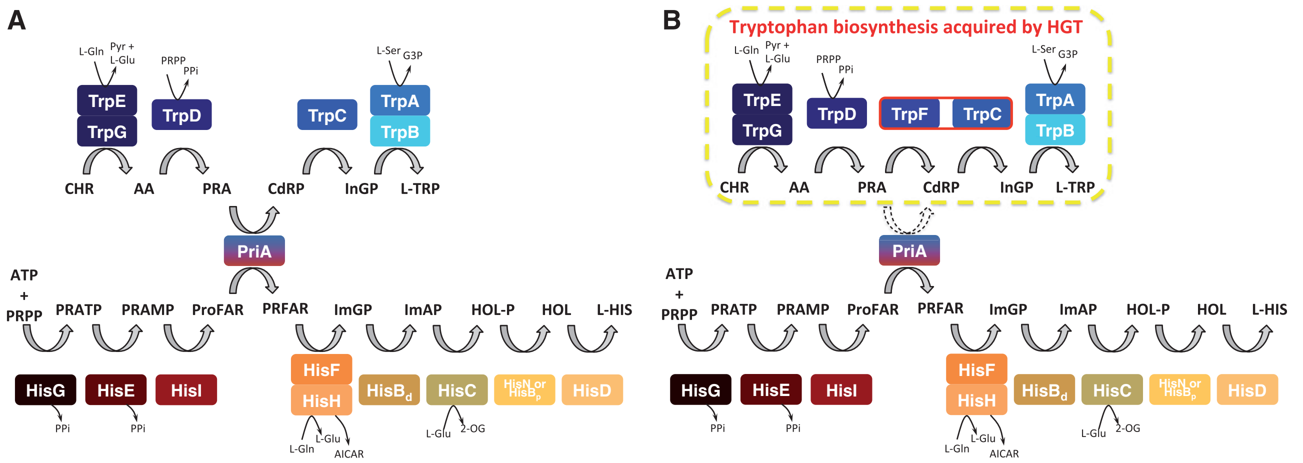


Fig. 1. L-histidine and L-tryptophan biosynthesis in *Actinobacteria*. (*A*) Convergent pathways, as found in *Streptomyces coelicolor* and *Mycobacterium tuberculosis*. (*B*) Independent pathways, as found in *Corynebacterium diphtheriae* and *C. glutamicum*. The enzymes for L-histidine and L-tryptophan biosynthesis are shown. Phosphoribosyl isomerase A (PriA), at which these pathways converge in (*A*), is shown between these pathways. Names and details of the pathways' intermediaries and enzymes are provided as supplementary table S1, Supplementary Material online.

corynebacterial *trp* genes following a homologous gene displacement that rendered synteny at this locus almost impossible to recognize (Xie et al. 2003, 2004).

Corynebacterium species are closely related to *Streptomyces coelicolor* and *Mycobacterium tuberculosis*, where L-histidine and L-tryptophan biosynthesis have been shown to converge (fig. 1) (Barona-Gomez and Hodgson 2003; Kuper et al. 2005). *Streptomyces coelicolor* and *M. tuberculosis* lack a *trpF* gene and the *his* and *trp* genes seem to cluster (Barona-Gomez and Hodgson 2003). The function encoded by the missing *trpF* is compensated by a dual-substrate $(\beta\alpha)_8$-barrel phosphoribosyl isomerase, encoded by the *priA* gene, a close homolog (~50% ID) of the *hisA* gene. Thus, the product of *priA* participates in the biosynthesis of both L-tryptophan and L-histidine [HisA, *N*′-[(5′-phosphoribosyl) formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) isomerase].

Recent biochemical and biophysical analyses demonstrate that the dual-substrate specificity of PriA seems to have evolved by means of active site conformational diversity. The residues located at flexible β to α loops 1, 5, and 6 mediate the metamorphosis of PriA's highly constrained active site, allowing the same cavity to adopt two different architectures specific for each activity (Wright et al. 2008; Noda-Garcia et al. 2010; Due et al. 2011).

The two contrasting biosynthetic scenarios described earlier, implying different evolutionary hypothesis, are illustrated in figure 1. We utilized comparative genomics, phylogenetic reconstructions, Michaelis Menten enzyme kinetics, site-directed mutagenesis, and structural characterization to discriminate between these two evolutionary hypotheses. Five selected PriA isomerases were comprehensively functionally characterized and classified according to their substrate specificities and metabolic pathway contributions. We found that following HGT, narrowing of substrate specificity occurred in a gene-duplication independent fashion, involving analogous rather than homologous enzymes. This enzyme specialization process was found to involve acquisition of conserved mutations surrounding the active site. Moreover, molecular dynamic simulations showed the role of protein conformational diversity, independent of an induced-fit mechanism, on the evolution of enzyme substrate specificity. Thus, we provide the first evidence for the evolution of substrate specificity following HGT in a recipient's enzyme.

Results

To investigate the relationship between HGT and the evolution of substrate specificity, we used comparative genomics of the *his* and *trp* genes together with phylogenetic reconstructions of PriA homologs from *Mycobacterium* and *Corynebacterium* species. These analyses revealed that members of the genus *Mycobacterium*, as well as a certain subclade of the genus *Corynebacterium*, lack a WPTO and encode the *his* and *trp* genes (*hisD*, *hisC*, *hisB*, *hisH*, *priA*, *hisF*, *hisI*, *trpE*, *trpC*, *trpB*, and *trpA*) within a single locus smaller than 15 kb. We refer to this as the *his-trp* gene cluster. Remarkably, *his* and *trp* gene fusions were found in *C. kroppenstedtii*, rendering a HisF-HisI-TrpE polypeptide, indicative of the full integration of L-histidine and L-tryptophan biosynthesis (fig. 2, blue clades). In contrast, and as previously reported (Xie et al. 2003, 2004), we confirmed the existence of a subclade of the genus *Corynebacterium* with an HGT-acquired WPTO that correlates with deterioration of the *his-trp* gene cluster (fig. 2). The deterioration of this cluster includes loss of *trpB* and *trpA* genes, and mutation of *trpC*, leaving exclusively *his* genes. As a consequence, the *his* and *trp* genes in these organisms are separated by at least 800 kb.

On the basis of this observation, we asked whether HGT could shape enzyme substrate specificity. Specifically, given the existence of a *trpF* gene encoding for redundant PRA isomerase activity in certain *Corynebacterium* species, narrowing of substrate specificity of PriA was hypothesized.

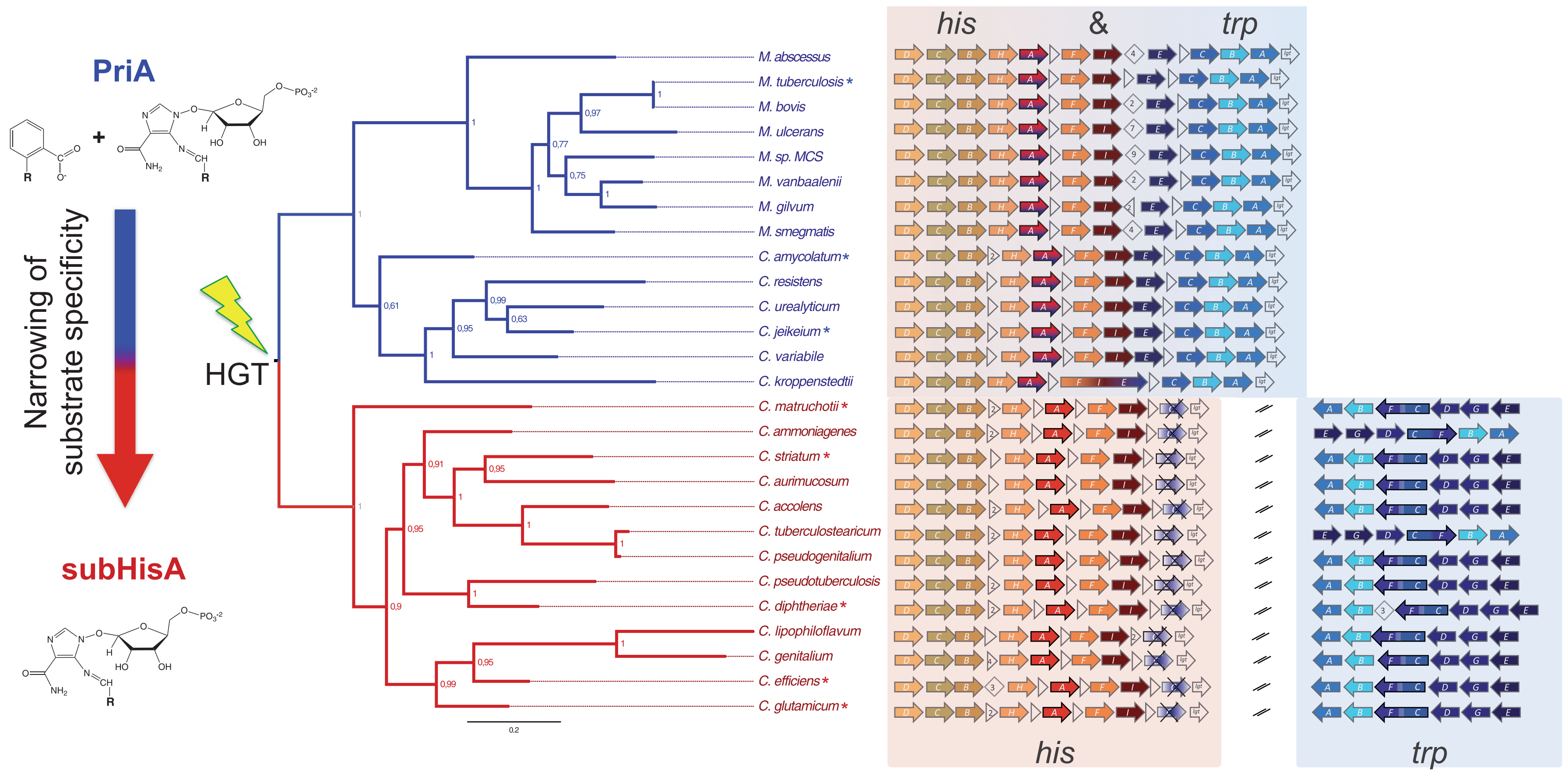


Fig. 2. *his* and *trp* comparative genomics and *priA* phylogeny. (*A*) PriA-based phylogeny. Numbers at nodes are the approximate likelihood ratio test supporting each branch. The branches with *priA* genes, as these sequences co-occur with lack of a *trpF* gene, are shown in blue. The branches with *subhisA* genes, as they co-occur with HGT-acquired WPTO *trpFs*, are shown in red. Enzymes selected for further functional analyses are highlighted with an asterisk. (*B*) Genomic context analysis of *his* and *trp* genes. The *lgt* gene, shown in white, was adopted as a genetic marker to define conservation of gene context. Genes of unknown function, or unrelated to L-His or L-Trp biosynthesis, are marked with white triangles (same directionality) or diamonds (both directionalities). The numbers within these triangles and diamonds indicate how many predicted genes are in this category. When this is higher than 15, disruption of the *his* and *trp* gene cluster is marked with two diagonal black lines. Gene's nomenclature and colors are used as in supplementary table S1 (Supplementary Material online) and figure 1.

Biochemical Analysis of Selected PriA Homologs

In vivo characterization of selected enzymes was conducted by testing the ability of any given PriA homolog to complement *hisA* and *trpF* minus *Escherichia coli* mutants (Wright et al. 2008). Enzyme in vitro characterizations were performed using coupled enzyme assays when proteins could be expressed and purified to homogeneity, as we have previously done for other PriA enzymes (Noda-Garcia et al. 2010). Three independent assays were performed to obtain Michaelis Menten kinetic parameters (table 1 and fig. 3). As hypothesized, these experiments allowed us to confirm the dual-substrate specificity of the PriA enzyme obtained from two organisms, *Corynebacterium jeikeium* and *C. amycolatum*, belonging to the subclade containing the *his-trp* gene cluster but lacking a *trpF* gene (fig. 2). The kinetic parameters obtained for the enzyme from *C. jeikeium*, from which the enzyme could be purified, were found to be similar to those previously obtained for PriA enzymes from *M. tuberculosis* and *S. coelicolor*.

The PriA homologs from *C. diphtheriae*, *C. efficiens*, *C. glutamicum*, *C. matruchotii*, and *C. striatum*, whose genomes encode functional HGT-acquired WPTO *trpFs*, were comprehensively characterized (fig. 2). These PriA homologs were found to completely lack PRA isomerase activity. Conversion of PRA could not be detected, either with highly sensitive in vivo complementation assays based in high copy number plasmids with strong promoters or by active site saturation conditions in vitro. This result contrasts with the catalytic efficiency of these enzymes when conversion of ProFAR was tested (table 1 and fig. 3). Technical problems with purified proteins used in the assays aimed at detection of PRA conversion in vitro thus can be ruled out.

On the basis of these results, we propose to rename the PriA homologs from this subclade as subHisA, a more appropriate name that reflects the function of these enzymes, and the subfunctionalization process involved in the narrowing of their substrate specificity. We next utilized X-ray crystallography and molecular dynamic analyses to address the structural foundations of the functional shift from PriA to subHisA.

Identification of Active Site Mutations in SubHisA

To compare subHisA with PriA at the structural level, we attempted to elucidate the structure of several subHisA homologs. We crystallized and solved the structure of subHisA from *C. efficiens* (2.25 Å resolution, PDB: 4AXK; supplementary table S2, Supplementary Material online). Detailed structural comparisons, taking into account all previous functional, structural, and site-directed mutagenesis knowledge, revealed important differences between PriA and subHisA, as discussed further in the following paragraphs. The changes identified during these analyses include both different 3D positions and identity of key active-site residues. Additionally, although potentially interesting as we have previously postulated (Wright et al. 2008; Noda-Garcia et al. 2010), including the specialized HisA enzyme within these comparisons was considered, but without giving good results. This was attributed to the impossibility of properly comparing HisA with PriA and subHisA, and drawing conclusions from such comparisons, as the available HisA structures lack substrate analogs and are too divergent from organisms unrelated to the Actinobacteria.

Residues known to be catalytically important for conversion of PRA, namely, Arg143, His50, and Ser81 (Due et al. 2011), were found to be different between PriA and subHisA. In PriA, Arg143 interacts with the catalytic general acid Asp175, allowing its correct polarization and thus preventing clashes between Asp130 and the carboxylate of PRA. In subHisA, Arg143 is replaced by Asn142, which not only lacks the correct charge to perform an analogous role but also is at least 10 Å away from the active site (fig. 4). Furthermore, the equivalent to Asp130 in PriA, that is, Asp127 in subHisA, is shifted two positions toward the N-terminus and 6 Å away from the active site. Although the exact role of Asp130, a HisA and PriA universally conserved residue, remains to be clarified, it is known to be functionally essential (Wright et al. 2008; Due et al. 2011).

In PriA, the specific binding of PRA occurs through the residues His50 and Ser81. Although Ser81 is conserved in PriA, this position contains a threonine in subHisA. A change from serine to threonine may seem a subtle change, but the methyl group of the threonine may affect the contact made between the hydroxyl group common to these residues and PRA. Indeed, mutation of serine to threonine in PriA has been shown to abolish PRA isomerase activity without affecting conversion of ProFAR (Wright et al. 2008; Noda-Garcia et al. 2010). However, drawing conclusions about His50 after structural analysis turned out to be more complicated than with Ser81. Although His50 is a conserved residue between PriA and subHisA, the residues next to it are different in both enzymes. Of potential relevance may seem to be a change of a Leucine in PriA into a Phenylalanine in subHisA. This modification is likely to alter the protonation state and electronegativity of His50, with a concomitant effect upon its binding capabilities (fig. 4).

The earlier-mentioned differences between PriA and subHisA, potentially accounting for the functional shift between these enzymes, may have been selected during evolution to avoid conversion of PRA by subHisA. An implication of this hypothesis would be that mutation of the residues next to His50, and Ser80 itself (which together bind the carboxylic group unique to PRA) may reverse the natural evolutionary process that led to narrowing of substrate specificity in subHisA. Thus, guided by the multiple sequence alignment of figure 5, a triple Leu48Ile, Phe50Leu, and Ser80Thr mutant of subHisA from *C. diphtheria* was constructed. As hypothesized, these mutations were found to generate an enzyme capable of converting PRA into CdRP. Moreover, the gain of PRA isomerase function in this triple subHisA mutant occurs without compromising its original ProFAR isomerase activity (table 1 and fig. 3).

Loss of Conformational Diversity in SubHisA

As a way to compare the conformational diversity of subHisA and PriA, we performed molecular dynamics simulations.

Table 1. Functional Characterization of PriA Homologs.

Organism (Enzyme)	In Vivo Activity (complementation)		In Vitro Activity[a]					
	HisA	TrpF	HisA			TrpF		
			K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1}\ s^{-1}$)	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M ($\mu M^{-1}\ s^{-1}$)
Escherichia coli (HisA)[b]			1.6	4.9	3.1			
E. coli (TrpF)[b]						12.2	34.5	2.82
Mycobacterium tuberculosis[c] (PriA)	+	+	19	0.23	0.012	21	3.6	0.17
Streptomyces coelicolor[c,d] (PriA)	+	+	3.6 ± 0.7	1.3 ± 0.2	0.36	5.0 ± 0.08	3.4 ± 0.09	0.68
Corynebacterium amycolatum (PriA)	+	+						
C. jeikeium (PriA)	+	+	2.3 ± 0.2	0.9 ± 0.08	0.39	5.1 ± 1.0	1.6 ± 0.16	0.31
C. diphtheriae (subHisA)	+	–	4.4 ± 0.5	2.6 ± 0.3	0.59	n.d.	n.d	n.d.
C. efficiens (subHisA)	+	–	1.9 ± 0.3	2.7 ± 0.5	1.42	n.d.	n.d.	n.d.
C. glutamicum (subHisA)	+	–						
C. matruchotii (subHisA)	+	–						
C. striatum (subHisA)	+	–	6.9 ± 0.7	2.1 ± 0.5	0.3	n.d	n.d.	n.d
C. diphtheriae (subHisA*) (Leu48Ile-Phe50Leu-Thr80Ser)	+	+	4.5 ± 1.5	0.6 ± 0.08	0.13	133 ± 10	0.05 ± 0.01	0.0004

NOTE.—The discrepancy between the *M. tuberculosis* data and all other PriA enzymes reported here may relate to the fact that suboptimal conditions were used for determination of the *M. tuberculosis* enzyme kinetic parameters, leading to an underestimation of its K_M. Standard deviation is not provided for data obtained from previously published works.

[a] Each data point comes from at least three independent determinations using freshly purified enzyme. n.d., activity not detected, even using active-site saturation conditions. Empty entries reflect our inability to properly express and/or solubilize these proteins.

[b] Data obtained from Henn-Sax et al. (2002) for *E. coli* HisA and Sterner et al. (1996) for *E. coli* TrpF.

[c] In vitro data obtained from Noda-Garcia et al. (2010) and Due et al. (2011) for *S. coelicolor* and *M. tuberculosis*, respectively.

[d] In vivo data obtained from Barona-Gomez and Hodgson (2003).

Given that the catalytic efficiency of the enzyme with solved structure, that is, subHisA from *C. efficiens* (PDB: 4AXK) seems to differ from all other subHisA enzymes that were biochemically characterized (table 1), we constructed a homology model of subHisA from *C. diphtheriae* (80% ID). After systematic searches, we obtained an ad hoc set of optimized conditions for the molecular dynamics study of PriA from *M. tuberculosis* (PDB: 2Y89) and subHisA from both *C. efficiens* and *C. diphtheriae*. Notably, the same thermodynamic behavior was found for the two subHisA enzymes (supplementary fig. S3, Supplementary Material online). This result suggests that narrowing of substrate specificity follows a common molecular mechanism. Hence, from this point onward, we will refer to subHisA indistinctively of the *Corynebacterium* species it comes from.

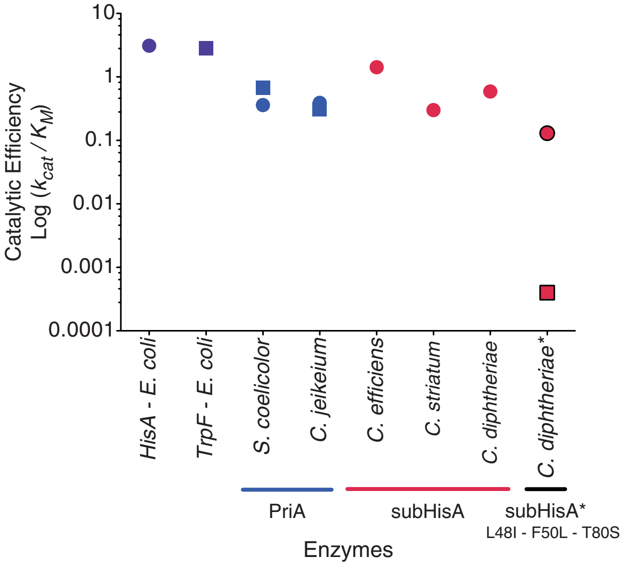


Fig. 3. Selected PRA and ProFAR isomerase catalytic efficiencies. ProFAR isomerase (HisA) and PRA isomerase (TrpF) activities are shown in circles and squares, respectively. Data from *Escherichia coli* were obtained from Henn-Sax et al. (2002) and Sterner et al. (1996). Data from *Streptomyces coelicolor* were obtained from Noda-Garcia et al. (2010). subHisA* (Leu48Ile, Phe50Leu and Thr80Ser) activities are highlighted. The detailed enzyme kinetic parameters and in vivo characterization is provided as table 1.

We found subHisA to have a more compact tertiary structure than PriA, despite the fact that both enzymes show similar overall thermodynamic stable structures in solution, indicated by kindred backbone root-mean-square deviations (RMSD). Moreover, the internal hydrogen bonding networks seem to be equivalent in PriA and subHisA (supplementary fig. S3, Supplementary Material online). In contrast, individual side-chain RMSD of residues contained on β to α loops 1 and 6, as well as in α helix 7, were found to be significantly higher in PriA than in subHisA (fig. 6*A* and supplementary movie S1, Supplementary Material online). This observation is in agreement with the fact that PriA adopts different conformational substates related to its dual-substrate specificity (Wright et al. 2008; Due et al. 2011). More importantly, subHisA, which can only accept ProFAR as a substrate, may have lost conformational diversity during the process leading to the narrowing of this enzyme's substrate specificity.

To further investigate the importance of the enzymes accessible conformational states, we performed a principal component analysis (PCA) on the molecular dynamics of PriA and subHisA. This approach allowed us to cluster all conformations adopted by the enzymes throughout their corresponding dynamics in solution. Indeed, our analyses revealed the existence of four most populated conformational states in PriA, and only one in subHisA. Interestingly, two of the four conformational states predicted for PriA after these analyses were previously reported using cocrystal structures with PRA and ProFAR analogs (Due et al. 2011). The latter observation strongly supports the validity of our findings, which are highlighted in figure 6*B*. Given that substrates were not used for these analyses, moreover, the conformational space explored by PriA thus appears to be independent of an induced-fit mechanism

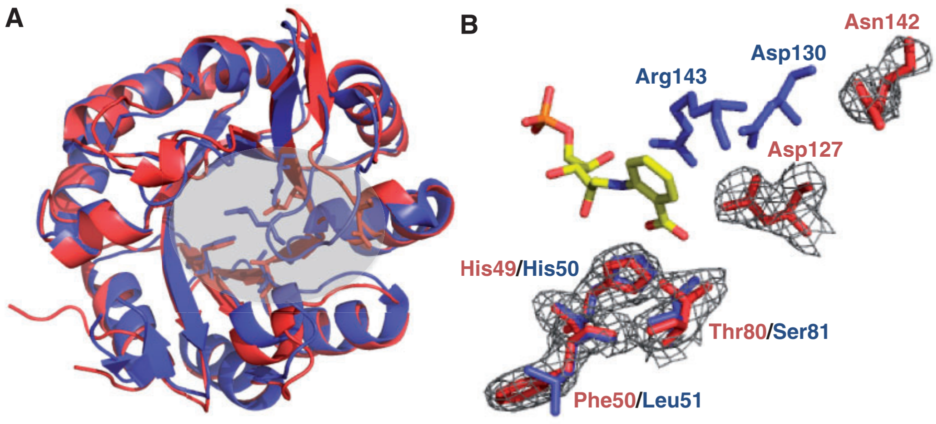
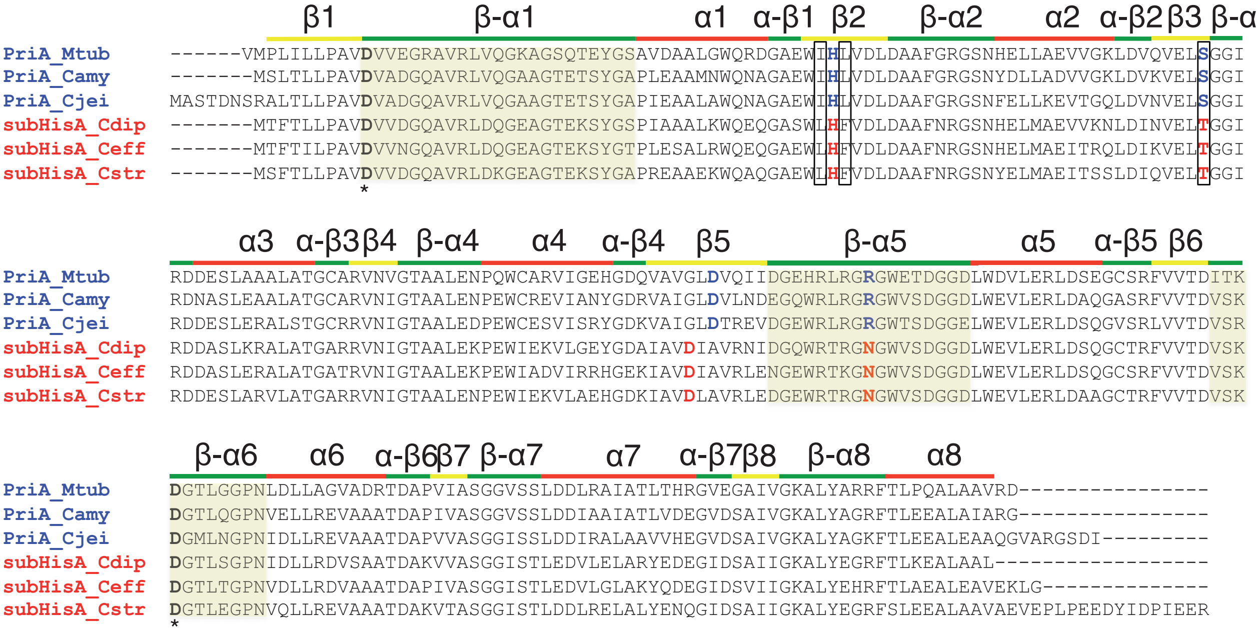


Fig. 4. X-ray structural and sequence analysis of subHisA. (*A*) Structure of PriA (PDB: 2Y85, blue) superimposed on *Corynebacterium efficiens* subHisA (chain A, PDB: 4AXK, red). Key residues in the active site are highlighted. (*B*) Zoom-in of the superimposed active-site residues of PriA (blue) and subHisA (red), with rCdRP (yellow), showing at the bottom of the active site the substrate binding residues His49, Phe50 and Thr80, as well as variant residues Asp127 and Asn142, which adopt a novel architecture at the top of the active site.



```
PriA_Mtub     ------VMPLILLPAVDVVEGRAVRLVQGKAGSQTEYGSAVDAALGWQRDGAEWIHLVDLDAAFGRGSNHELLAEVVGKLDVQVELSGGI
PriA_Camy     -------MSLTLLPAVDVADGQAVRLVQGAAGTETSYGAPLEAAMNWQNAGAEWIHLVDLDAAFGRGSNYDLLADVVGKLDVKVELSGGI
PriA_Cjei     MASTDNSRALTLLPAVDVADGQAVRLVQGAAGTETSYGAPIEAALAWQNAGAEWIHLVDLDAAFGRGSNFELLKEVTGQLDVNVELSGGI
subHisA_Cdip  -------MTFTLLPAVDVVDGQAVRLDQGEAGTEKSYGSPIAAALKWQEQGASWLHFVDLDAAFNRGSNHELMAEVVKNLDINVELTGGI
subHisA_Ceff  -------MTFTILPAVDVVNGQAVRLDQGEAGTEKSYGTPLESALRWQEQGAEWLHFVDLDAAFNRGSNHELMAEITRQLDIKVELTGGI
subHisA_Cstr  -------MSFTLLPAVDVVDGQAVRLDKGEAGTEKSYGAPREAAEKWQAQGAEWLHFVDLDAAFNRGSNYELMAEITSSLDIQVELTGGI
                              *

PriA_Mtub     RDDESLAAALATGCARVNVGTAALENPQWCARVIGEHGDQVAVGLDVQIIDGEHRLRGRGWETDGGDLWDVLERLDSEGCSRFVVTDITK
PriA_Camy     RDNASLEAALATGCARVNIGTAALENPEWCREVIANYGDRVAIGLDVLNDEGQWRLRGRGWVSDGGDLWEVLERLDAQGASRFVVTDVSK
PriA_Cjei     RDDESLERALSTGCRRVNIGTAALEDPEWCESVISRYGDKVAIGLDTREVDGEWRLRGRGWTSDGGELWEVLERLDSQGVSRLVVTDVSR
subHisA_Cdip  RDDASLKRALATGARRVNIGTAALEKPEWIEKVLGEYGDAIAVDIAVRNIDGQWRTRGNGWVSDGGDLWEVLERLDSQGCTRFVVTDVSK
subHisA_Ceff  RDDASLERALATGATRVNIGTAALEKPEWIADVIRRHGEKIAVDIAVRLENGEWRTKGNGWVSDGGDLWEVLERLDSQGCSRFVVTDVSK
subHisA_Cstr  RDDESLARVLATGARRVNIGTAALENPEWIEKVLAEHGDKIAVDLAVRLEDGEWRTRGNGWVSDGGDLWEVLERLDAAGCTRFVVTDVSK

PriA_Mtub     DGTLGGPNLDLLAGVADRTDAPVIASGGVSSLDDLRAIATLTHRGVEGAIVGKALYARRFTLPQALAAVRD----------------
PriA_Camy     DGTLQGPNVELLREVAAATDAPIVASGGVSSLDDIAAIATLVDEGVDSAIVGKALYAGRFTLEEALAIARG----------------
PriA_Cjei     DGMLNGPNIDLLREVAAATDAPVVASGGISSLDDIRALAAVVHEGVDSAIVGKALYAGKFTLEEALEAAQGVARGSDI---------
subHisA_Cdip  DGTLSGPNIDLLRDVSAATDAKVVASGGISTLEDVLELARYEDEGIDSAIIGKALYEGRFTLKEALAAL------------------
subHisA_Ceff  DGTLTGPNVDLLRDVAAATDAPIVASGGISTLEDVLGLAKYQDEGIDSVIIGKALYEHRFTLAEALEAVEKLG--------------
subHisA_Cstr  DGTLEGPNVQLLREVAAATDAKVTASGGISTLDDLRELALYENQGIDSAIIGKALYEGRFSLEEALAAVAEVEPLPEEDYIDPIEER
              *
```

Fig. 5. Multiple sequence alignment of PriA and subHisA sequences. Catalytic residues, Asp11 and Asp 175, are marked with an asterisk. PRA binding residues are also highlighted. subHisA* gain-of function residues are framed. The secondary structures, including loops, α helixes and β sheets, are shown at the top of the sequence. The sequence regions corresponding to loops 1, 5, and 6 is highlighted.

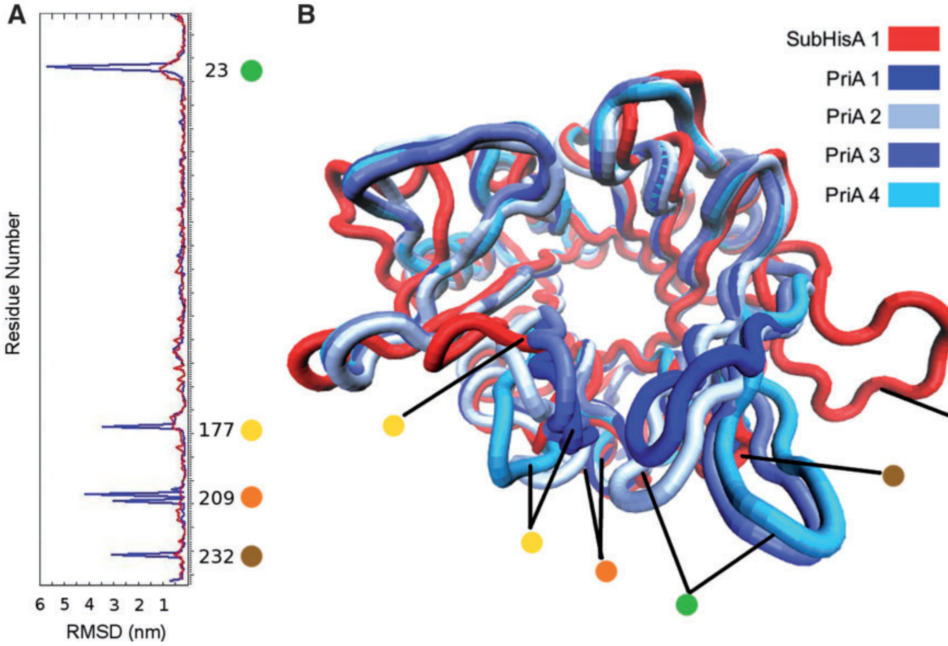


Fig. 6. Molecular dynamics of subHisA and PriA. (*A*) RMSD per residue of PriA (blue) and subHisA (red) with respect to equilibrated initial structures. (*B*) Different average structures, or conformational states, found for PriA (four shades of blue) and subHisA (red) after clustering of the molecular dynamics trajectories based in a PCA.

Discussion

The most accepted hypothesis regarding enzyme evolution embraces enzyme substrate ambiguity and the idea that modern enzymes are the result of specialization processes prompted by gene duplication (Jensen 1976; Piatigorsky 2007). We found, however, that a generalist enzyme, PriA, is present in approximately 50% of the organisms belonging to the closely related genera *Mycobacterium* and *Corynebacterium*. Lack of TrpF, and occurrence of a PriA enzyme with dual-substrate specificity, is in agreement with early suggestions of a metabolic interlock and common ancestry between L-histidine and L-tryptophan biosynthesis (Jensen 1969; Kane and Jensen 1970; Nester and Montoya 1976). The evidence provided by these reports suggest cross-regulation, potentially involving the common biosynthetic precursor phosphoribosyl pyrophosphate, in *Bacillus subtilis*. Therefore, specialization of these ancient biosynthetic pathways during the course of evolution in most Actinobacteria must have been impeded by strong physiological constrains that outweigh the benefits of enzyme proficiency and pathway specialization.

If enzyme specialization in the subclades containing PriA enzymes is constrained by strong factors, then gene duplication and subsequent divergence can be expected to occur at low frequency, making it an unlikely event. HGT, as a driving force for specialization of L-histidine and L-tryptophan biosynthesis in the *Corynebacterium* lineage receiving the WPTO may have overcome the limitations of evolution through gene duplication. Indeed, organisms are believed to have evolved regulation of metabolism in a pathway-specific manner only possible in the absence of substrate ambiguity (Jensen 1976). Interestingly, in *C. glutamicum*, where we confirmed the existence of a subHisA enzyme, feedback gene regulation of L-tryptophan (Ikeda 2006; Brune et al. 2007; Xie et al. 2003, 2004) and L-histidine (Jung et al. 2010) biosynthesis, which are specialized pathways, seems to have evolved. The foregoing physiological regime contrasts with the proficiency of its broad substrate PriA ancestor, which is encoded within a tightly packed, conserved and constitutively expressed *his-trp* gene cluster (Hu et al. 1999; Hodgson 2000; Parish 2003).

The occurrence of an ancestral-like scenario in modern organisms, that is, a generalist enzyme relying in a single active site that supports committed pathways, not only challenges the view of duplication followed by functionalization as a mandatory process in enzyme evolution (Hughes 1994; Depristo 2007; Des Marais and Rausher 2008) but also raises the interesting question of functional trade-off. Narrowing of enzyme substrate specificity in subHisA shows that loss of one of the ancestral activities can occur without compromising the catalytic efficiency of the remaining enzyme function, even within a highly constrained active site. The conserved substitutions in the branch where subHisA has evolved suggests that narrowing of substrate specificity in purely biochemical processes may involve positive selection. In the presence of a PRA isomerase encoded by a WPTO *trpF*, the highly conserved mutations leading to subHisA (e.g., Leu48Ile, Phe51Leu, Ser80Thr, Arg142Asn, and shift of Asp127) may have provided an adaptive mechanism to avoid productive binding of PRA.

This idea is in agreement with the regulatory and physiological regime described for subHisA-containing *Corynebacterium* species in previous paragraphs. Therefore, given the promiscuous-prone active site of subHisA, as demonstrated by our site-directed mutagenesis experiment, mutations that would restrain PRA from binding—without affecting binding of ProFAR—must have been selected for. Solution to such conundrum speaks out of a complex evolutionary history shaped by the unknown mechanisms by which HGT operates. Moreover, this may be the reason why subHisA has not been able to re-specialize to the levels encountered in mono-functional HisA enzymes, implying a trade-off in terms of evolvability, rather than in absolute enzyme proficiencies. Investigating the reversibility of subHisA into PriA, to identify mutations involved in this functional trade-off beyond those that could be pinpointed after our structural analyses, may shed some light into the raising issues of reversibility (Tokuriki et al. 2012).

Our molecular dynamics analyses allowed us to compare the extent of conformational diversity between two closely related enzymes with broad (or “generalist”) and narrow substrate (or “specialist”) specificity. As conformational diversity has been hypothesized to serve as evolutionary raw material (James and Tawfik 2003; Tokuriki and Tawfik 2009), our discovery that this conformational diversity is lost in the narrow substrate, or ‘specialist’ enzyme, is remarkable. PriA has been previously postulated to accommodate and convert two different substrates through conformational changes (Wright et al. 2008; Noda-Garcia et al. 2010; Due et al. 2011). The molecular dynamics results are consistent with these observations. It should be noted, however, that the conformational states explored by PriA exist irrespective of the presence of substrates, questioning the likelihood of an induced-fit mechanism.

In conclusion, during dynamic genome processes, which may include HGT and differential gene loss, positive selection may be needed to drive both 1) evolution of narrowing of enzyme substrate specificity from a generalist enzyme; and 2) efficient assembly of HGT-acquired biosynthetic pathways within the receiving metabolic network, as previously postulated (Klassen 2009). Our results also emphasize the need for an integrated view on the evolution of enzyme substrate specificity, which should include prokaryotic physiology and genetics. Incorporating HGT into current models of enzyme evolution, including its formalization within population genetics, seems both a necessity and an opportunity for evolutionary biology. Finally, our results demonstrate the importance of multidisciplinary approaches as a powerful conceptual framework to investigate complex evolutionary mechanisms in biochemical and biophysical processes (Dean and Thornton 2007).

Materials and Methods

Bioinformatics Analysis

The Blast algorithm was used for database searches. The sequences were aligned with Muscle version 3.6 and edited with Jalview. ProtTest v1.4 (Abascal et al. 2005) was used to select, out of 56 different models, the protein evolution model that best fit the protein alignments of PriA. According to the statistical AIC, this model was WAG + I + G + F. The selected protein evolution model and its parameters were used for the reconstruction of protein phylogenies using the maximum likelihood methods (Guindon et al. 2010). The genome context analyses were done using the Artemis Comparative Tool (Carver et al. 2005).

Functional Characterization of PriA Homologs

PriA coding sequence from *C. jeikeium* was synthesized by our group (supplementary fig. S1, Supplementary Material online) and subHisA coding sequences from *C. amycolatum*, *C. efficiens*, *C. matruchotii*, and *C. striatum* were synthesized by GeneART; in both cases, codons were optimized for its over expression in *E. coli* (table S3, Supplementary Material online). subHisA from *C. diphtheriae* and *C. glutamicum* were cloned from genomic DNA gently gifted by Androulla Efstratiou (Health Protection Agency, UK) and from the ATCC collection, respectively. All enzymes were cloned in a pQE-30 derivative (Qiagen) and pET22b (Novagen) using the enzymes NdeI and HindIII. In vivo *E. coli trpF* and *hisA* complementation assays were done as previously reported (Wright et al. 2008) other than pQE-30 (Qiagen) derivatives were used, and M9 minimal medium was enriched with a mixture of all the amino acids at 50 μg/ml other than L-histidine and L-tryptophan. Enzyme purification by Nickel affinity chromatography was performed as previously reported (Noda-Garcia et al. 2010). In vitro Michaelis Menten enzyme kinetic parameters of both PRA and ProFAR isomerase activities were measured as previously reported using as controls known enzymes, both active (positive control) and inactive (negative control) (Henn-Sax et al. 2002; Noda-Garcia et al. 2010).

Construction of the subHisA* Mutant

The mutant subHisA_Cdip_Leu48Ile-Phe50Leu-Thr80Ser was constructed using the site-directed mutagenesis kit from Stratagene. The triple mutant was constructed using the pQEI_subHisA_Cdip_Thr80Ser as a template (Noda-Garcia et al. 2010) and the oligonucleotides Leu48Ile-Phe50Leu_For 5′ggggcatcgtggattcatctggtggatttagat and Leu48Ile-Phe50Leu_Rev 5′atctaaatccaccagatgaatccacgatgcccc. subHisA* was cloned in pET22b (Novagen) using the enzymes NdeI and HindIII and sequenced before functional analysis.

X-Ray Crystallography

Overexpressed sub-HisA from *Corynebacterium efficiens* was purified as a 6X His-tagged fusion from plasmid pET22-Ceff in *E. coli* strain BL21star (DE3) in LB broth. Soluble protein was obtained as reported previously for PriA (Wright et al. 2008; Noda-Garcia et al. 2010). Initial crystallization trials were performed with screens from Molecular Dimensions Ltd, Hampton Research and Emerald Bio-structures Inc. using the sitting drop vapor diffusion technique. Needle-shaped crystals were obtained with conditions 20 (0.1 M HEPES pH 7.5, 1.4 M Sodium Citrate), 70 (0.1 M Bis-Tris pH 5.5, 0.2 M $MgCl_2$, 25% w/v PEG 3350), and 71 (0.1 M *bis*–tris pH 6.5, 0.2 M $MgCl_2$, 25% w/v PEG 3350) of the Hampton Research screen using 0.2 μl of protein at 15 mg/ml mixed with an equal volume of mother liquor. After optimization, crystals grew after 1 or 2 days at 291 K in mother liquor consisting of 0.1 M *bis*–tris pH 7.5, 25% v/w PEG 3350 and 0.2 M $MgCl_2$, and mixing 1 μl of protein at 15 mg/ml with an equal volume of mother liquor.

Prior to data collection, PriA crystals were cryoprotected by dipping in mother liquor containing 30% of glycerol and immediately frozen in the N_2 cryostream. X-ray data were collected on the I04 beamline at the Diamond synchrotron (UK) using an ADSC Q315 CCD detector. All data were indexed, integrated, and scaled using the XDS package. Subsequent data handling was carried out using the CCP4 software package (1994). Molecular replacement was carried out using the coordinates of *S. coelicolor* PriA (PDB: 2vep) as a search model with the PHASER program (McCoy et al. 2007). Refinement of the structure was carried out by alternate cycles of REFMAC (Murshudov et al. 1997) using noncrystallographic symmetry (NCS) restraints and manual rebuilding in O (Jones et al. 1991). Water molecules were automatically added to the atomic model by Arp/wARP (Perrakis et al. 1997) and in the last steps of refinement all the NCS restraints were released. A summary of the data collection and refinement statistics is given in supplementary table S2, Supplementary Material online.

Molecular Dynamics Simulations

To find the best protocol to perform the molecular dynamics analysis, an optimization protocol specified in supplementary text S1 and table S5, Supplementary Material online, was followed. Missing loops from the crystal structure of PriA from *M. tuberculosis* (PDB: 2Y89) and a comparative model of subHisA from *C. diphtheriae* based on the crystal structure of *C. efficiens* (PDB: 4AXK, this study) were constructed using Rosetta 3.2.1 (Leaver-Fay et al. 2011). Addition of missing side-chains and protons was achieved with the WHATIF tools (Vriend 1990) keeping its predicted protonation state for His residues and assuming a neutral pH. Topology files, computational cubic box, solvation, system neutralization by addition of NaCl, system minimization, equilibration, and molecular dynamics simulations were carried out using GROMACS 4.5.3 (Hess et al. 2008). For this, CHARMM27 all-atom force field (with CMAP) version 2.0 (MacKerell et al. 1998) and explicit TIP3P water (Jorgensen et al. 1983) were used. Systems were minimized for 5,000 conjugate gradient steps and heated up to 300 K during 600 ps with protein atoms harmonically restrained. This was followed by equilibration steps done under NvT conditions (300 K) and then under NpT conditions (1 atm), during 1 ns each, using the V-rescale and isotropic Berendsen barostat methods without atom restraints.

Long-range electrostatics interactions were included using the Reaction Field method. A cutoff for the van der Waals interactions was applied with a 1.2-nm radius, and the LINCS method was used to restrain all bonds involving hydrogen atoms. Molecular dynamics of 300 ns was performed with a time step of 2 fs. Trajectories were obtained by saving the atomic coordinates of the whole system every 80 ps. Generation of DCD and PSF files was done with VMD's psfgen plugin (Humphrey et al. 1996). Calculation of global RMSDs, radius of gyration, and hydrogen bond formation as a function of time, and average RMSDs per residue were estimated with tools from GROMACS 4.5.3 (Hess et al. 2008).

Cross-correlation matrix, PCA, clustering, and average structures were obtained using Carma 1.0 (Glykos 2006). All numerical simulations and corresponding analysis were performed at the supercomputing center (mazorka) at Langebio. Structure, dynamics, and PCA comparisons among subHisA from *C. efficiens*, subHisA from *C. diphtheriae*, and PriA from *M. tuberculosis* are specified in supplementary text S1 and S2, Supplementary Material online.

Supplementary Material

Supplementary tables S1–S4, figures S1 and S2, text S1 and S2, and movie S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors are indebted to Prof. Therese Markow for useful discussions and critical reading of this manuscript. The authors thank Karina Verdel-Aranda, Helena Wright, Hilda E. Ramos-Aboites, Dean Rea, and Ralf Flaig for technical support, preliminary experiments, and help with X-ray data collection. This work was supported by a UCMEXUS Conacyt grant to F.B.-G. with Steven E. Brenner and Kimmen Sjölander as co-applicants, by Conacyt grants (No. 50952-Q and No. 83039) to F.B.-G. and (No. 132376) M.C.-T., by a joint international Royal Society (UK) grant to V.F. and F.B.-G., and by the Birmingham Warwick Science City Translational Medicine project to V.F.

References

Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.

Barona-Gomez F, Hodgson DA. 2003. Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* 4:296–300.

Brune I, Jochmann N, Brinkrolf K, Huser AT, Gerstmeir R, Eikmanns BJ, Kalinowski J, Puhler A, Tauch A. 2007. The IclR-type transcriptional repressor LtbR regulates the expression of leucine and tryptophan biosynthesis genes in the amino acid producer *Corynebacterium glutamicum*. *J Bacteriol.* 189:2720–2733.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. 2005. ACT: the Artemis comparison tool. *Bioinformatics* 21:3422–3423.

Collaborative Computational Project, Number 4 (CCP4). 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr.* 50:760–763.

Dean AM, Thornton JW. 2007. Mechanistic approaches to the study of evolution: the functional synthesis. *Nat Rev Genet.* 8:675–688.

Depristo MA. 2007. The subtle benefits of being promiscuous: adaptive evolution potentiated by enzyme promiscuity. *HFSP J.* 1:94–98.

Des Marais DL, Rausher MD. 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* 454:762–765.

Dittmar K, Liberles D. 2010. Evolution after gene duplication. Hoboken (NJ): Wiley-Blackwell.

Due AV, Kuper J, Geerlof A, von Kries JP, Wilmanns M. 2011. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. *Proc Natl Acad Sci U S A.* 108:3554–3559.

Glykos NM. 2006. Software news and updates—Carma: a molecular dynamics analysis program. *J Comput Chem.* 27:1765–1768.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 59:307–321.

Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K, Sterner R. 2002. Two (betaalpha)(8)-barrel enzymes of histidine and tryptophan biosynthesis have similar reaction mechanisms and common strategies for protecting their labile substrates. *Biochemistry* 41:12032–12042.

Hess B, Kutzner C, van der Spoel D, Lindahl E. 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput.* 4: 435–447.

Hodgson DA. 2000. Primary metabolism and its control in streptomycetes: a most unusual group of bacteria. *Adv Microb Physiol.* 42: 47–238.

Hu DS, Hood DW, Heidstra R, Hodgson DA. 1999. The expression of the trpD, trpC and trpBA genes of *Streptomyces coelicolor* A3(2) is regulated by growth rate and growth phase but not by feedback repression. *Mol Microbiol.* 32:869–880.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci.* 256:119–124.

Humphrey W, Dalke A, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graph Model.* 14:33–38.

Ikeda M. 2006. Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. *Appl Microbiol Biotechnol.* 69:615–626.

James LC, Tawfik DS. 2003. Conformational diversity and protein evolution—a 60-year-old hypothesis revisited. *Trends Biochem Sci.* 28:361–368.

Jensen RA. 1969. Metabolic interlock. Regulatory interactions exerted between biochemical pathways. *J Biol Chem.* 244:2816–2823.

Jensen RA. 1976. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol.* 30:409–425.

Jones TA, Zou JY, Cowan SW, Kjeldgaard M. 1991. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr A.* 47: 110–119.

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. 1983. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* 79:926–935.

Jung S, Chun JY, Yim SH, Lee SS, Cheon CI, Song E, Lee MS. 2010. Transcriptional regulation of histidine biosynthesis genes in *Corynebacterium glutamicum*. *Can J Microbiol.* 56:178–187.

Kane JF, Jensen RA. 1970. Metabolic interlock. The influence of histidine on tryptophan biosynthesis in *Bacillus subtilis*. *J Biol Chem.* 245: 2384–2390.

Klassen JL. 2009. Pathway evolution by horizontal transfer and positive selection is accommodated by relaxed negative selection upon upstream pathway genes in purple bacterial carotenoid biosynthesis. *J Bacteriol.* 191:7500–7508.

Kuper J, Doenges C, Wilmanns M. 2005. Two-fold repeated (betaalpha)4 half-barrels may provide a molecular tool for dual substrate specificity. *EMBO Rep.* 6:134–139.

Leaver-Fay A, Tyka M, Lewis SM, et al. (32 co-authors). 2011. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* 487:545–574.

Lerat E, Daubin V, Ochman H, Moran NA. 2005. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol.* 3:e130.

MacKerell AD, Bashford D, Bellott M, et al. (29 co-authors). 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B.* 102:3586–3616.

McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *J Appl Crystallogr.* 40: 658–674.

Murshudov GN, Vagin AA, Dodson EJ. 1997. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr.* 53:240–255.

Nester EW, Montoya AL. 1976. An enzyme common to histidine and aromatic amino acid biosynthesis in *Bacillus subtilis*. *J Bacteriol.* 126: 699–705.

Noda-Garcia L, Camacho-Zarco AR, Verdel-Aranda K, Wright H, Soberon X, Fulop V, Barona-Gomez F. 2010. Identification and analysis of residues contained on beta → alpha loops of the dual-substrate (beta alpha)8 phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci.* 19: 535–543.

Ohno S. 1970. Evolution by gene duplication. New York (NY): Springer-Verlag.

Pal C, Papp B, Lercher MJ. 2005. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet.* 37: 1372–1375.

Parish T. 2003. Starvation survival response of *Mycobacterium tuberculosis*. *J Bacteriol.* 185:6702–6706.

Perrakis A, Sixma TK, Wilson KS, Lamzin VS. 1997. wARP: Improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. *Acta Crystallogr D Biol Crystallogr* 53:448–455.

Piatigorsky J. 2007. Gene sharing and evolution: the diversity of protein function. Cambridge (MA): Harvard University Press.

Sterner R, Kleemann GR, Szadkowski H, Lustig A, Hennig M, Kirschner K. 1996. Phosphoribosyl anthranilate isomerase from *Thermotoga maritima* is an extremely stable and active homodimer. *Protein Sci.* 5:2000–2008.

Tokuriki N, Jackson CJ, Afriat-Jurnou L, Wyganowski KT, Tang R, Tawfik DS. 2012. Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme. *Nat Commun.* 3:1257.

Tokuriki N, Tawfik DS. 2009. Protein dynamism and evolvability. *Science* 324:203–207.

Treangen TJ, Rocha EP. 2011. Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* 7: e1001284.

Vriend G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph.* 8:52–56, 29.

Wright H, Noda-Garcia L, Ochoa-Leyva A, Hodgson DA, Fulop V, Barona-Gomez F. 2008. The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem Biophys Res Commun.* 365:16–21.

Xie G, Bonner CA, Song J, Keyhani NO, Jensen RA. 2004. Inter-genomic displacement via lateral gene transfer of bacterial trp operons in an overall context of vertical genealogy. *BMC Biol.* 2:15.

Xie G, Keyhani NO, Bonner CA, Jensen RA. 2003. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. *Microbiol Mol Biol Rev.* 67:303–342.