Published in final edited form as:

Biochem J. 2016 May 01; 473(9): 1141–1152. doi:10.1042/BJ20151271.

Co-occurrence of analogous enzymes determines evolution of a novel ($\beta\alpha$)₈-isomerase sub-family after non-conserved mutations in flexible loop

Ernesto A. Verduzco-Castro^{*}, Karolina Michalska^{†,‡}, Michael Endres[†], Ana L. Juárez-Vazquez^{*}, Lianet Noda-García^{*,1}, Changsoo Chang^{†,‡}, Christopher S. Henry[§], Gyorgy Babnigg[†], Andrzej Joachimiak^{†,‡,||}, Francisco Barona-Gómez^{*,2}

^{*}Evolution of Metabolic Diversity Laboratory, Unidad de Genómica Avanzada (Langebio), Cinvestav-IPN, Irapuato, CP36821, México

[†]Midwest Center for Structural Genomics, Argonne National Laboratory, Argonne, Illinois 60439, U.S.A.

[‡]Structural Biology Center, Biosciences Division, Argonne National Laboratory, Argonne, Illinois 60439, U.S.A.

[§]Mathematics and Computer Science, Argonne National Laboratory, Argonne, Illinois 60439, U.S.A.

^{||}Department of Biochemistry and Molecular Biology, University of Chicago, Illinois 60637, U.S.A.

Abstract

We investigate the evolution of co-occurring analogous enzymes involved in L-tryptophan and L-histidine biosynthesis in *Actinobacteria*. Phylogenetic analysis of *trpF* homologues, a missing gene in certain clades of this lineage whose absence is complemented by a dual-substrate HisA homologue, termed PriA, found that they fall into three categories: (i) *trpF-1*, an L-tryptophan biosynthetic gene horizontally acquired by certain *Corynebacterium* species; (ii) *trpF-2*, a paralogue known to be involved in synthesizing a pyrrolopyrrole moiety and (iii) *trpF-3*, a variable non-conserved orthologue of *trpF-1*. We previously investigated the effect of *trpF-1* upon the evolution of PriA substrate specificity, but nothing is known about the relationship between *trpF-3* and *priA*. After *in vitro* steady-state enzyme kinetics we found that *trpF-3* encodes a phosphoribosyl anthranilate isomerase. However, mutation of this gene in *Streptomyces sviveus* did not lead to auxothrophy, as expected from the biosynthetic role of *trpF-1*. Biochemical characterization of a dozen co-occurring TrpF-2 or TrpF-3, with PriA homologues, explained the

²To whom correspondence should be addressed (fbarona@langebio.cinvestav.mx).

¹Present address: Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel

AUTHOR CONTRIBUTION

Ernesto Verduzco-Castro performed biochemical and molecular biology experiments and analysed results. Ernesto Verduzco-Castro, Ana Juárez-Vazquez and Francisco Barona-Gómez designed and performed the sequence and phylogenetic analyses. Ernesto Verduzco-Castro and Michael Endres cloned, overexpressed and purified all proteins for crystallization trials. Ernesto Verduzco-Castro screened for crystallization conditions for all proteins. Karolina Michalska and Changsoo Chang collected X-ray diffraction data, solved and refined the protein structures. Ernesto Verduzco-Castro, Karolina Michalska, Lianet Noda-García, Christopher Henry, Gyorgy Babnigg, Andrzej Joachimiak and Francisco Barona-Gómez helped draft the manuscript. Ernesto Verduzco-Castro, Christopher Henry, Andrzej Joachimiak and Francisco Barona-Gómez conceived the study, participated in its design and coordination. All authors read and approved the final manuscript.

prototrophic phenotype, and unveiled an enzyme activity trade-off between TrpF and PriA. X-ray structural analysis suggests that the function of these PriA homologues is mediated by non-conserved mutations in the flexible L5 loop, which may be responsible for different substrate affinities. Thus, the PriA homologues that co-occur with TrpF-3 represent a novel enzyme family, termed PriB, which evolved in response to PRA isomerase activity. The characterization of co-occurring enzymes provides insights into the influence of functional redundancy on the evolution of enzyme function, which could be useful for enzyme functional annotation.

Keywords

analogous enzymes; ($\beta\alpha$)₈-barrel; PriA; *Streptomyces*; substrate specificity; TrpF

INTRODUCTION

L-tryptophan and L-histidine biosynthesis have a common chemical transformation involving an Amadori rearrangement. These reactions are catalysed by TrpF and HisA enzymes, which convert the aminoaldoses phosphoribosylanthranilate (PRA) and *N*'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (ProFAR) to the aminoketoses 1-(*O*-carboxyphenylamino)-1-deoxyribulose-5-phosphate (CdRP) and *N*'-[(5'-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide (PRFAR) respectively. Structurally these proteins adopt the same ($\beta\alpha$)₈-barrel fold (also commonly known as a TIM-barrel), despite sharing very little sequence identity (~10%) [1,2]. In *Escherichia coli* and relatives, TrpF and HisA are encoded by two different genes and belong to two independent feedback-regulated biosynthetic pathways (Figure 1A). Nevertheless, HisA can gain specificity for the TrpF substrate, PRA, after random gain-of-function mutations selected for in a *trpF* background [1,3]. Moreover, PRA specificity has been found to be highly evolvable in other mechanistically unrelated protein folds [4,5]. This poses PRA isomerase activity as a suitable enzymatic function to investigate functional shifts involving co-occurring analogous enzymes.

Here we take advantage of the convergent nature of L-tryptophan and L-histidine biosynthesis in the ancestral *Actinobacteria* phylum, which includes the model organisms *Streptomyces coelicolor* and *Mycobacterium tuberculosis*, to study enzyme co-evolution. Most *Actinobacteria* lack a *trpF* gene and their histidine and tryptophan biosynthetic pathways utilize a single dual-substrate ($\beta\alpha$)₈-barrel isomerase (Figure 1A), termed phosphoribosyl isomerase A (PriA) [6]. PriA evolved from its homologue HisA (~45% sequence identity), as shown after comprehensive phylogenetic analyses [7]. PriA's structure and function relationship has been also thoroughly investigated [8–11], and possible connections with other His and Trp enzymes have been proposed [12]. Moreover, given its physiologically relevant dual-substrate specificity, PriA has been adopted as a model system to study the evolution of the ($\beta\alpha$)₈-fold [12,13], and previously, to investigate genetic models driving enzyme evolution [3,14,15].

The absence of a *trpF* gene in *Actinobacteria*, however, is not universally conserved, as a complete tryptophan operon that includes a *trpF* gene exists in a sub-clade of the genus

Corynebacterium. The acquisition of this *trp* operon has been previously proposed to occur after horizontal gene transfer (HGT) [16], and we have recently shown that this HGT-acquired *trpF* correlates with lack of PRA affinity of the PriA homologues present in these organisms. The resulting product of this process is a monofunctional HisA, or 'subHisA', a name that reflects the subfunctionalization suffered by this enzyme. Interestingly, the altered specificity of subHisA was found to be due to both active site mutations and decrease in conformational flexibility [14].

As seen in many ($\beta\alpha$)₈-barrels [13], the core of PriA and subHisA is rigid, whereas loops exhibit substantial flexibility [11,12]. For instance, the functionally important L5 loop of PriA from *M. tuberculosis* (PriA_Mtub) adopts a β -hairpin-like structure (consisting of strands herein referred to as β 5a and β 5b), which assumes, depending on which molecule is present, either a relaxed (PRA-related product analogue rCdRP, which is the reduced form of CdRP) (Figure 1B) or twisted knot-like (ProFAR-related product PRFAR) (Figure 1C) conformation. In these two arrangements, Arg-143 and Trp-145, which anchor the aromatic fragment of the substrate molecule through stacking interactions, exchange their positions, leaving either Trp pointing into the active site and Arg facing the solvent, or *vice versa* [11]. Importantly, besides substrate binding, Arg-143 has been also postulated to participate in positioning of the catalytic Asp-175 for PRA isomerization, but it does not play such a role in the equivalent transformation of ProFAR [10,11].

Analysis of currently available actinobacterial genomes reveals that some of them contain a *trpF* homologue that is unrelated to that acquired by HGT in *Corynebacterium*. Therefore we decided to determine the phylogenetic distribution of the enzymes, and then to confirm their PRA isomerase activity. We specifically asked the question of whether the presence of these *trpF* genes could influence substrate specificity of the PriA homologues occurring in the same organisms. Thus, functional and detailed X-ray structural characterization of selected closely related PriA homologues (~88% sequence identity), representing different background scenarios, were undertaken. Notably, non-conserved residues present in the L5 loop were found to be responsible for functional shifts of up to two orders of magnitude. A novel enzyme sub-family, termed phosphoribosyl isomerase B (*priB* gene), which is functionally defined due its co-occurrence with *trpF*, is postulated. The implications of these functional and structural results are discussed in terms of the mechanisms driving enzyme functional evolution and how annotation could be advanced beyond sequence similarity.

EXPERIMENTAL PROCEDURES

Bioinformatics analysis

The sequences used for the phylogenetics analysis are included in Supplementary Tables S1 and S2. The sequences were aligned with Muscle version 3.6 and edited with Jalview. ProtTest v1.4 [17] was used to select out of 56 different models, the protein evolution model that best fit the protein alignments. The selected evolution model and its parameters were used for reconstruction of protein phylogenies using the Bayesian methods. The genome context analyses were done using the SEED [18].

Construction of a *trpF* (SSEG_02541) mutant of *Streptomyces svicieus*

The *Ssvi TrpF* strain was constructed using REDIRECT® [19]. The protocol, plasmids and strains were provided by PBL Biomedical Laboratories. The following oligonucleotides were used: GCTGTT CACACATACGAATTAGATACGGTGAGC GGC ATG ATT CCG GGG ATC CGT CGA CC and ACG CTG CCC CGG CCG AGC ACC GGC GCC TCC TGG GAA TCA TGT AGG CTG GAG CTG CTT C (the bases that are identical in the SSEG_02541 sequence are underlined). The disruption cassette was obtained by PCR using Taq DNA polymerase (New England Biolabs). SSEG_02541 was mutagenized by homologous recombination in the plasmid TOPOTA:FRegMgI, which contains SSEG_02541 plus the 1500 bp upstream and downstream regions. Double crossovers would replace the whole 588 bp of SSEG_02541 by a 1328 bp cassette containing the apramycin resistance cassette *aac(3)IV* and an origin of replication *oriT*. The resulting construct was introduced into *S. svicieus* by RP4-based conjugation [20] and selected for using apramycin (50 µg/ml). The presence of the mutated version of SSEG_02541 was detected by PCR, and sequencing was used to confirm the presence of the desired replacement.

PRA isomerase assay using cell-free extracts

Cell-free extracts (CFE) of *S. svicieus* and its *TrpF* mutant were prepared as follow. Spores were germinated in 10 ml YEME medium at 30°C. After three days, the cultures were harvested at 3000 × *g* at 4°C for 10 min, washed twice, and resuspended in minimal medium (50 ml) with or without tryptophan (50 µg/ml). The cultures were allowed to grow for four days at 30°C and then harvested at 3000 × *g* at 4°C for 10 min, washed and resuspended in 1.5 ml of 50 mM Tris/HCl (pH 8.0), 5% glycerol, 0.5 mM EDTA and 0.1 mM DTT. The cells were disrupted by sonication. The lysate was then centrifuged at 13000 × *g* for 10 min at 4°C. The CFE was taken for activity determination. The protein concentration in the CFE was determined by the Bradford method. The PRA isomerase activity assay was performed using PRA synthesized *in situ* after complete conversion of 30 µM anthranilic acid using 10 molar excess of phosphoribosyl pyrophosphate (PRPP) with TrpD from *Saccharomyces cerevisiae*, a total of 30 µM PRA was estimated to be produced. TrpC from *Thermotoga maritima* was added to avoid product inhibition. Both reactions (PRA production and its subsequent conversion) were followed fluorometrically in 96-well plates (Nuc 96-Well Optical Bottom Plates) in a TECAN infinite M1000 plate reader (excitation at 310 nm and emission at 400 nm) [10].

Complementation assays and enzyme kinetics of TrpF and PriA homologues

TrpF and PriA or PriB DNA sequences from *Mycobacterium smegmatis* (*priA*: NC_008596.1:3287922..3288698), *Streptomyces globisporus* (*priA*: AJUO01000211.1: 28477..29199, *trpF-2*: AJUO01000190.1: 39164..39790), *Streptomyces* sp. Mg1 (*priB*: NZ_DS570392.1:4082..4813, *trpF-3*: NZ_DS570471.1:15700..16290), *S. svicieus* (*priB*: NZ_CM000951.1:2602879..2603616, *trpF-3*: NZ_CM000951.1:4956766..4957383), *Streptomyces ipomoeae* (*priB*: AEJC01000296.1: 12777..13499, *trpF-3*: AEJC01000183.1:3975..4601), *Jonesia denitrificans* (*trpF-3*: NC_013174.1:954098..954712) and *Arthrobacter aurences* (*priB*: NC_008711.1:1753327..1754070, *trpF-3*: NC_008711.1:3892978..3893583) were amplified

by PCR using genomic DNA as a template. PriB from *S. sp. C* (*priB*: ACEW01000224.1:11860..11135) was synthesized by GeneART, where codons were optimized for its overexpression in *E. coli*. The products were cloned in pQE-30 (NdeI and HindIII) derivative (Qiagen) and pColdI (Takara, SacI and HindIII). *In vivo E. coli trpF* and *hisA* complementation assays were done as previously reported [9] with the exception that pQ3-30 (Qiagen) derivatives were used, and M9 minimal medium was enriched with a mixture of all the amino acids at 50 $\mu\text{g/ml}$ apart from L-histidine and L-tryptophan. pColdI constructions were used for protein expression, and enzyme purification using nickel affinity chromatography was performed as previously reported [10]. *In vitro* Michaelis–Menten enzyme kinetic parameters of both PRA and ProFAR isomerase activities were measured as previously reported using as control known enzymes, both active (positive control) and inactive (negative control) [11].

Gene cloning and protein expression and purification for crystallization

To overproduce TrpF-3 from *J. denitrificans*, PriA from *S. coelicolor* and PriB from *Streptomyces* sp. Mg1 and *Streptomyces svicens*, their cognate genes were subcloned into the plasmid pMCSG68 (Midwest Center for Structural Genomics). BL21Magic cells carrying the plasmid pMCSG68:TrpF-3_Jden, pMCSG68:PriA_Scoe, pMCSG68:PriB_SMg1 and pMCSG68:PriB_Ssvi respectively, were grown at 310 K and shaken at 200 $\text{rev}\cdot\text{min}^{-1}$ in enriched LB plus phosphate medium (pMCSG68:TrpF-3_Jden, pMCSG68:PriA_Scoe and pMCSG68:PriB_Mg1) or in a modified M9 salts ‘pink’ medium (pMCSG68:PriB_Ssvi) until OD_{600} of 1 was reached. The cultures were transferred to 277 K for 1 h. Subsequently, the cultures were transferred to 291 K and protein expression was induced with 0.5 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) alone for pMCSG68:TrpF-3_Jden, pMCSG68:PriA_Scoe, pMCSG68:PriB_Mg1 or with 150 $\text{mg}\cdot\text{l}^{-1}$ each of L-valine, L-isoleucine, L-leucine, L-lysine, L-threonine, L-phenylalanine, 90 $\text{mg}\cdot\text{l}^{-1}$ selenomethionine (Medicilon) and 1 mM IPTG for pMCSG68:PriB_Ssvi. The cells were incubated overnight, harvested and resuspended in lysis buffer [500 mM NaCl, 5%(v/v) glycerol, 50 mM HEPES–NaOH pH 8.0, 20 mM imidazole, 10 mM 2-mercaptoethanol]. The proteins were purified as described previously [21]. Specifically, the purification protocol included immobilized metal-affinity chromatography (IMAC) on an ÄKTAexpress system (IMAC-I: GE Healthcare Life Sciences) followed by His₆-tag cleavage using recombinant His-tagged TEV protease and a second IMAC step (IMAC-II) to remove the protease, the uncut protein and the affinity tag. The pure proteins were concentrated using Amicon Ultra filters (Millipore) in 20 mM HEPES–NaOH pH 8.0, 250 mM NaCl, 2 mM DTT.

Protein crystallization

Crystallization of SeMet-labelled PriB_Ssvi and native PriB_SMg1, PriA_Scoe and TrpF-3_Jden was performed at 16°C by sitting drop vapour diffusion technique in 96-well Crystal Quick plates (Greiner Bio-one), which were set up using a Mosquito liquid dispenser (TTP LabTech). The MCSG 1–4 screens from Microlytic were used to search for crystallization hits. For each condition, 0.4 μl of crystallization formulation and 0.4 μl of protein at a concentration of 10 mg/ml were mixed and then equilibrated against 140 μl of the reservoir solution. The crystals appeared under a number of conditions. The best apo

(throughout the text ‘apo’ refers to the structure with a ligand-free or sulfate/phosphate-bound active site) PriB_Ssvi crystals grew from a solution containing 1.8 M NaH₂PO₄/K₂HPO₄, pH 8.2 (PriB_Ssvi_apo). The PriB_Ssvi complex with AICAR ([5′-(5-aminoimidazole-4-carboxamide) ribonucleotide, PriB_Ssvi_AICAR) gave crystals from 0.2 M NaCl, 0.1 M imidazole/HCl, pH 8, 0.4 M NaH₂PO₄/1.6 M K₂HPO₄ and 2 mM ProFAR. The PriA_Scoe complex with AICAR (PriA_Scoe_AICAR) crystallized from a solution containing 0.2 M NaCl, 0.2 M sodium cacodylate, pH 6.5, and 2.0 M (NH₄)₂SO₄ with 2 mM ProFAR. PriB_SMg1 crystallized from 2 M (NH₄)₂SO₄, 0.1 M Bis/Tris pH 5.5 (PriB_SMg1_X1) and from 0.1 M Tris/HCl, pH 8.5, 1.5 M Li₂SO₄ and 2 mM rCdRP (PriB_SMg1_X2). The best TrpF-3_Jden crystals grew from a solution containing 0.1 M Bis/Tris propane:NaOH pH 7.0, 2 M ammonium citrate dibasic.

Diffraction data collection

The TrpF-3_Jden crystals were derivatized with [Ta₆Br₁₂]Br₂ to facilitate experimental phasing. Specifically, the crystals were soaked in a mixture of well solution (1.6 μ l) and 15 mM water suspension of the tantalum cluster (0.4 μ l). After 24 h, 0.2 μ l of tantalum cluster stock suspension were added to the drop. The crystals were soaked for an additional 3 days. The TrpF-3_Jden crystals (derivatized or native) did not require cryoprotection and were transferred to liquid nitrogen directly from the drop. Prior to data collection all PriB and PriA crystals were cryo protected in mother liquor supplemented with either glycerol (30%, PriB_Ssvi_apo 20% PriA_Scoe_AICAR) or 28% sucrose (PriB_Ssvi_AICAR, PriB_SMg1_X1, PriB_SMg1_X2) and subsequently plunged into liquid nitrogen. Diffraction data were collected at 100 K at the Structural Biology Center at the Advanced Photon Source, Argonne National Laboratory [22], either at 19-ID equipped with an ADSC quantum Q315r CCD detector (both PriB_Ssvi crystals, PriB_SMg1_X2, PriA_Scoe_AICAR and native TrpF_Jden) or 19-BM beamline (PriB_SMg1_X1 and TrpF-3_Jden derivative) equipped with a ADSC quantum Q210r CCD detector. For all PriB and PriA crystals, the datasets were collected at the wavelength corresponding to the selenium peak. For derivatized crystals of TrpF-3_Jden two datasets were collected at tantalum peak and inflection energies. The images were processed using the HKL3000 suite of programs [23]. For TrpF-3_Jden, data from three crystals were scaled together. Data collection statistics are summarized in Tables 3 and 4.

Structure solution and refinement

The TrpF-3_Jden structure was solved by multiple anomalous dispersion (MAD) approach as implemented in autoSHARP [24]. 12 tantalum sites corresponding to two clusters were localized and the results of phase determination are provided in Table 3. The initial model was built in Buccaneer [25]. The structure was further modified by iterations of manual rebuilding in the graphics program COOT [26] and fully anisotropic crystallographic refinement in PHENIX against the native data set with hydrogen atoms in riding positions [27]. The final structure of TrpF-3_Jden consists of residues Ala-0–Arg-204. The only residues that are missing from the model due to poor definition in the electron density maps belong to the cloning artefact (Ser-2–Asn-1). In addition to the protein chain, 266 water molecules and 3 citrate moieties have been identified. The $mF_o - DF_c$ difference map near

His-82 indicates a positive peak that may correspond to another disordered citrate molecule, however the density was not defined enough to allow confident modelling.

Both PriB_Ssvi structures were solved by single-wavelength anomalous diffraction (SAD) method. All procedures for heavy atom search in SHELXD [28], SAD phasing in SHELXE and MLPHARE [29], phase improvement by density modification in DM [30] and initial protein model building in Arp/wArp [31] were done by the structure module of the HKL3000 software package [23]. The structure of PriB_SMg1_X1 was solved by molecular replacement (MR) with the 1VZW PDB deposit [8] used as a template in Phaser [32]. The initial model was then improved by the automatic rebuilding protocol in Arp/wArp. The structure of PriB_SMg1_X2 was solved by MR with PriB_SMg1_X1 as a model. The PriA_Scoe_AICAR structure was solved by MR in Molrep [33] with the apo model (2VEP). All structures were further modified by iterations of manual rebuilding in the graphics program COOT [26] and crystallographic refinement in REFMAC5 [34] from the CCP4 suite [35] (PriB_Ssvi_apo) or PHENIX [27] (PriB_Ssvi_AICAR, PriB_SMg1). All structures contain a single protein molecule, with a different level of completeness. In the final PriB_Ssvi_apo model residues Ser-2–Leu-5 and Arg-143–Glu-148 have not been modelled due to poor electron density. Similarly, PriB_Ssvi_AICAR is missing N-terminal Ser-2–Asn-1, whereas PriB_SMg1_X1 does not include Ser-2–Val-3 and Gly-142–Glu-146. For the same reasons, in PriB_SMg1_X2, Ser-2–Pro-2 are omitted from the model. In the PriA_Scoe_AICAR the entire protein chain has been modelled starting from the N-terminus methionine. In addition to the protein chain, the PriB_Ssvi_apo structure contains two phosphate ions and 352 water molecules whereas PriB_Ssvi_AICAR has two sulfate ions, 336 water molecules and one molecule of the broken ProFAR, AICAR, an intermediate of the histidine biosynthesis pathway and a chemical degradation product of ProFAR (cleavage of C7–N6 bond) [36]. Similarly, PriA_Scoe_AICAR contains one AICAR molecule, two sulfate ions, 256 water molecules and 1 glycerol molecule. In the PriB_SMg1_X1 model four sulfate ions were included and 227 water molecules, whereas PriB_SMg1_X2 contains five sulfate ions and 229 water molecules. The rCdrP ligand added to crystallization is not bound to the protein molecule.

RESULTS

Identification of actinobacterial PRA isomerases using phylogenomics

To evaluate the co-occurrence of enzymes with PRA isomerase activity involved in histidine and tryptophan biosynthesis in *Actinobacteria*, we used comparative genomics together with phylogenetic reconstructions (Figures 2 and 3A). These analyses revealed that approximately 30% of the actinobacterial genome sequences, available at the time of our analyses, encoded *trpF* homologues. Bayesian phylogenetic reconstructions using the protein products of these *trpF* homologues, including selected sequences throughout the Eubacteria kingdom, allowed us to resolve at least three different classes of actinobacterial TrpF homologues. For the sake of clarity, but only within the present paper, we will refer to them as TrpF-1, TrpF-2 and TrpF-3. These sub-clades are consistent with two different bacterial phyla as previously defined by Jensen and co-workers [16]. The TrpF-1 sub-clade (red branch, Figure 2) represents the corynebacterial TrpF xenologues present in the

Gammaproteobacteria. Thus, and as introduced above, TrpF-1 co-occurs with a sub-functionalized PriA, subHisA, which has lost its PRA isomerase function [14,15].

The other two sub-clades (TrpF-2 and TrpF-3 branches, orange and green respectively) are grouped within the *Actinobacteria* (Figure 2). From these, the deep-rooted orange sub-clade contains TrpF homologues coming only from a limited number of actinobacterial genera, including *Nocardia*, *Actinokineospora* and *Streptomyces*. Gene context analyses revealed that these organisms show synteny in the surroundings of these *trpF* homologues. These loci include tryptophan biosynthetic homologues, namely, *trpD*, *trpA* and *trpC* genes, as well as an *ectC* homologous gene annotated as L-ectoine synthase, which together form an operon (Figure 3B). This operon is found within a previously reported natural products biosynthetic gene cluster, which directs the synthesis of a pyrrolopyrrole moiety in the actinobacterium *Micromonospora* sp. TP-A0468 (M468) [37], rather than L-tryptophan or L-ectoine. Indeed, this *trpF*-like enzyme homologue (*kstC3* gene or TrpF-2), has been shown to catalyse the isomerization of 4*N*'-(5'-phosphoribosyl) 4-aminopyrrole-2-carboxylate (PRaP) (orange branches, Figures 3A and 3C).

The remaining sub-clade within the *Actinobacteria* (green branches or TrpF-3) includes *trpF* genes found in distantly related genera, such as *Rhodococcus*, *Arthrobacter* or *Streptomyces*. Subsequent gene conservation analyses revealed that almost all of the *Arthrobacter* (100%) and *Rhodococcus* (95%) species conserve a *trpF* gene, although in *Streptomyces* the degree of conservation drops down to 25%. The differential occurrence of this *trpF* gene suggests an ancestral origin, and that it is indeed an orthologous TrpF enzyme (Figures 2 and 3A). However, further gene context analysis of these *trpF* orphan loci reveals a lack of synteny, including the absence of other *trp* genes (Figure 3B and Supplementary Table S3). Moreover, the *trpF-3* genes show high sequence diversity among orthologues, which contrasts with other L-Trp synthesis enzymes from their corresponding *trp* operons that lack a *trpF* gene. For instance, whereas the TrpF-3 enzymes share 65% sequence identity, other tryptophan biosynthetic enzymes such as TrpC share 80% sequence identity. These observations raise doubts about the involvement of the TrpF-3 proteins in tryptophan biosynthesis.

In order to understand the function of the *trpF-3* homologues, we designed and performed mutagenesis experiments. A gene knockout mutant of *S. sviveus* deficient for this gene was obtained after homologous gene displacement. The *S. sviveus trpF-3* strain was found to be prototrophic, that is, capable of growing on minimal medium without addition of L-tryptophan. Moreover, lack of auxotrophy was found to coincide with a 50% residual PRA isomerase activity in this mutant, as tested in cell-free extracts (Figure 4). Thus, it was concluded that the TrpF-3 enzymes are functional, and that they may support tryptophan biosynthesis. The same results were obtained when L-tryptophan was added to the medium, showing that the expression of this enzyme is not controlled by tryptophan. Therefore, it was hypothesized that the residual PRA isomerase activity can be explained by constitutive expression of the cognate PriA homologue present in this organism.

Biochemical analysis of PriA and TrpF homologues: identification of functional trade-offs

Having identified a previously unrecognized TrpF enzyme sub-family within *Actinobacteria*, which seems to be functional *in vivo*, we investigated how substrate specificities of co-occurring analogous PriA enzymes may be affected by the presence of TrpF-3. *In vivo* characterization of selected enzymes was conducted by testing the ability of any given PriA and TrpF homologues to complement *E. coli* mutants lacking the *trpF* (strain FBGW_f) or *hisA* (strain HfrG6) genes. In addition, when possible, some of them were purified to homogeneity and characterized, as previously described [10,14,15]. Moreover, conversion of PRA by the TrpF-2 homologue from *S. globisporus* (PRaP_Sglo) was investigated. An experiment using highly sensitive *in vivo* complementation assays based on high copy number plasmids with strong promoters, confirmed lack of PRA isomerase function of this TrpF-2, in agreement with its proposed functional divergent nature (Table 1).

Five TrpF-3 homologues from *J. denitrificans* (TrpF-3_Jden), *A. aurences* (TrpF-3_Aaur), *Streptomyces* sp. Mg1 (TrpF-3_SMg1), *S. svisceus* (TrpF-3_Ssvi) and *S. ipomoeae* (TrpF-3_Sipo) were characterized and shown to have PRA isomerase activity. These TrpF proteins were able to complement the *trpF* mutant of *E. coli*, which contrasts with the inability of these proteins to rescue growth of the *hisA E. coli* mutant (Table 1). However, we found that the catalytic efficiencies of the purified enzymes vary broadly. Whereas the kinetic parameters obtained for TrpF-3_Jden, TrpF-3_Aaur and TrpF-3_SMg1 were similar to those previously reported for TrpF from *E. coli* (TrpF_Ecoli) [12], TrpF-3_Sipo and TrpF-3_Ssvi were 6- and 12-fold less active respectively (Table 1 and Figure 5). Analysis of the kinetic parameters reveals that the observed variability seems to be related to slower turnovers (k_{cat}) rather than to bigger K_M values. This suggests that the tradeoff of PRA isomerase activities observed may not necessarily be related to substrate specificity.

We also found that all PriA homologues characterized here preserve dual-substrate specificity. First, the ProFAR and PRA isomerase activities of the PriA enzymes from organisms lacking a *trpF-3* gene, PriA_Sglo and PriA_Msme, were confirmed to be similar to that of *bona fide* PriAs. The discrepancy between data for PriA_Mtub [11] and that of PriA_Msme (85% identity), assayed here at optimal pH 7.5 [10], may relate to the fact that suboptimal pH (pH 8.5) was previously used for PriA_Mtub, leading to an underestimation of the K_M value of this enzyme [11]. Second, the ProFAR isomerase activities of the PriA homologues that co-occur with TrpF-3 enzymes are similar within this subgroup and also comparable to efficiencies of other PriA enzymes. In contrast, however, the PRA isomerase activity of this sub-set ranges from $0.0022 \mu\text{M}^{-1} \cdot \text{s}^{-1}$ to $0.14 \mu\text{M}^{-1} \cdot \text{s}^{-1}$ (Table 2 and Figure 5). From this point onwards, therefore, the sub-family of PriA homologues present in organisms whose genomes encode a *trpF-3* gene will be referred to as PriB, from *phosphoribosyl isomerase B*.

The abovementioned data are consistent with the results obtained from the *S. svisceus trpF-3* gene knockout experiment, and provide an enzyme system suitable for gaining insights into functional trade-offs during enzyme co-evolution. For instance, catalytic efficiency for conversion of PRA by PriB_Ssvi and PriB_SspC is one and two orders of magnitude, respectively, below that of the ProFAR isomerase activity of all characterized PriA enzymes. This clear difference contrasts with the high sequence identity shared by these enzymes,

which ranges from 88% to 92%, accounting for only 12–20 variable amino acids. Moreover, the K_M and k_{cat} parameters for PriB enzymes suggest that their reduced PRA isomerase activity is due to poorer specificity for PRA, as indicated by a 4–20-fold increase in K_M . Thus, in contrast to the changes at the level of k_{cat} seen in TrpF-3 enzymes, the observed functional tradeoff of PriBs relates to changes in K_M .

X-ray crystallographic structural analysis of PriB: functional and evolutionary implications

To place functional insights into structural perspective, we elucidated the structures of TrpF-3 and PriB enzymes. We were able to solve one structure of a TrpF-3, that from *J. denitrificans* [1.09 Å (1 Å=0.1 nm), TrpF-3_Jden, PDB 4WUI, Table 3], as well as high-resolution structures of two PriB homologues from *Streptomyces*. The PriB structures elucidated include PriB_SMg1 in two apo¹ forms (PriB_SMg1_X1 and PriB_SMg1_X2), and PriB_Ssvi in both its apo form (PriB_Ssvi_apo) and in a complex with [5'-(5-aminoimidazole-4-carboxamide) ribonucleotide or AICAR (PriB_Ssvi_AICAR), an intermediate of the histidine biosynthesis pathway and a chemical degradation product of the ProFAR substrate (cleavage of C7 N6 bond) (Table 4). In addition, we have *de novo* determined the structure of the PriA homologue from *S. coelicolor*, but in complex with AICAR (PriA_Scoe_AICAR). All the structures are monomeric ($\beta\alpha$)₈ barrels and have clear electron density along the entire chain with the exception of the N-termini (in all structures) and fragments of the L5 loop (in PriB_Ssvi_apo and PriB_SMg1_X1). Moreover, the structures obtained in the absence of ligand mimics contain phosphate or sulfate ions in positions corresponding to the phosphate groups of ProFAR or PRA.

Although we aimed at obtaining structures of both TrpF and PriB enzymes from the same organism, in order to complement our functional enzyme trade-offs analyses within the same cell, this was not possible despite comprehensive screening for crystallization conditions. We therefore focused on comparing the structures of PriBs with PriA structures from *S. coelicolor* and *M. tuberculosis* (PDBs 2Y85, 2Y88, 2VEP and 5DN1). Structural analysis using these PDBs revealed subtle differences between PriA and PriB sub-families in the active site region (Figure 6). The changes include different positions of residues away from the catalytic framework, with the most pronounced rearrangements and various degrees of order seen in the L5 loop. Interestingly, although loops 1 and 6 have been shown to adopt different conformations in PriA_Mtub, defining the apo and halo states [15,16], the PriA and PriB structures from *Streptomyces* species do not show these movements. So we focused on the several interactions that contribute to conformation and position of the L5 loop in PriB proteins, as further described below, with an emphasis on the contacts related to ProFAR and PRA substrates.

First, we looked into the similarities between PriA and PriB with regards to the ProFAR-specific L5 loop states. As introduced, in the PRFAR-bound state (and most likely ProFAR-bound state, as suggested by recent comparisons with newly available HisA structures, [38]) the L5 loop of PriA_Mtub assumes an ordered β -hairpin structure with a knot-like conformation at its tip. Although different amino acid residues are present in PriA and PriB, similar arrangement of the main chain is observed in PriA_Scoe_AICAR and PriB_Ssvi_AICAR (Figure 6A). Specifically, in all three structures the loop adopts a

geometry consistent with type I' β -turn, involving residues Gly-142–Trp-145 (same numbering for Ssvi and Mtub, Gly-138–Trp-141 in Scoe) and with the main-chain (C=O)–main-chain (N) hydrogen bond between the flanking residues. Moreover, in the knot-like conformation of PriB_Ssvi_AICAR and PriA_Scoe_AICAR a salt bridge involving Arg-143 (Arg-139 in Scoe) and Glu-113 (Glu-109 in Scoe) is present (Figure 6A). In PriB_Ssvi_AICAR, the loop conformation is stabilized by a salt bridge between Lys-141 from β 5a and Glu-148 from β 5b, whereas in PriA_Mtub_PRFA, Lys-141 is replaced by Arg-141 making an equivalent interaction via the guanidinium group of Arg-141, which forms hydrogen bonds with the main chain carbonyl groups of Trp-145 and Glu-146. Overall, this structural analysis shows conservation of the ProFA-specific structural elements, despite lack of sequence identity.

Second, we tried to identify differences between PriA and PriB in the relaxed conformation states, which may provide insights into PRA binding. As PriB_SMgl_apo_X1 shows less detail than the X2 form, our structural analysis was done on the latter structure, which also displays two conformations of the L5 loop, one of them disordered within the β 5b strand. In the PriB_SMgl_apo-observed ordered state, the L5 loop adopts a relaxed hairpin conformation with a type II β -turn, as defined previously [39,40], which is stabilized by a hydrogen bond between Gly-140 and the backbone of Trp-143 (C = O ... N distance 3.2 Å, Figure 6B). This observation suggests a reduction in the intrinsic mobility of the PriB enzymes. In contrast, the L5 loop structure of PriA_Scoe, which is more flexible [9], does not exactly correspond to the shape observed in PriB_SMgl_apo. In PriA_Scoe, Arg-137 replaces Lys-139, and a hydrogen bond is observed between the guanidinium group of Arg-137 from β 5a (through N ϵ) and the main-chain carbonyl group of the neighbouring Gly-138, equivalent to Gly-140 in PriB_SMgl_apo (Figure 6B). Consequently, in PriA_Scoe, Gly-138 cannot participate in the interaction stabilizing the β -turn between β 5a and β 5b that we report herein for PriB. Thus, the latter strand remains fully flexible in PriA, allowing for both the knot-like conformation needed for ProFA conversion, as well as for the sliding of loop L5 into the active site to allow the functionally essential Arg-143 to promote PRA binding. We therefore propose that this is actually not possible in PriB due to the interaction between Gly-140 and Trp-143, which albeit weak, appears to provide sufficient support to reduce the β 5b movement, and thus preventing the L5 loop from sliding inside the active site.

Based on the previous structural observations, we aimed at gaining evolutionary insights into the relationship between the functional trade-offs reported in the previous section and L5 conformational states. We were able to identify key differences between PriA and PriB, namely, Arg141Lys, Arg147Ser and Asp148Glu (58% frequency, numbering of PriB_Ssvi). For instance, these mutations are found in PriB_SMgl, PriB_SspC, PriB_Ssvi, but not in PriB_Sipo, which has in these positions two arginine and glutamate residues respectively, characteristic of PriA_Scoe (Figure 6C). It should be noted that these observations would have been obscured using only PriA_Mtub as a structural reference due to taxonomic distance between *Streptomyces* and *Mycobacterium*. Indeed, in agreement with this observation, the PriB_Sipo homologue has the highest PRA isomerase activity among the tested PriBs, but is at least 2-fold lower than PriA_Scoe and PriA_Sglo. These changes may have some consequences for the L5 loop behaviour, but this cannot be easily assessed from

available structural static data, suggesting that enzyme dynamics plays a key role in evolutionary processes, as we have previously suggested for subHisA [14].

DISCUSSION

An ancestral origin of *trpF-3* is supported by the occurrence of an orthologue in the genome of the deep-rooted organism *Acidimicrobium ferrooxidans*, which has been shown to have a HisA monofunctional enzyme [7]. Its wide taxonomic distribution hints towards differential loss of TrpF-3 as the most parsimonious explanation for the evolution of this gene.

Worthwhile to note is the observation that in most *Streptomyces* species, including those with a *trpF-3* gene, the *trp* genes tend to cluster with the *his* genes. The only exception is *trpF-3* itself, which is present in a genomic region unrelated to tryptophan biosynthesis. Although PRA isomerase activity could be demonstrated for TrpF-3, these enzymes may be evolving other yet-to-be discovered functions, resembling adaptation of *trpF-2* for synthesis of a natural product (*kstC3*; [37]). Unfortunately, the sole TrpF-3 structure that we were able to elucidate (TrpF-3_Jden, PDB 4WUI) is far too distant from available TrpF structures, hampering our ability to perform informative structural comparisons.

Given the different specificities of PriA and PriB, it would be interesting to compare the regulatory mechanisms linking tryptophan and histidine biosynthesis in their cognate genetic backgrounds. The corresponding biosynthetic enzymes in *S. coelicolor*, where a PriA enzyme exists, have been reported to be constitutively co-expressed [41,42]. This scenario contrasts with that found in *Corynebacterium*, where a *trpF-1* gene acquired after HGT prompting loss of PRA specificity in the co-evolving PriA homologues, is feedback regulated. As a result, it is possible that the model organism of this genus, *C. glutamicum*, evolved pathway-specific regulatory mechanisms [43,44], which is only feasible in the absence of substrate ambiguity. Our results suggesting lack of feedback regulation in *S. sviveus* wild type, and its *trpF* mutant, where TrpF-3 and PriB are co-evolving, seems a suitable system to start addressing this issue by means of metabolomics analysis.

Co-evolution of PRA isomerase activities encoded by *trpF-3* and *priB* homologues showed a functional tradeoff. Low PRA isomerase activity in PriB seemed to be compensated by more active TrpF enzymes, as their combined activities reach levels similar to those found in PriAs. These observations would be in agreement with previous theoretical models of quantitative, rather than qualitative, subfunctionalization of duplicate (or analogous) genes [45]. However, retention of PRA isomerase activity encoded by *trpF-3*, as a response to suboptimal activity in PriB enzymes for reaching the required metabolic flux, disagrees with the notion that ‘imperfect’ enzymes are still orders of magnitude more efficient than needed [46]. Co-evolution of TrpF-3 and PriB, therefore, may be driven by the interplay of neutral evolution, environmental constraints and the need for higher metabolic fluxes. All these factors have been implicated in intragenic epistasis [47], which is expected to take place during functional shifts. Within this context, analogous enzyme genes may increase exploration of sequence space in search of an optimal metabolic flux.

One could expect then, that if TrpF-3 fully specializes into another metabolic function, a PriA enzyme, with proficient dual-substrate specificity, has to co-evolve with the

specialization of TrpF. The residues interaction networks discovered after our structural analyses suggest that indeed PriB from *S. ipomoeae*, PriB_Sipo, may be evolving towards a PriA, as its L5 loop includes residues found conserved in the latter. This is in agreement with the fact that loss of PRA isomerase activity in PriB is associated with modifications of the L5 loop, which contributes to substrate interactions. The loop mutations are very similar chemically, and their influence on the loop conformation was only apparent after detailed structural analysis. Even though it is difficult to attribute specific roles of the side chains, it is clear that Lys/Glu combination distorts the PRA-specific loop conformation, which may explain reduced affinities. These findings suggest that modification of the loop dynamics [13], even if these mutations are non-conserved, mediates the dual-substrate specificity of PriB. Indeed, the amino acids found are in agreement with recent ancestral sequence reconstructions of HisA and PriA enzymes, reported while the present paper was under review [48].

Among currently known bacteria, *Actinobacteria* show the highest diversity within the HisA enzyme family: HisA [7], PriA [6], subHisA [14], and now PriB. Remarkably, these sub-families could not be discovered by sequence comparisons. On one hand, high sequence similarity does not necessarily mean identical enzyme functions, which results in divergent enzyme families. On the other hand, it is also true that enzymes with no sequence similarity can perform the same catalytic conversions. Thus, in the absence of functional and structural data, which cannot be readily obtained for large sets of homologues, our results hint towards what data could be used to assign enzyme function more accurately. Our functional and structural characterization of co-evolving TrpF and PriB enzymes, identified after phylogenomic and comparative genomic analyses, implies that genomic data, including gene context, phylogenetic occurrence, as well as regulatory motifs, may be of great value for functional annotation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We acknowledge Hilda E. Ramos-Aboites for technical support.

FUNDING

This work was supported by the Conacyt Mexico [grant number 179290 (to F.B.-G.)]; the National Institutes of Health [grant number GM094585 (to A.J.)]; the U.S. Department of Energy, Office of Biological and Environmental Research [grant number DE-AC02*06CH11357].

Abbreviations

AICAR	5'-(5-aminoimidazole-4-carboxamide) ribonucleotide
CdRP	1-(<i>O</i> -carboxyphenylamino)-1-deoxyribulose-5-phosphate
CFE	cell-free extracts
CHR	chorismate

HGT	horizontal gene transfer; IMAC, immobilized metal-affinity chromatography
IPTG	isopropyl β -D-1-thiogalactopyranoside
PRA	phosphoribosylanthranilate
PRaP	4N'-[(5'-phosphoribosyl) 4-aminopyrrole-2-carboxylate
PRFAR	4N'-[(5'-phosphoribulosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide
ProFAR	4N'-[(5'-phosphoribosyl)formimino]-5-aminoimidazole-4-carboxamide ribonucleotide
PRPP	phosphoribosyl pyrophosphate

REFERENCES

1. Jurgens C, Strom A, Wegener D, Hettwer S, Wilmanns M and Sterner R (2000) Directed evolution of a (beta alpha)8-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A* 97, 9925–9930 [PubMed: 10944186]
2. Henn-Sax M, Thoma R, Schmidt S, Hennig M, Kirschner K and 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 [PubMed: 12356303]
3. Nasvall J, Sun L, Roth JR and Andersson DI (2012) Real-time evolution of new genes by innovation, amplification, and divergence. *Science* 338, 384–387 [PubMed: 23087246]
4. Patrick WM and Matsumura I (2008) A study in molecular contingency: glutamine phosphoribosylpyrophosphate amidotransferase is a promiscuous and evolvable phosphoribosylanthranilate isomerase. *J. Mol. Biol.* 377, 323–336 [PubMed: 18272177]
5. Evran S, Telefoncu A and Sterner R (2012) Directed evolution of (betaalpha)(8)-barrel enzymes: establishing phosphoribosylanthranilate isomerisation activity on the scaffold of the tryptophan synthase alpha-subunit. *Protein Eng. Des. Sel* 25, 285–293 [PubMed: 22490958]
6. Barona-Gomez F and Hodgson DA (2003) Occurrence of a putative ancient-like isomerase involved in histidine and tryptophan biosynthesis. *EMBO Rep.* 4, 296–300 [PubMed: 12634849]
7. Noda-García L, Juárez-Vázquez AL, Ávila-Arcos MC, Verduzco-Castro EA, Montero-Morán G, Gaytán P, Carrillo-Tripp M and Barona-Gómez F (2015) Insights into the evolution of enzyme substrate promiscuity after the discovery of (β α)8 isomerase evolutionary intermediates from a diverse metagenome. *BMC Evol. Biol* 15, 107 [PubMed: 26058375]
8. Kuper J, Doenges C and Wilmanns M (2005) Two-fold repeated (betaalpha)4 half-barrels may provide a molecular tool for dual substrate specificity. *EMBO Rep.* 6, 134–139 [PubMed: 15654319]
9. Wright H, Noda-Garcia L, Ochoa-Leyva A, Hodgson DA, Fulop V and Barona-Gomez F (2008) The structure/function relationship of a dual-substrate (betaalpha)8-isomerase. *Biochem. Biophys. Res. Commun* 365, 16–21 [PubMed: 17967415]
10. Noda-Garcia L, Camacho-Zarco AR, Verdel-Aranda K, Wright H, Soberon X, Fulop V and Barona-Gomez F (2010) Identification and analysis of residues contained on beta $\rightarrow$ alpha loops of the dual-substrate (beta alpha)8 phosphoribosyl isomerase A specific for its phosphoribosyl anthranilate isomerase activity. *Protein Sci.* 19, 535–543 [PubMed: 20066665]
11. Due AV, Kuper J, Geerlof A, von Kries JP and 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 [PubMed: 21321225]

12. List F, Sterner R and Wilmanns M (2011) Related (betaalpha)8-barrel proteins in histidine and tryptophan biosynthesis: a paradigm to study enzyme evolution. *Chembiochem* 12, 1487–1494 [PubMed: 21656890]
13. Ochoa-Leyva A, Barona-Gomez F, Saab-Rincon G, Verdel-Aranda K, Sanchez F and Soberon X (2011) Exploring the structure-function loop adaptability of a (beta/alpha)(8)-barrel enzyme through loop swapping and hinge variability. *J. Mol. Biol.* 411, 143–157 [PubMed: 21635898]
14. Noda-Garcia L, Camacho-Zarco AR, Medina-Ruiz S, Gaytan P, Carrillo-Tripp M, Fulop V and Barona-Gomez F (2013) Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Mol. Biol. Evol.* 30, 2024–2034 [PubMed: 23800623]
15. Noda-Garcia L and Barona-Gomez F (2013) Enzyme evolution beyond gene duplication: a model for incorporating horizontal gene transfer. *Mol. Genet. Elements* 3, e26439 [PubMed: 24251070]
16. Xie G, Bonner CA, Song J, Keyhani NO and 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 [PubMed: 15214963]
17. Abascal F, Zardoya R and Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21, 2104–2105 [PubMed: 15647292]
18. Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crecy-Lagard V, Diaz N, Disz T, Edwards R et al. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702 [PubMed: 16214803]
19. Gust B, Challis GL, Fowler K, Kieser T and Chater KF (2003) PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci. U.S.A* 100, 1541–1546 [PubMed: 12563033]
20. Keijsers BJ, van Wezel GP, Canters GW, Kieser T and Vijgenboom E (2000) The ram-dependence of *Streptomyces lividans* differentiation is bypassed by copper. *J. Mol. Microbiol. Biotechnol.* 2, 565–574 [PubMed: 11075933]
21. Kim Y, Dementieva I, Zhou M, Wu R, Lezondra L, Quartey P, Joachimiak G, Korolev O, Li H and Joachimiak A (2004) Automation of protein purification for structural genomics. *J. Struct. Funct. Genomics* 5, 111–118 [PubMed: 15263850]
22. Rosenbaum G, Alkire RW, Evans G, Rotella FJ, Lazarski K, Zhang RG, Ginell SL, Duke N, Naday I, Lazarski J et al. (2006) The Structural Biology Center 19ID undulator beamline: facility specifications and protein crystallographic results. *J. Synchrotron Radiat* 13 (Pt 1), 30–45 [PubMed: 16371706]
23. Minor W, Cymborowski M, Otwinowski Z and Chruszcz M (2006) HKL-3000: the integration of data reduction and structure solution—from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* 62 (Pt 8), 859–866 [PubMed: 16855301]
24. Vonrhein C, Blanc E, Roversi P and Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol. Biol.* 364, 215–230 [PubMed: 17172768]
25. Cowtan K (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. Biol. Crystallogr.* 62 (Pt 9), 1002–1011
26. Emsley P and Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60 (Pt 12 Pt 1), 2126–2132 [PubMed: 15572765]
27. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* 66 (Pt 2), 213–221 [PubMed: 20124702]
28. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr. A* 64 (Pt 1), 112–122 [PubMed: 18156677]
29. Otwinowski Z (1991), Maximum likelihood refinement of heavy atom parameters. *Proceedings of the CCP4 Study Weekend*, 25–26 1 1991
30. Cowtan K (1994) Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography. 31, 34–38
31. Langer G, Cohen SX, Lamzin VS and Perrakis A (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc* 3, 1171–1179 [PubMed: 18600222]

32. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC and Read RJ (2007) Phaser crystallographic software. *J. Appl. Crystallogr* 40 (Pt 4), 658–674 [PubMed: 19461840]
33. Vagin A and Teplyakov A (1997) MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr* 30, 1022–1025
34. Sankaran K and Wu HC (1994) Lipid modification of bacterial prolipoprotein. Transfer of diacylglycerol moiety from phosphatidylglycerol. *J. Biol. Chem* 269, 19701–19706 [PubMed: 8051048]
35. CCP4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr* 50 (Pt 5), 760–763 [PubMed: 15299374]
36. Klem TJ and Davisson VJ (1993) Imidazole glycerol phosphate synthase: the glutamine amidotransferase in histidine biosynthesis. *Biochemistry* 32, 5177–5186 [PubMed: 8494895]
37. Ma HM, Zhou Q, Tang YM, Zhang Z, Chen YS, He HY, Pan HX, Tang MC, Gao JF, Zhao SY, Igarashi Y and Tang GL (2013) Unconventional origin and hybrid system for construction of pyrrolopyrrole moiety in kosinostatin biosynthesis. *Chem. Biol* 20, 796–805 [PubMed: 23790490]
38. Söderholm A, Guo X, Newton MS, Evans GB, Näsvall J, Patrick WM and Selmer M (2015) Two-step ligand binding in a ($\beta\alpha$)₈ barrel enzyme-substrate-bound structures shed new light on the catalytic cycle of HisA. *J. Biol. Chem* 290, 24657–24668 [PubMed: 26294764]
39. Venkatachalam CM (1968) Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units. *Biopolymers* 6, 1425–1436 [PubMed: 5685102]
40. de Beer TA, Berka K, Thornton JM and Laskowski RA (2014) PDBsum additions. *Nucleic Acids Res.* 42, D292–D296 [PubMed: 24153109]
41. Hu DS, Hood DW, Heidstra R and 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 [PubMed: 10361288]
42. Charaniya S, Mehra S, Lian W, Jayapal KP, Karypis G and Hu WS (2007) Transcriptome dynamics-based operon prediction and verification in *Streptomyces coelicolor*. *Nucleic Acids Res.* 35, 7222–7236 [PubMed: 17959654]
43. Brune I, Jochmann N, Brinkrolf K, Huser AT, Gerstmeir R, Eikmanns BJ, Kalinowski J, Puhler A and 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 [PubMed: 17259312]
44. Jung S, Chun JY, Yim SH, Lee SS, Cheon CI, Song E and Lee MS (2010) Transcriptional regulation of histidine biosynthesis genes in *Corynebacterium glutamicum*. *Can. J. Microbiol* 56, 178–187 [PubMed: 20237580]
45. Lynch M and Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* 154, 459–473 [PubMed: 10629003]
46. Newton MS, Arcus VL and Patrick WM (2015) Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes. *J. Royal Soc. Interface* 12, 20150036
47. Sanchez-Ruiz JM (2012) On promiscuity, changing environments and the possibility of replaying the molecular tape of life. *Biochem. J* 445, e1–3 [PubMed: 22702975]
48. Plach MG, Reisinger B, Sterner R and Merkl R (2016) Long-term persistence of bi-functionality contributes to the robustness of microbial life through exaptation. *PloS Genet* 12, e1005836 [PubMed: 26824644]

Figure 1. ($\beta\alpha$)₈ barrel isomerases involved in L-tryptophan and L-histidine biosynthesis
(A) Isomerization reactions in L-tryptophan and L-histidine biosynthetic pathways. L-tryptophan (blue) biosynthesis, which starts from chorismate (CHR), and L-histidine (red) biosynthesis, starting from ATP and PRPP, is shown. The committed reaction catalysed by PriA or phosphoribosyl isomerase A in *Actinobacteria* (dashed arrows), is independently catalysed by the enzymes TrpF or PRA isomerase, and HisA or ProFAR isomerase (dotted arrows) in most bacteria. **(B)** PriA from *M. tuberculosis* (PriA_Mtub) with L5 loop assuming a relaxed conformation upon binding of rCdRP. **(C)** PriA from *M. tuberculosis* (PriA_Mtub) with L5 loop assuming a twisted knot-like conformation upon binding of PRFAR. All the known active site residues [15] are shown in both complexes.

Figure 2. Phylogenetic tree of bacterial TrpF sequences

The seven different TrpF sub-clades are indicated by the names of the group of bacteria from which the sequences were obtained. Red (TrpF-1), orange (TrpF-2) and green (TrpF-3) branches represent TrpF sequences obtained from *Actinobacteria*. *Gammaproteobacteria* contain xenologues present in certain species belonging to the genus *Corynebacterium* (TrpF-1). See Supplementary Table S1 for details on the sequences included in the phylogenetic reconstruction.

Figure 3. Diversity of TrpF in *Streptomyces*

(A) RpoB, GyrB and AtpD concatenated phylogeny. Numbers at nodes are the approximate likelihood ratio test supporting each branch. The branches from *Streptomyces* that have the *trpF* gene are shown in orange and green. *Kitasatospora setae* (*kset*) was used as an out-group (see also Supplementary Table S2 for nomenclature of the branches). (B) Genome neighbourhood analysis of *Streptomyces trpF* genes. Arrows with the same colour represent orthologous genes. Colour nomenclature is only for illustrative purposes and does not indicate functional relationships (see also Supplementary Table S3). (C) Proposed enzyme activities for KstC3 or TrpF-2, PRaP isomerase (based on [37]), orange branches; TrpF-3, PRA isomerase (the present study), green branches.

Figure 4. PRA isomerase activity of cell free extract

PRA isomerase activity of cell free extract of wild type *S. svicens* (WT) and its *trpF* deletion mutant ($\Delta trpF$) in minimal media (MM) and with L-Trp supplementation (+Trp). PRA isomerase activity as the reduction in fluorescence (excitation at 310 nm and emission at 400 nm). Each data point represents three biological replicates and three technical replicates. Student *t* test, with an alpha level of 0.05.

Figure 5. PRA and ProFAR isomerase catalytic efficiencies of TrpF and PriA homologues
The detailed enzyme kinetic parameters are provided in Tables 1 and 2. The detection limits for the PRA and ProFAR isomerase assay used in the present study are $0.0001 \mu\text{M}^{-1} \cdot \text{s}^{-1}$ and $0.001 \mu\text{M}^{-1} \cdot \text{s}^{-1}$, respectively. An asterisk (*) symbol is used to denote proteins that were structurally compared in the present study.

Figure 6. X-ray structure and sequence analysis of PriB and PriA

(A) Structure of L5 loop with a ProFAR-related ligand from PriB_Ssvi_AICAR (PDB 4TX9), PriA_Scoe_AICAR (PDB 5DN1) and PriA_Mtub_PRFAR (PDB 2Y88, D11N mutant). For clarity only a fragment of the ligand is shown for PriA_Mtub_PRFAR. (B) Structure of the L5 loop in the apo or PRA-specific conformation from PriB_SMg1_apo_X2 (PDB 4X9S), PriA_Scoe_apo (PDB 2VEP) and PriA_Mtub_rCdRP (PDB 2Y85). The numbering is indicated as the actual position for each protein sequence. (C) Multi-sequence alignments of L5 loop (green) and the surrounding residues of PriA and PriB. Substrate-anchoring arginine and tryptophan residues are marked with an asterisk. Highly conserved

residues in PriB are framed. The numbering is indicated as the position of the residues in PriA_Mtub and PriB_Ssvi.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Functional characterization of TrpF homologues

Enzyme	<i>In vivo</i> activity		<i>In vitro</i> activity [*]				Reference
	HisA	TrpF	PRA isomerase				
			K_M (μ M)	k_{cat} (s^{-1})	k_{cat}/K_M (μ M $^{-1} \cdot s^{-1}$)		
TrpF-1_Ecoli	–	+	28.9	33.8	1.2		[13]
TrpF-3_Iden	–	+	16.8 $\pm$ 3.3	27 $\pm$ 1.6	1.6		This work
TrpF-3_Aaur	–	+	9.3 $\pm$ 0.4	29.2 $\pm$ 5.2	3.3		This work
TrpF-3_SMgl	–	+	8.4 $\pm$ 1.7	10.5 $\pm$ 2.4	1.25		This work
TrpF-3_Ssvi	–	+	9.4 $\pm$ 0.47	0.82 $\pm$ 0.08	0.088		This work
TrpF-3_Sipo	–	+	23.0 $\pm$ 1.3	4.3 $\pm$ 0.35	0.19		This work
TrpF-2_Sglo	–	–	Not detected				This work

* Each data point comes from at least three independent determinations using freshly purified enzyme.

Table 2

Functional characterization of PriA and PriB homologues

Enzyme	<i>In vivo</i> assay		<i>In vitro</i> activity*				PRA isomerase			Reference
	HisA	TrpF	ProFAR isomerase		k_{cat}/K_M ($\mu M^{-1} \cdot s^{-1}$)	PRA isomerase				
			Complementation	K_M (μM)		k_{cat} (s^{-1})	K_M (μM)	k_{cat} (s^{-1})		
HisA_Afer	+	−	1.1 ± 0.2	0.05 ± 0.001	0.045		Not detected			[7]
PriA_Blon	+	+	2.7 ± 0.5	0.4 ± 0.1	0.1	6.1 ± 0.1	2.1 ± 0.5	0.3		[7]
PriA_Sglo	+	+	4.2 ± 0.8	0.74 ± 0.03	0.18	11 ± 1.0	3.8 ± 0.2	0.34		This work
PriA_Scoe	+	+	3.6 ± 0.7	1.3 ± 0.2	0.4	5.0 ± 0.08	3.4 ± 0.09	0.7		[10]
PriA_Mtub	+	+	19	0.23	0.012	21	3.6	0.17		[11]
PriA_Msme	+	+	2.6 ± 0.5	0.85 ± 0.04	0.33	7.9 ± 2.4	3.1 ± 0.43	0.39		This work
PriA_Cjei	+	+	2.3 ± 0.2	0.9 ± 0.08	0.4	5.1 ± 1.0	1.6 ± 0.16	0.3		[14]
PriB_Aaur	+	+	2.1 ± 0.5	1.8 ± 0.2	0.9	26.3 ± 6.3	0.37 ± 0.09	0.014		This work
PriB_Sipo	+	+	3.8 ± 0.2	0.82 ± 0.02	0.21	60.8 ± 1.1	8.25 ± 0.4	0.14		This work
PriB_Ssvi	+	+	3.9 ± 0.89	0.69 ± 0.04	0.18	24.5 ± 4.0	1.6 ± 0.29	0.067		This work
PriB_SspC	+	+	11.4 ± 3.4	2.53 ± 0.74	0.22	149.9 ± 29	1.4 ± 0.12	0.009		This work
PriB_SMgl	+	+	13.2 ± 3.4	0.92 ± 0.19	0.069	129.6 ± 34	0.29 ± 0.04	0.0022		This work

*. Each data point comes from at least three independent determinations using freshly purified enzyme.

Table 3

Data processing and refinement statistics for TrpF-3 from *J. denitrificans*

Data collection			
Crystal	TrpF-3_Jden Ta peak	TrpF-3_Jden Ta inf	TrpF-3_Jden native
Space group	$P2_12_12_1$		$P2_12_12_1$
Cell dimensions [Å]	$a = 50.47, b = 65.41, c = 67.11$		$a = 49.44, b = 64.31, c = 66.40$
Temperature [K]	100		100
Radiation source	APS 19-BM		APS 19-ID
Wavelength [Å]	1.2554	1.2559	0.9793
Resolution [Å] *	30.00–2.00(2.03–2.00)	30.00–2.00(2.03–2.00)	25.00–1.09(1.11–1.09)
Unique reflections	15492 (746)	15581 (757)	88243 (4156)
R_{merge}	0.085 (0.556)	0.130 (above 1)	0.066 (0.795)
$\langle I \rangle / \langle \sigma \rangle$	16.7 (2.2)	9.6 (0.9)	32.3 (2.4)
Completeness [%]	99.4 (100)	99.2 (98.4)	99.6 (95.7)
Redundancy	4.7 (4.1)	4.4 (3.3)	9.6 (6.4)
Phasing (for resolution range 27.38–2.00 Å)			
Phasing power (acentric/centric)	1.39/1.09		
Phasing power (anomalous)	3.55	1.94	
FOM (acentric/centric)	0.66/0.34		
Refinement			
Resolution [Å]			24.71–1.09
Reflections work/test set			86373/1782
$R_{\text{work}} / R_{\text{free}}$			0.128/0.138
No. of atoms protein/ligand/water			1591/39/266
Average B factor [Å ²] protein/ligand/water			13.2/20.2/29.8
Rmsd			
Bond lengths [Å]			0.011
Bond angles[°]			1.514
Ramachandran plot [%]			
Favoured			96.74

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Outliers	0.5
Molprobability score	1.11
Clashscore	1.25
PDB code	4WUI

* Values in parentheses correspond to the highest resolution shell.

$R_{\text{merge}} = \frac{\sum h \sum j |I_{hj} - \langle I_{hj} \rangle|}{\sum h \sum j I_{hj}}$, where I_{hj} is the Intensity of observation j of reflection h .

$R = \frac{\sum h |F_o| - |F_c|}{\sum h |F_o|}$ for all reflections, where F_o and F_c are observed and calculated structure factors respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement.

Table 4

Data processing and refinement statistics for PriA homologues

Data collection				
Crystal	PriB_Ssvi_apo	PriB_Ssvi_AICAR	PriB_SMgl_X1	PriB_SMgl_X2
Space group	$P6_522$	$P4_32_12$	$P4_32_12$	$P3_12$
Cell dimensions [Å]	$a = 69.11, c = 175.04$	$a = 72.55, c = 143.29$	$a = 69.33, c = 126.93$	$a = 69.55, c = 127.32$
Temperature [K]	100	100	100	100
Radiation source	APS 19-ID	APS 19-ID	APS 19-BM	APS 19-ID
Wavelength [Å]	0.9793	0.9793	0.9793	0.9792
Resolution [Å] *	59.85–1.33 (1.38–1.33)	30.00–1.60 (1.63–1.60)	30.00–1.57 (1.60–1.57)	50–1.95 (1.98–1.95)
Unique reflections	55033 (3659)	51524 (2548)	43584 (2100)	19264 (936)
R_{merge}	0.071 (0.834)	0.078 (over 1)	0.066 (over 1)	0.148 (0.591)
$\langle I \rangle / \langle \sigma \rangle$	26.5 (2.0)	27.8 (2.0)	34.9 (2.1)	18.7 (2.0)
Completeness [%]	94.74 (64.61)	99.7 (99.5)	98.9 (98.0)	100 (100)
Redundancy	9.1 (6.7)	7.8 (7.8)	11.7 (8.1)	4.6 (4.1)
Refinement				
Resolution [Å]	26.63–1.33	26.84–1.60	27.86–1.57	28.94–1.60
Reflections work/test set	52240/2793	50137/1319	42291/1249	40416/1299
$R_{\text{work}} / R_{\text{free}}$	0.127 / 0.151	0.152/0.181	0.155/0.184	0.155/0.190
No. of atoms protein/ligand/water	1862/10/352	1870/32/336	1752/20/227	1820/25/229
Average B factor [Å ²] protein/ligand/water	15.3/20.5/29.9	24.0/22.0/34.3	28.1/49.3/38.2	24.4/28.5/33.4
Rmsd				
Bond lengths [Å]	0.006	0.014	0.011	0.005
Bond angles [°]	1.24	1.62	1.44	0.98
Ramachandran plot [%]				
Favoured	98	98	98	97
Outliers	0	0	0	0
Molprobit score	1.29	0.77	0.69	1.4
Clashscore	4.01	0.53	0.57	4.7
PDB code	4U28	4TX9	4W9T	5DNI

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

* Values in parentheses correspond to the highest resolution shell.

[†] $R_{\text{merge}} = \Sigma h \Sigma j |I_{hj} - \langle I_{hj} \rangle| / \Sigma h \Sigma j I_{hj}$, where I_{hj} is the intensity of observation j of reflection h .

[‡] $R = \Sigma h |F_O - F_C| / \Sigma h F_O$ for all reflections, where F_O and F_C are observed and calculated structure factors, respectively. R_{free} is calculated analogously for the test reflections, randomly selected and excluded from the refinement.