

Evolutionary diversification of protein–protein interactions by interface add-ons

Maximilian G. Plach^a, Florian Semmelmann^a, Florian Busch^b, Markus Busch^a, Leonhard Heizinger^a, Vicki H. Wysocki^b, Rainer Merkl^{a,1}, and Reinhard Sterner^{a,1}

^aInstitute of Biophysics and Physical Biochemistry, University of Regensburg, D-93040 Regensburg, Germany; and ^bDepartment of Chemistry and Biochemistry, The Ohio State University, Columbus, OH 43210

Edited by David Baker, University of Washington, Seattle, WA, and approved August 18, 2017 (received for review May 3, 2017)

Cells contain a multitude of protein complexes whose subunits interact with high specificity. However, the number of different protein folds and interface geometries found in nature is limited. This raises the question of how protein–protein interaction specificity is achieved on the structural level and how the formation of nonphysiological complexes is avoided. Here, we describe structural elements called interface add-ons that fulfill this function and elucidate their role for the diversification of protein–protein interactions during evolution. We identified interface add-ons in 10% of a representative set of bacterial, heteromeric protein complexes. The importance of interface add-ons for protein–protein interaction specificity is demonstrated by an exemplary experimental characterization of over 30 cognate and hybrid glutamine amidotransferase complexes in combination with comprehensive genetic profiling and protein design. Moreover, growth experiments showed that the lack of interface add-ons can lead to physiologically harmful cross-talk between essential biosynthetic pathways. In sum, our complementary *in silico*, *in vitro*, and *in vivo* analysis argues that interface add-ons are a practical and widespread evolutionary strategy to prevent the formation of nonphysiological complexes by specializing protein–protein interactions.

protein–protein interactions | protein interfaces | interface add-on | glutamine amidotransferases | protein evolution

Protein–protein interactions are essential for key cellular processes, ranging from the formation of molecular machineries to the assembly of signal transduction networks. A huge number of interactions have evolved to accomplish the various biological functions. For example, experimental and *in silico* methods projected the interactome of yeast to comprise ~18,000 binary protein–protein interactions (1). In the light of such dense protein networks, it is not trivial for a cell to guarantee interaction specificity in crucial cases such as toxin–antitoxin, antibody–antigen, protease inhibitor, or multienzyme complexes. This problem is aggravated by the limited number of different interface geometries that mediate protein–protein interactions. It is estimated that not more than 1,000 such geometries exist (2, 3) and that their number is restricted by the same biophysical constraints that limit the number of protein folds (4, 5).

Understanding the principles of protein–protein interactions in general, and the assurance of interaction specificity despite the limited number of interface geometries in particular, is an important biological challenge. A recent *in silico* analysis indicated that relatively small insertions and deletions in protein interfaces can differentiate between monomers and homodimers, and that these elements may preclude undesired interactions (6). However, the interfaces in homooligomers are unique with respect to amino acid composition and residue-to-residue contact preferences and differ significantly from those found in other types of protein complexes such as heterooligomers (7). For the latter, interaction specificity is most crucial in cases where two or more proteins with similar interface geometries exist that compete for the same interaction partner (8). An example would be a complex A:B, in which A can interact with several homologous potential interaction partners B, B', and B''. Often, the interfaces of B' and

B'', whose binding has to be avoided, are similar to that of the genuine partner B, creating the risk of erroneous and potentially harmful A–B' and A–B'' interactions.

To contribute to the understanding of interaction specificity in heterooligomers, we started with a systematic *in silico* survey of the interfaces from 305 representative heteromeric protein complexes. Not considering small differences in secondary structure elements or slightly different quaternary structures, in about 10% of this sample, interface geometries vary significantly between related complexes that share homologous subunits. In these cases, interfaces are extended by additional loops and entire secondary structure elements that contain residues crucial for complex stability. We designated these elements “interface add-ons” and presumed that they differentiate interfaces between related complexes that share homologous subunits, and thus contribute to interaction specificity, much like additional bits turn a master key into a special key.

To substantiate this assumption, we comprehensively analyzed protein interaction specificities in a family of glutamine amidotransferase complexes (GATases) that are part of the tryptophan and folate biosynthesis pathways. These heteromeric enzyme complexes comprise glutaminase and synthase subunits, which interact to transfer ammonia from glutamine to an acceptor substrate (9). The synthase subunits of these GATases, as well as the glutaminase subunits, respectively, are homologs, share high sequence similarity, and belong to the same folds. A subset of synthase subunits exclusively involved in tryptophan biosynthesis contains an interface add-on, which is absent in all other homologous synthase subunits, including those exclusively involved in folate biosynthesis. We experimentally characterized 54 combinations of nine synthases (three containing the interface add-on) with six different glutaminases, as well as a rationally

Significance

Proteins adopt no more than a thousand folds, and the number of different protein–protein interface geometries is restricted to around 1,000, too. Given this limited structural repertoire, it has remained elusive how hundreds of thousands of specific protein–protein interactions evolved and how unspecific interactions are avoided. We report on a strategy to solve this dilemma, which is the integration of additional structural elements at the interface periphery that guarantee specificity. We named these elements “interface add-ons” to reflect the benefit they provide to protein interfaces, as software add-ons do to web browsers or as additional bits turn a master key into a special key.

Author contributions: M.G.P., V.H.W., R.M., and R.S. designed research; M.G.P., F.S., and F.B. performed research; M.G.P., F.B., M.B., and L.H. contributed new reagents/analytic tools; M.G.P., F.S., F.B., L.H., V.H.W., R.M., and R.S. analyzed data; and M.G.P., F.S., F.B., V.H.W., R.M., and R.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence may be addressed. Email: rainer.merk1@ur.de or reinhard.sterner@ur.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1707335114/-DCSupplemental.

designed synthase with a deletion in its interface add-on. Our results show that glutaminase–synthase interaction specificity is determined by the presence or absence of the interface add-on, independent of the phylogenetic origin of the proteins and their participation in tryptophan or folate biosynthesis.

The profiling of more than 15,000 bacterial and archaeal genomes highlights the greater biological relevance of this finding: Most species possess two homologous synthases for tryptophan and folate biosynthesis, and the lack of interface add-ons in these subunits enables the binding of the same glutaminase. In contrast, in species that possess a synthase with an interface add-on and a second synthase without an interface add-on, an additional, specifically adapted glutaminase is present. We assume that positive selection favored this diversification of the synthases as it allowed for an effective separation of tryptophan and folate biosynthesis. In vivo experiments show that this separation is physiologically relevant as its override is detrimental for cellular growth.

Results

Identification of Interface Add-Ons. Interfaces of protein complexes can be partitioned into highly conserved core and more variable rim regions (10, 11). We speculated that this variability can lead to diverse peripheral geometries that might contribute to protein–protein interaction specificities within heteromeric complexes. For a systematic computational assessment of interface geometries, we combined structural information from Protein Data Bank (PDB) entries with the sequences provided by the corresponding InterPro families, which contained, on average, 12,000 homologs. Our protocol (Fig. 1A) comprises seven filter routines that systematically refine the specification of an interface add-on and simultaneously narrow down the number of candidates (SI Appendix, Table S1).

The survey was based on those 1,739 heteromeric, bacterial protein complex structures deposited in the PDB that were devoid of nonprotein macromolecules and had subunit stoichiometries of AB, A₂B₂, A₃B₃, A₄B₄, A₆B₆, ABC, and A₂B₂C₂. Removing

identical proteins and focusing on structures associated with superordinate family entries in InterPro reduced the number to 305 complexes. In the following, we name them “reference complexes” and use *SU* to address one of the subunits A, B, or C. For each subunit, *SU*, and all its InterPro homologs, *H*, we computed pairwise sequence alignments, PW(*SU*, *H*) (Fig. 1B), and eliminated all PW(*SU*, *H*) of poor quality as well as heavily fragmented ones. Subsequently, those PW(*SU*, *H*) were identified that showed in *SU* or *H* an additional fragment containing at least eight residues. Thus, our approach excludes minor alterations of interface topologies that are related to shorter indels and identifies fragments that can fold into defined secondary structure elements.

For the computational analysis of the remaining PW(*SU*, *H*), two cases had to be distinguished. First, *SU* can contain one or more additional fragments (insertions) that are not present in the aligned homolog *H*. Second, *H* can contain one or more additional fragments that are not present in *SU* (deletions). To assess the first case, the insertions resulting from all PW(*SU*, *H*) were mapped onto the sequence of *SU* and a histogram was computed that specifies the number of insertions overlapping each residue position (Fig. 1B). Subsequently, only those insertions were considered further that exceeded a predefined significance threshold; compare the gray area in the histogram shown in Fig. 1B and the examples in SI Appendix, Fig. S1A. Thus, 209 insertions were found in 117 of the reference complexes, with most insertions comprising between 10 and 20 amino acids (Fig. 1C). To assess the second case, a histogram specifying the number of deletions beginning at each residue position of *SU* was analyzed analogously. In total, we identified 392 deletions associated with 212 reference complexes; examples are shown in SI Appendix, Fig. S1B. A detailed characterization of the corresponding insertions in *H* is difficult because their structures are not known and the local topology of such large insertions is often unreliable in homology models (12). Moreover, the subsequent

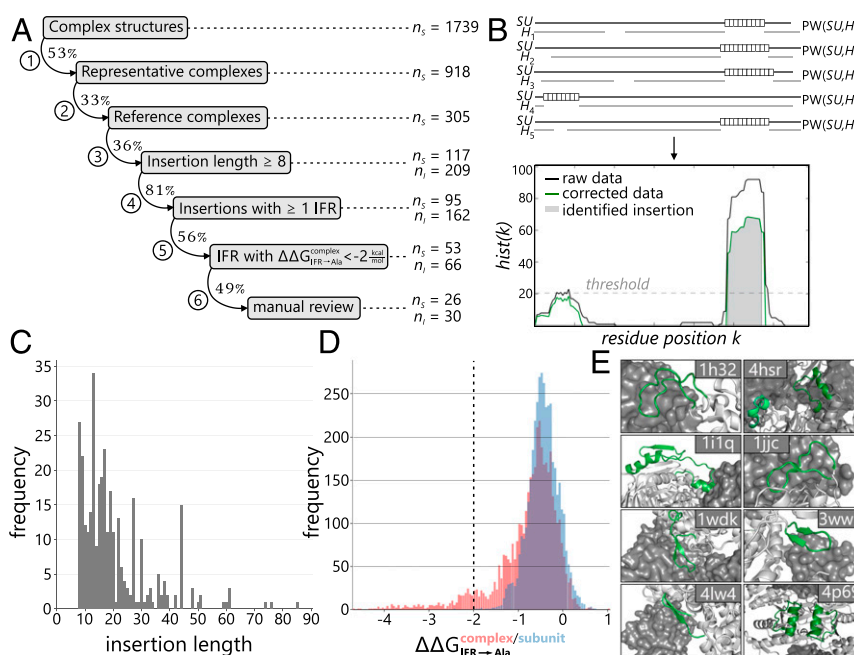


Fig. 1. Survey of interface add-ons in heteromeric protein complexes. (A) Step-by-step refined identification of interface add-ons in 1,739 bacterial, heteromeric protein complexes. The abbreviations n_s and n_i refer to the numbers of structures and insertions, respectively, remaining after the different steps of our analysis. Details are given in SI Appendix, Table S1; the final set of interface add-ons is listed in SI Appendix, Table S2. (B) Analysis of pairwise alignments PW(*SU*, *H_i*). Insertions of at least eight residues (white boxes) were mapped onto each residue position k of *SU*, and resulting $hist(k)$ counts were plotted. (C) Histogram of insertion lengths. (D) Histogram of predicted protein–protein affinity change values $\Delta\Delta G_{\text{complex/subunit}}^{\text{complex}}$ (red) and $\Delta\Delta G_{\text{complex/IFR} \rightarrow \text{Ala}}^{\text{subunit}}$ (blue) for all IFRs of all insertions. The dashed vertical line indicates the threshold of -2 kcal/mol used for classifying interface add-ons. (E) Examples of heteromeric protein complexes that contain interface add-ons (green).

docking of the model into the given reference complex introduces further errors; thus, we did not further examine these cases.

In contrast, the 209 fragments that correspond to an insertion in *SU* could be analyzed in detail. To begin with, we identified 162 insertions in 95 reference complexes that contained at least one interface residue (IFR). Next, we assessed in silico the contribution of these insertions to complex stability. The algorithm mCSM estimates the effect of mutations on protein stability and protein–protein affinity, and the resulting $\Delta\Delta G$ values correlate well with experimental findings (13). We used mCSM for an alanine scanning of all 162 insertions by mutating each nonalanine IFR individually to alanine. According to mCSM classification, a large number of the $IFR \rightarrow Ala$ mutations are highly destabilizing for the complex ($\Delta\Delta G_{IFR \rightarrow Ala}^{complex} < -2$ kcal/mol; note that negative values indicate destabilizing mutations) (Fig. 1D). Consistently, alanine mutations of IFRs that are crucial for complex stability (hot spots) typically result in similar $\Delta\Delta G$ values (14). It is important to note that these effects are not caused by destabilized subunits, because 90% of the corresponding $\Delta\Delta G_{IFR \rightarrow Ala}^{subunit}$ values deduced from subunit structures are above -1.0 kcal/mol (Fig. 1D).

Assuming that interface add-ons contribute little to the stability of the subunits but considerably to the stability of the complex, we classified an insertion as an interface add-on if at least one $\Delta\Delta G_{IFR \rightarrow Ala}^{complex}$ value was below -2 kcal/mol. After a final manual control, 30 interface add-ons remained in 26 of the 305 reference complexes (SI Appendix, Table S2). Assuming that the reference complexes are a representative sample of the full structural repertoire, it can be estimated that $\sim 10\%$ of bacterial protein complexes contain interface add-ons. This estimation might be skewed because protein complexes with simpler stoichiometry (e.g., AB, A_2B_2) are over-represented in the PDB, and thus constitute 69% of the reference complexes. However, the finding that the interfaces in each of the seven stoichiometry groups that contribute to the reference complexes contain, on average, $9.5 \pm 3.9\%$ add-ons argues against such a bias.

Interface Add-Ons in Heteromeric, Bacterial Protein Complexes. Interface add-ons comprise, on average, 23 amino acids, of which 63%

are involved in protein–protein interactions and generally form well-defined secondary-structure elements (50% α -helices, 11.5% β -strands, and 38.5% loop regions). Typically, interface add-ons are characterized by a strong internal sequence conservation (sequence logo bit scores >3) and are present in about 8% of the respective InterPro homologs. The corresponding complexes are engaged in a variety of key biological functions in amino acid and pyrimidine biosynthesis, β -oxidation of fatty acids, biosynthesis of aminoacyl-tRNAs, sulfur metabolism, respiratory chains, biosynthesis of antibiotics, and the citric acid cycle. Eight examples of interface add-ons are shown in Fig. 1E and described in detail in SI Appendix, Table S3.

In the following, we paradigmatically characterized the interface add-on found in anthranilate synthases (AS). These enzyme complexes are heterotetrameric GATases, which catalyze the initial step in the biosynthesis of the essential amino acid tryptophan. Their glutaminase subunits, TrpG, hydrolyze glutamine to glutamate. The concomitantly formed ammonia is subsequently channeled to the synthase subunits, TrpE, where it reacts with chorismate (CH) to anthranilate (AA) (Fig. 2A). The interface add-on is located in TrpE and, with a length of 51 residues, is one of the largest add-ons identified in our survey. It folds into two α -helices connected by two β -strands and prominently protrudes into the TrpE:TrpG dimer interface with loop L2 and α -helix H2 (Fig. 2B). Although the interface add-on as a whole contains only a few conserved residues, the two α -helices contain several invariable hydrophobic (H1) and charged/polar (H2) amino acids, respectively (Fig. 2C). In contrast to the AS from *Salmonella typhimurium*, those from *Mycobacterium tuberculosis* and *Sulfolobus solfataricus*, for example, contain TrpE subunits without this interface add-on. In the following, we thus use the term “TrpE” to refer to AS synthase subunits that do not contain the interface add-on and use the term “TrpEx” (“extended”) to refer to AS synthase subunits that contain the interface add-on.

Some organisms not only possess AS (TrpEx₂:TrpG₂) but also the related complex aminodeoxychorismate synthase (ADCS; PabB:PabA), which catalyzes the first step of folate biosynthesis

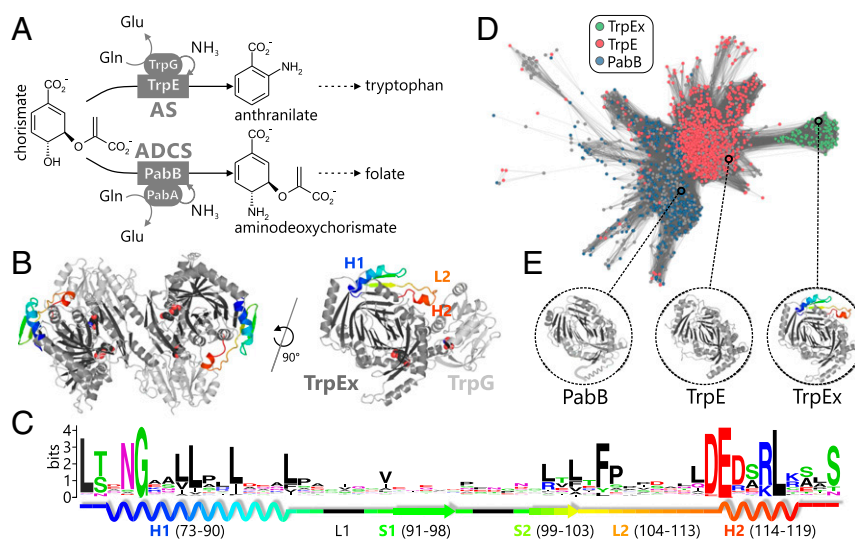


Fig. 2. Sequence–structure–function relationship in the TrpE(x)/PabB family. (A) Reactions catalyzed by AS and ADCS in tryptophan and folate biosynthesis. Both the glutaminase subunits (ellipses) and the synthase subunits (rectangles) are homologs. (B) TrpEx₂:TrpG₂ AS complex from *S. typhimurium* (PDB ID code 1i1q) and one TrpEx:TrpG heterodimer. The interface add-on is colored in a rainbow gradient. Active sites are indicated by superimposed space-filling ligand models. (C) Sequence logo of the TrpEx interface add-on, with its 2D elements indicated. Numbering is based on PDB ID code 1i1q. (D) Main cluster of the TrpE (x)/PabB SSN (IPR019999) generated at an E-value cutoff of 1E-77. It contains all TrpE(x) and most TrpE/PabB sequences. Nodes are colored according to the annotation of InterPro (TrpE or PabB). Gray nodes represent sequences with ambiguous annotation. TrpEx nodes are colored green after manual identification of the interface add-on in nodes annotated as TrpE. (E) Crystal structures of TrpEx from *S. typhimurium* (PDB ID code 1i1q), TrpE from *M. tuberculosis* (PDB ID code 4pen), and PabB from *E. coli* (PDB ID code 1k0e, a helix that is not resolved is sketched in cartoon representation).

(15) (Fig. 2A). Its synthase subunit, PabB, is homologous to TrpEx, but lacks the interface add-on, and its glutaminase subunit, PabA, is homologous to TrpG. A sequence similarity network (SSN) of the InterPro family subsuming TrpEx, TrpE, and PabB sequences shows that TrpE and PabB proteins, which both do not contain the interface add-on, are more similar in sequence to each other than to TrpEx, as they tightly cluster together (Fig. 2D). In contrast, TrpEx sequences are grouped into a distinct subcluster due to their common interface add-on and their close phylogenetic relationship (discussed in the next section). Removal of the interface add-on region from TrpEx sequences increases their similarity to TrpE or PabB sequences (SI Appendix, Fig. S2A, B, and E), and a statistical analysis of the corresponding SSNs indicates that the interface add-on is more conserved than the rest of the TrpEx sequences (SI Appendix, Fig. S2D). This sequence-based characterization of the interface add-on is further emphasized by the comparison of representative TrpEx, TrpE, and PabB structures, indicating that the interface add-on is also the only major structural difference between these three homologs (Fig. 2E).

Three observations suggest that the interface add-on in TrpEx acts as an explicit negative design element to differentiate TrpEx from PabB, and thus ensure the specific formation of AS (TrpEx₂:TrpG₂) and ADCS (PabB:PabA) complexes: first, its location in the interface between TrpEx and TrpG; second, the high conservation of charged and polar residues in its interface helix H2, a characteristic of insertions that modulate the association of protein complexes (6); and, third, the fact that some organisms like *Escherichia coli* contain homologous ADCS complexes whose glutaminase subunit, PabA, is highly similar to the TrpG glutaminase subunit of AS. Consequently, the interface add-on in TrpEx may prevent the putative cross-interactions between TrpEx-PabA and PabB-TrpG, and thus the formation of nonphysiological complexes.

Phylogenetic Distribution of AS and ADCS Complexes. In contrast to the genomes of γ -Proteobacteria like *E. coli* and *S. typhimurium* that contain the four genes coding for the individual AS and ADCS complexes, the genome of *Bacillus subtilis*, for example, contains the genes for TrpE and PabB but only a single glutaminase gene, annotated as *pabA*. This specific situation in *B. subtilis* has been investigated extensively (16, 17), and it has been shown that the single PabA glutaminase serves both TrpE and PabB. To get an overview of such gene cooccurrences, we determined phylogenetic distributions of *trpEx*, *trpE*, *trpG*, *pabB*, and *pabA* across more than 15,000 bacterial and archaeal species. To this end, we developed a computational genetic profiling routine (SI Appendix, Fig. S3A), which uses BLAST and hidden Markov models (HMMs) to find, annotate, and distinguish the homologous TrpEx/TrpE/PabB synthases and TrpG/PabA glutaminases in these species with high sensitivity and selectivity (SI Appendix, Fig. S3B and C). HMM approaches have proven successful to distinguish highly similar sequences or to find distantly related homologs (18, 19).

In brief, we created five HMMs to represent each type of synthase and each type of glutaminase. We applied BLAST to scan bacterial and archaeal genomes for homologs of the synthases (glutaminases) and compared the respective three (two) HMM scores of the hits to assign them as TrpEx, TrpE, or PabB (TrpG or PabA). The dataset was made nonredundant, and sequences with low HMM scores (ambiguous assignments) were rejected. Eventually, we determined the cooccurrences of synthases and glutaminases for 1,463 species that constitute the TrpEx subcluster (TrpEx species) and for 4,386 species that constitute the TrpE subcluster (TrpE species) of the SSN shown in Fig. 2D. The phylogenetic distribution of the cooccurrences was derived by mapping them onto a tree of life comprising a representative set of archaeal and bacterial species.

As expected, the majority of TrpEx species (84%) possess both of the synthase–glutaminase pairs TrpEx–TrpG and PabB–PabA (SI Appendix, Fig. S4A). In not more than 16% of the genomes, corresponding genes were present in multiple copies or missing (SI Appendix, Table S4). Most of the TrpEx species are γ -Proteobacteria and belong preferentially to the orders Vibrionales and Enterobacteriales (SI Appendix, Fig. S4B). The evolutionary oldest TrpEx species are *Shewanella*, an offspring of Pseudomonadales and Xanthomonadales. Most plausibly, TrpEx and TrpG have emerged in *Shewanella* and have been conserved in the γ -proteobacterial lineage since then. Apart from γ -Proteobacteria, TrpEx and/or TrpG is only present in *Helicobacter pylori*, *Pseudomonas aeruginosa*, and the Corynebacteria *Corynebacterium diptheriae*, *Corynebacterium efficiens*, and *Corynebacterium glutamicum*, which is a result of horizontal gene transfer (20, 21).

All other species, including archaea and all major bacterial phyla, are TrpE species. Fifty-nine percent of them (e.g., *B. subtilis*) possess TrpE and PabB, but only PabA and no TrpG glutaminases (SI Appendix, Fig. S4A). Some TrpE species (23%) lack PabB and contain only TrpE and PabA. Among them are Euryarchaeota and Crenarchaeota, which lack the classical folate-biosynthesis genes and use alternative biosynthetic pathways or rely on methanopterin-related methyl donors instead of tetrahydrofolate (22, 23). In not more than 18% of TrpE-species genomes, corresponding genes were present in multiple copies or missing (SI Appendix, Table S4). Taken together, TrpEx species generally contain a full set of synthases (TrpEx, PabB) and glutaminases (TrpG, PabA). In contrast, TrpE species generally contain one or both synthases (TrpE and/or PabB) and only one type of glutaminase (PabA). Consequently, in TrpE species, PabA has to interact with both synthases, whereas a more specific interaction pattern seems plausible for TrpEx species.

Protein–Protein Interaction Specificity in AS and ADCS Complexes.

We experimentally analyzed the influence of the TrpEx interface add-on on the specificity of glutaminase–synthase interactions in AS and ADCS complexes. For this purpose, we expressed and purified three TrpEx, two TrpE, and four PabB representatives from a phylogenetically diverse group of species: TrpEx from *S. typhimurium* (stTrpEx), *E. coli* (ecTrpEx), and *Serratia marcescens* (smTrpEx); TrpE from *Pseudomonas putida* (ppTrpE) and *S. solfataricus* (ssTrpE); and PabB from *P. putida* (ppPabB), *S. typhimurium* (stPabB), *E. coli* (ecPabB), and *B. subtilis* (bsPabB). The identity of the synthases was validated by an HPLC assay showing that all bona fide TrpEx, TrpE, and PabB enzymes formed the expected AA and ADC, respectively (SI Appendix, Fig. S5). Furthermore, we expressed and purified two TrpG and four PabA homologs: TrpG from *S. typhimurium* (stTrpG) and *E. coli* (ecTrpGD), and PabA from *P. putida* (ppPabA), *S. marcescens* (smPabA), *E. coli* (ecPabA), and *B. subtilis* (bsPabA). The ecTrpG could only be solubly expressed as a fusion construct with TrpD. The three TrpEx synthases share sequence identities between 72% and 88%, and the five synthases without an interface add-on (TrpE, PabB) share identities between 29% and 45%, with one outlier (ecPabB-stPabB) reaching 76% due to a close phylogenetic relationship. The identities between TrpEx and TrpE/PabB synthases are between 25% and 35%. Sequence identities between different glutaminases are around 97% (TrpG–TrpG), 65% (PabA–PabA), and 42% (TrpG–PabA) (SI Appendix, Table S5).

We analyzed the ability of the synthases to form complexes with the glutaminases by analytical size exclusion chromatography in combination with static light scattering (SEC-SLS). We made three striking observations, which are schematically illustrated in Fig. 3A. First, all three TrpEx synthases exclusively interacted with the two TrpG glutaminases. These combinations resulted in tetrameric complexes, in accordance with the oligomeric states deduced from AS crystal structures (PDB ID codes 1i1q and 1i7q). In other words, none of the TrpEx synthases

formed complexes with any of the PabA glutaminases. The absence of any interaction between TrpEx and PabA was confirmed by native mass spectrometry for the stTrpEx/ppPabA combination at a concentration similar to those that produced native TrpEx:TrpG complexes (*SI Appendix, Fig. S6*). Second, all TrpE and PabB synthases interacted with all PabA but none of the TrpG glutaminases. The TrpE:PabA complexes were dimers or tetramers, whereas the PabB:PabA complexes were exclusively dimeric. The complexes ppTrpE:ppPabA, ppPabB:ppPabA, and ecPabB:ecPabA were validated by native mass spectrometry, as was the absence of a complex for the ppTrpE-stTrpG combination (*SI Appendix, Fig. S6*). Third, interactions between TrpEx and TrpG on the one hand and between TrpE/PabB and PabA on the other hand are conserved across species and kingdom borders. All 15 tested enzymes were able to interact with a partner from a different organism, independent of their positions in the tree of life. For instance, TrpE from the crenarchaeon *S. solfataricus* formed complexes with PabA from phylogenetically distant Firmicutes (bsPabA) and γ -Proteobacteria (ppPabA, smPabA, and ecPabA).

Functional Characterization of Cognate and Hybrid AS and ADCS Complexes. We followed the glutamine-dependent conversion of chorismate (CH) to AA and aminodeoxychorismate (ADC) for all 54 possible combinations of synthases and glutaminases by continuous fluorimetric assays. The results are summarized in Fig. 3B. Notably, all stable complexes detected by SEC-SLS, except those

containing bsPabB, were catalytically active. AS complexes displayed catalytic efficiencies, $k_{\text{cat}}/K_m^{\text{CH}}$, between $8.7 \times 10^3 \text{ M}^{-1}\cdot\text{s}^{-1}$ and $1.3 \times 10^6 \text{ M}^{-1}\cdot\text{s}^{-1}$ with practically no differences in the highest efficiencies of TrpEx:TrpG and TrpE:PabA complexes. ADCS complexes (PabB:PabA) converted CH to ADC with catalytic efficiencies, $k_{\text{cat}}/K_m^{\text{CH}}$, between $9.9 \times 10^2 \text{ M}^{-1}\cdot\text{s}^{-1}$ and $2.8 \times 10^4 \text{ M}^{-1}\cdot\text{s}^{-1}$. These data show that functional AS and ADCS complexes can be formed by synthases and glutaminases from different species. This indicates a strong conservation of the synthase–glutaminase interface, because functional complexes require efficient channeling of nascent ammonia between the subunits. We could not detect measurable AS or ADCS activity for any of the “noninteracting” pairs. The only exception was ssTrpE, which did not only display AS activity with all four PabA glutaminases but also, to a certain degree, with ecTrpGD, indicating the formation of a transient complex during catalysis.

It is known that glutaminase activity in GATases is allosterically stimulated by the synthase (24–26). To quantify this effect for the 54 combinations of glutaminases and synthases, we incubated TrpG and PabA with glutamine and monitored its hydrolysis before and after the addition of TrpEx, TrpE, or PabB (*SI Appendix, Fig. S7*). The determined stimulation factors are illustrated in Fig. 3C. We found that both TrpG representatives do not display measurable activity in the absence of TrpEx, in accordance with previous findings on stTrpG (27). The presence of each of the TrpEx synthases, however, leads to apparent turnover rates for glutamine hydrolysis of up to 0.07 s^{-1} . In most

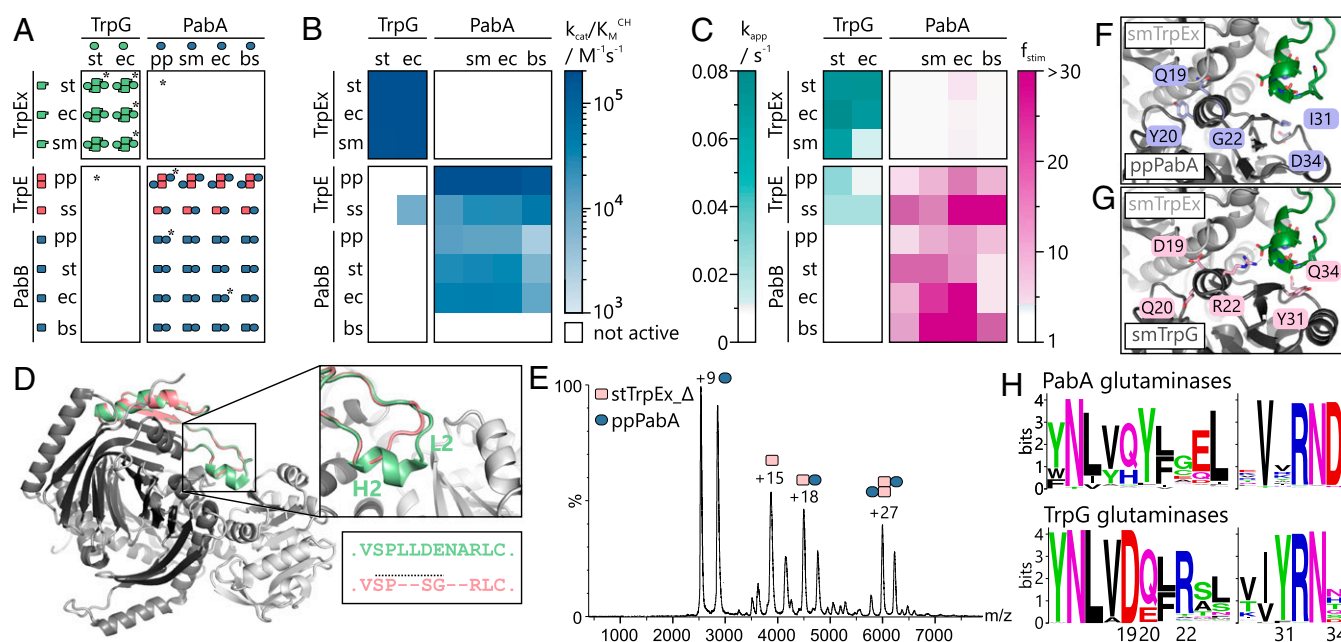


Fig. 3. Characterization of synthase–glutaminase interactions as well as design principles and characterization of stTrpEx Δ and ppPabA*. (A) Oligomeric states of glutaminases, synthases, and cognate and hybrid complexes as determined by SEC-SLS and native mass spectrometry (*). Blank spaces indicate no complex formation. Molecular weights and mass spectra are provided in *SI Appendix, Tables S6 and S7* and in *SI Appendix, Fig. S6*. (B) Catalytic efficiencies for the glutamine-dependent conversion of CH to AA (combinations involving TrpEx/TrpE) and ADC (combinations involving PabB). Each combination was assayed in triplicate. Exact values with SDs are provided in *SI Appendix, Table S8*. The bsPabB did not display ADCS activity with any of the available glutaminases under the applied experimental conditions. (C) Apparent turnover rates of glutamine hydrolysis by TrpG [apparent glutamine hydrolysis rate (k_{app})] and stimulation of PabA glutamine hydrolysis by TrpEx, TrpE, or PabB [stimulation factor (f_{stim})]. Each combination was assayed in triplicate in the presence of 4 mM glutamine. Exact values with SDs are provided in *SI Appendix, Table S9*. (D) Overlay of a stTrpEx:stTrpG dimer (light gray) with a model of stTrpEx Δ (dark gray). The cutout shows a detailed view of the modified interface add-on part (green in stTrpEx, red in stTrpEx Δ). The motif LLDENA in stTrpEx was replaced with SG to mimic a type I β -turn. (E) Representative mass spectrum of an equimolar mixture of stTrpEx Δ and ppPabA (20 μM each). Charges of the most populated species are included. Mass spectra for mixtures of stTrpEx Δ and other PabA or TrpG glutaminases are shown in *SI Appendix, Fig. S8A*. (F) Predicted interface region of an artificial smTrpEx:ppPabA complex. Positions selected for mutation are highlighted in blue, and the smTrpEx interface add-on is highlighted in green. (G) Interface region of one smTrpEx:smTrpG dimer of the tetrameric AS complex from *S. marcescens* (PDB ID code 1i7q). Residues highlighted in pink correspond to those chosen for mutations in ppPabA. Residue numbering in G and H is according to PDB ID code 1i7q. (H) Sequence logos of the highlighted residues and their surroundings.

cases, the presence of TrpE/PabB, as expected, did not lead to glutamine hydrolysis activity. Only the presence of ppTrpE and ssTrpE activated the TrpGs to a certain degree.

All PabA glutaminases proved to be slightly active already in the absence of any synthase; the apparent turnover rates were between 0.003 s^{-1} and 0.015 s^{-1} . When supplemented with TrpE or PabB, the rates increased up to 70-fold. Thus, PabA glutaminases are stimulated by allosteric signals from both TrpE and PabB. As expected, the presence of the three TrpEx synthases did not stimulate their glutaminase activity.

The Interface Add-On Determines Interaction Specificity in AS and ADCS Complexes. The hitherto presented data strongly indicate that the interface add-on in TrpEx synthases is the structural element that prevents the binding of PabA glutaminases and only allows for the formation of correct TrpEx/TrpG pairs. As an ultimate test, we deleted six residues in α -helix H2 and the adjacent loop L2 of the interface add-on from stTrpEx while leaving intact the remaining interface to stTrpG (Fig. 3D). The resulting stTrpEx Δ variant could be expressed in soluble form and purified. It was enzymatically active for the formation of AA with ammonium chloride as the nitrogen source; the catalytic efficiency ($k_{cat}/K_m^{\text{CH}} = 1.6 \pm 0.1 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$) was identical to that of stTrpEx ($k_{cat}/K_m^{\text{CH}} = 1.5 \pm 0.01 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$).

We first observed that the propensity of stTrpEx Δ to bind to stTrpG and ecTrpGD was reduced (SI Appendix, Fig. S84). Moreover, in striking contrast to stTrpEx, stTrpEx Δ was able to bind to ppPabA, smPabA, and ecPabA, leading to the formation of tetrameric TrpEx Δ_2 :PabA $_2$ complexes (Fig. 3E and Table 1 and SI Appendix, Fig. S84). No interaction of stTrpEx Δ could be detected with bsPabA. Although the reasons for this deviating behavior are unclear, pairwise sequence comparisons show that bsPabA bears the lowest sequence similarity to the other PabA glutaminases (SI Appendix, Table S5). The binding of stTrpEx Δ to ppPabA leads to a twofold stimulation of ppPabA glutaminase activity (Table 1), compared with the 4.5-fold stimulation of ppPabA by its native interaction partners ppTrpE and ppPabB. This demonstrates that the interaction between stTrpEx Δ and ppPabA is productive, as the catalytic activity of the glutaminase is enhanced by the binding of the synthase. Although we could also detect modest activation of smPabA and ecPabA by stTrpEx Δ , the stimulation factors were too low to be reliably quantified. None of the stTrpEx Δ_2 -glutaminase complexes were catalytically

active in the context of glutamine-dependent AA formation. Most plausibly, the channel required for transferring ammonia from the glutaminases to stTrpEx Δ is not functional. In any case, however, our data clearly support the notion that the TrpEx interface add-on determines interaction specificity in AS and ADCS complexes.

Generation of Functional TrpEx–PabA Interactions by Interface Design. So far, we have demonstrated that the interactions in TrpEx $_2$:TrpG $_2$ and PabB/TrpE:PabA complexes are specific and that synthase–glutaminase cross-interactions are prevented by the interface add-on in TrpEx. Thus, possible cross-interactions are physiologically highly unlikely. This raises the questions as to how the transition from TrpE to TrpEx took place and how the orthogonal TrpEx $_2$:TrpG $_2$ /PabB:PabA system has evolved in γ -Proteobacteria. One possible scenario is that TrpG evolved in response to the emergence of the TrpEx interface add-on. However, this scenario implies a cell that temporarily tolerates the existence of a noninteracting, and thus nonfunctional, TrpEx–PabA pair, which seems improbable, given the essential metabolic function of AS in tryptophan biosynthesis.

A more plausible evolutionary route is in line with recent findings by Laub and coworkers (28). They showed that novel toxin–antitoxin complexes, which are important for cell defense and viability, can evolve without nonfunctional interstages by passing through promiscuous intermediates with relaxed interaction specificity. A similar evolutionary path might have led from PabA to TrpG via intermediates that display relaxed interaction specificity toward synthases. This would imply that few mutations in PabA are sufficient to allow for TrpEx binding without compromising its function and the interaction with its cognate partner PabB.

To test the plausibility of this evolutionary scenario, we first generated a homology model of ppPabA and superimposed it with smTrpG in the smTrpEx $_2$:smTrpG $_2$ crystal structure to yield an artificial smTrpEx:ppPabA complex (Fig. 3F). Using this model and a comprehensive sequence analysis of the interface regions of TrpG and PabA glutaminases (SI Appendix, Fig. S9), we chose five residues of ppPabA. These residues are located in structural elements of TrpG that either directly contact the interface add-on or are in close proximity to add-on contacting residues (regions 1 and 2 in SI Appendix, Fig. S9A) and are conserved within either the TrpG- or PabA-type glutaminases. Other IFRs in these structural elements were not considered, either because they were conserved across the two glutaminase types

Table 1. Structural and enzymatic characteristics of synthase–glutaminase complexes comprising either the stTrpEx Δ synthase or the ppPabA* glutaminase

Synthase–glutaminase pairs	Complex formation, [†] native MS	Kinetic parameters of AS formation [‡]			Glutaminase activity [§] f_{stim}
		k_{cat} , s^{-1}	K_m^{CH} , μM	k_{cat}/K_m^{CH} , $\text{M}^{-1} \cdot \text{s}^{-1}$	
stTrpEx Δ + stTrpG	Ex Δ_2 :G $_2$ tetramer	—	—	—	n.a.
stTrpEx Δ + ecTrpGD	Ex Δ_2 :GD $_2$ tetramer	—	—	—	n.a.
stTrpEx Δ + ppPabA	Ex Δ_2 :A $_2$ tetramer	—	—	—	2.0 ± 0.1
stTrpEx Δ + smPabA	Ex Δ_2 :A $_2$ tetramer	—	—	—	<2.0
stTrpEx Δ + ecPabA	Ex Δ_2 :A $_2$ tetramer	—	—	—	<2.0
stTrpEx Δ + bsPabA	Ex Δ_2 :A $_2$ tetramer	—	—	—	<2.0
ppPabA* + stTrpEx	Ex $_2$:A $_2$ tetramer	0.31 ± 0.05	6.2 ± 3.5	5.6×10^4	41.8 ± 3.0
ppPabA* + ecTrpEx	Ex $_2$:A $_2$ tetramer	0.63 ± 0.05	9.6 ± 2.0	6.7×10^4	45.2 ± 3.3
ppPabA* + smTrpEx	Ex $_2$:A $_2$ tetramer	0.04 ± 0.004	5.2 ± 2.8	8.8×10^3	49.0 ± 2.7
ppPabA* + ppTrpE	E $_2$:A $_2$ tetramer	0.9 ± 0.06	8.4 ± 1.8	1.1×10^5	11.5 ± 3.0
ppPabA* + ppPabB	B:A dimer	0.2 ± 0.03	25.0 ± 2.6	6.7×10^3	19.1 ± 0.6

[†]Representative spectra and molecular weights are provided in SI Appendix, Fig. S8 and Table S7.

[‡]Values are the mean and SD from at least three independent measurements. stTrpEx Δ was inactive (—).

[§]Stimulation factors (f_{stim}) are based on the apparent glutamine hydrolysis rate of the listed glutaminases (SI Appendix, Table S9) and that of ppPabA* ($k_{app} = 0.005 \text{ s}^{-1}$). Glutaminase stimulation at the lower end of the assay detection limit is <2.0. n.a., not applicable (TrpG glutaminases do not display glutaminase activity in the absence of a synthase).

(regions 3–6 in *SI Appendix*, Fig. S9A) or because the corresponding positions did not show any preferences for certain amino acids. We replaced these five residues with the corresponding ones of smTrpG (Fig. 3G), which resulted in the variant ppPabA*. Four of the substitutions (Q19D, Y20Q, G22R, and I31Y) lead to residues that are mainly conserved within TrpG glutaminases, whereas the D34Q substitution leads to a residue that is only present in a minor fraction of TrpG homologs (Fig. 3H) but removes the highly conserved aspartate found in PabA glutaminases.

The ppPabA* variant could be expressed in soluble form and purified. It formed tetrameric complexes with all three TrpEx homologs (Table 1 and *SI Appendix*, Fig. S8B). The complexes were functional and converted CH to AA in a glutamine-dependent assay with catalytic efficiencies, k_{cat}/K_m^{CH} , between $8.8 \times 10^3 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $6.7 \times 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$. These values are only one order of magnitude lower than those of native TrpEx₂:TrpG₂ complexes. Moreover, the glutaminase activity of ppPabA* is stimulated about 45-fold by the presence of stTrpEx, smTrpEx, or ecTrpEx. Therefore, ppPabA* possesses the three hallmark features of wild-type glutaminases: (i) formation of stable glutaminase–synthase complexes, (ii) channeling of nascent ammonia from glutaminase to synthase, and (iii) allosteric stimulation of glutaminase activity.

Notably, ppPabA* displayed a relaxed interaction specificity toward synthases, as it retained its ability to form stable and functional complexes with both ppTrpE and ppPabB, which do not contain the interface add-on (Table 1). The composition and catalytic efficiencies of these complexes were comparable to those of genuine ppTrpE:ppPabA and ppPabB:ppPabA complexes. Thus, the five substitutions only marginally impair the native ppPabA function but permit an additional functional interaction with TrpEx. PpPabA* is therefore promiscuous with regard to interacting with synthases that contain an interface add-on and those that do not. It may thus represent a promiscuous evolutionary intermediate that fills the gap between the TrpE/PabB-specific PabA glutaminases and the TrpEx-specific TrpG glutaminases.

Importantly, the crucial effect of these mutations is to allow for the accommodation of the interface add-on of TrpEx by the ppPabA* glutaminase, because they are located in structural elements of the glutaminase that lie close to the add-on in a TrpEx:TrpG complex. Moreover, the remaining interface regions of TrpEx, TrpE, and also PabB synthases are highly similar. A comprehensive sequence analysis of these regions (regions 2–5 in *SI Appendix*, Fig. S9B) clearly shows that there are no major differences in the interface composition between the three types of synthases and that there is no other decisive negative design element present in TrpEx that is missing in TrpE and PabB. The single IFR of TrpEx outside of the add-on that is specific for TrpEx is Arg283. However, this residue is distant from the TrpEx–TrpG interface and is rather involved in interactions necessary for tetramer formation in AS complexes (29). Taken together, these results clearly support the view that the interface add-on of TrpEx is the essential specificity determinant for TrpEx–TrpG interactions and that no broader change in sequence led to the isolation of these interactions.

Experimental Evidence for Harmful Metabolic Cross-Talk in TrpE Species. Our genetic profiling of over 15,000 bacterial and archaeal species has shown that Vibrionales and Enterobacteria have evolved a conserved orthogonal system of GATases for the two important biosynthetic pathways leading to tryptophan and folate. The observation that TrpEx and TrpG have been retained in all descendants after the split between *Pseudomonas* and *Shewanella* species ~950 Mya (30) suggests that two orthogonal AS and ADCS complexes entail some kind of selective advantage like the prevention of metabolic cross-talk.

In principle, such cross-talk is conceivable in bacteria that do not possess TrpEx but do possess TrpE. As shown above, these

species only contain PabA glutaminases that form functional AS and ADCS complexes with both TrpE and PabB. This raises the question as to how ammonia flow is directed toward either tryptophan or folate biosynthesis in these organisms. The existence of sophisticated regulatory mechanisms in Firmicutes illustrates the effort that has been invested by nature to solve this problem (*SI Appendix*, Fig. S10A and B).

The central player in regulating tryptophan and folate biosynthesis in *B. subtilis* is the tryptophan-sensing protein TRAP (31), which binds excessive tryptophan and exercises transcriptional attenuation and translational control on both the tryptophan and the folate operon. A problematic situation develops if biosynthesis of folate is required while cellular levels of tryptophan are high. Under such conditions, the production of the single available glutaminase, PabA, required for folate biosynthesis is blocked by TRAP. However, a dual-promoter system allows for the translation-mediated displacement of TRAP, resulting in the full ADCS complex required for folate biosynthesis (32).

As long as a cell contains TrpE and PabB concurrently, the TRAP-based mechanism cannot guarantee specific direction of ammonia flow to one pathway or the other. Moreover, this sophisticated regulatory network could easily be disturbed, resulting in potentially harmful metabolic cross-talk. We simulated such a situation by expressing a plasmid-borne copy of *trpE* in *B. subtilis* and hypothesized that the resulting high amounts of TrpE synthase will take up all PabA glutaminase, and thus deduct it from folate biosynthesis. Vice versa, we also assumed that a similar expression of *pabB* will lead to a shortage of PabA glutaminases in tryptophan biosynthesis.

We transformed the prototrophic *B. subtilis* strain SB491 with expression plasmids that contained either *bspabB* or *bstpE*. Plasmids containing *bspabA*, *sttpEx*, or *smtrpEx* were used as controls. To test the phenotypic effects of isopropyl β-D-1-thiogalactopyranoside (IPTG)-induced overexpression, we let the transformants grow on defined minimal medium and determined average colony sizes (Fig. 4A). Expression of *bspabB* had no effect on the average colony size. However, expression of *bstpE* resulted in significantly smaller colonies, which grew to only about 20% of the size observed in the absence of IPTG. We did not observe similar growth deficiencies for the empty plasmid or for the expression of *bspabA*, *sttpEx*, and *smtrpEx*. Thus, the growth deficiency observed upon expression of *bstpE* is not caused by toxic effects from IPTG or high protein concentrations. To test if the growth deficiency is caused by compromised folate biosynthesis, we performed the same growth experiments in the presence of folate or its precursor *p*-aminobenzoic acid (Fig. 4B). Indeed, the supplementation of these metabolites resulted in normally sized colonies, thus offsetting the effect of *bstpE* expression.

Discussion

Interface Add-Ons Are an Evolutionary Tool for the Diversification of Protein Interfaces and Protein–Protein Interactions. It is intriguing how protein complexes form with the specificity and selectivity required for their proper function in almost all biological processes. A profound understanding of this specificity and selectivity is not possible without detailed knowledge about how protein–protein interfaces and interactions between proteins change and evolve. Given the limited number of different protein architectures (4, 6), quaternary structure topologies (33), and interface geometries (2, 3), the diversification of most protein–protein interactions is the result of adaptational mutations (34) of residues that did not change interface geometry. For instance, many bacteria contain several large families of paralogous toxin–antitoxin complexes, which interact in a highly specific manner (35). Interaction specificity in one particular family has been narrowed down to five residues in the toxin and four residues in the antitoxin, which can mutate and adapt to each other without compromising protein function (28). Moreover, considerable

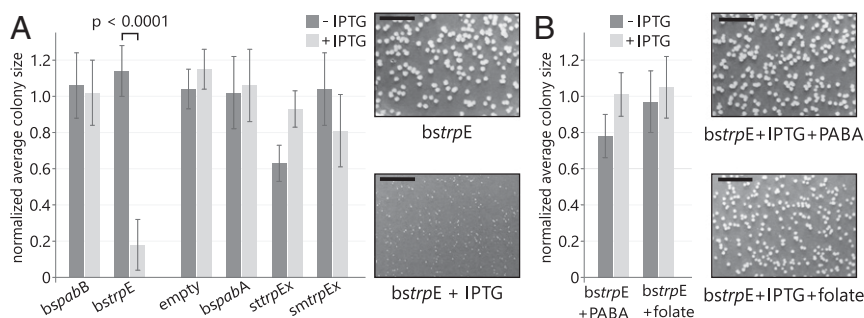


Fig. 4. *B. subtilis* growth experiments. Average sizes of *B. subtilis* colonies grown from cells transformed with the indicated plasmids. Colony sizes were determined after growth on minimal medium lacking (dark gray) or containing (light gray) 2 mM IPTG for 48 h. Values are normalized to the pDG(−) sample; error bars indicate the SD from four independent replicates. (Insets) Representative colonies for selected samples. (Scale bars: 1 cm.) Representative images of all samples are shown in *SI Appendix, Fig. S10 C and D*. (A) Overexpression of *bstrpE* leads to a significant reduction of colony size ($P < 0.0001$ at a confidence interval of 99%). (B) Presence of PABA or folate offsets the effect of *bstrpE* overexpression.

specificity in paralogous families of bacterial histidine kinase-response regulator complexes, which, for example, sense and respond to changes in phosphate availability, is generated by just three to four residues in the complex interfaces (36).

Mutational diversification of interactions is, however, inherently restrained if the interface does not only serve as a binding surface but also performs an additional function like the propagation of an allosteric signal, the completion of an active site, or the channeling of reaction intermediates between the interacting proteins. In such cases, mutations of IFRs, while increasing interaction specificity, may concurrently compromise function. This greater selective constraint imposed on interfaces compared with other regions of proteins is reflected in the lower mutational rate of IFRs relative to non-IFRs (37, 38).

Our systematic survey of protein–protein interfaces in heteromeric complexes highlights one solution to such dilemmas: the addition of add-ons to existing interfaces. These interface add-ons have a typical length of 10–20 aa and also mostly a well-defined secondary structure. All of them contain at least one residue that falls into the category of a binding hot spot, such that a mutation of this residue to alanine decreases the binding free energy by at least 2 kcal/mol. In fact, many of the interface add-ons contain three or more, and some even up to nine, hot-spot residues. Interface add-ons seem to be quite frequent, consistent with the assumption that negative design elements are important evolutionary traits (8). Under very stringent filter conditions, we found them in about 10% of the structures in our representative dataset. Interface add-ons are also not limited to certain phyla as they are present in complexes from Actinobacteria, the Deinococcus-Thermus group, Firmicutes, and Thermotogae, as well as several classes of Proteobacteria.

Large interface insertions with similar effects on protein–protein interaction specificity have, to the best of our knowledge, not been identified so far. Although significant insertions have been described in other enzymes, they merely modulate self-association of homooligomers (6) or allosteric regulation (39), but have no impact on interaction specificity. Moreover, interface add-ons are to be discriminated from small, independently folding, interaction-mediating domains like ankyrin repeats (40), pox virus and zinc finger (POZ) domains (41), or bromodomains (42). These elements are stand-alone mediators of protein–protein interactions and do not change the specificity of an already existing interaction.

Subcellular localization or coordinated gene expression and protein synthesis can also help to assure specificity. For example, certain histidine kinases are present at a very low abundance inside *E. coli* cells; thus, the chances of them binding to a non-cognate response regulator are rather low (43, 44). It should also not be neglected that for some proteins, absolute interaction specificity or the avoidance of cross-talk is not required (44, 45).

We thus assume that interface add-ons have been integrated into protein complexes when adaptational mutations or spatial regulation mechanisms were not sufficient to avoid severe negative physiological effects caused by cross-talk with other structurally similar complexes.

One such case is most likely the interface add-on in the AS and its TrpEx subunit. First, with a length of 51 aa, the TrpEx interface add-on is particularly extensive. Such large insertions are rare because of their high risk of impairing protein stability. Consequently, insertions or deletions commonly comprise only one residue and most are shorter than eight residues (6, 46). Second, the synthase–glutaminase interface in AS does not only mediate complex formation but is also crucial for ammonia channeling and allosteric communication between the two subunits. Obviously, the TrpEx interface add-ons extend the interface in a way that does not compromise these functional properties. Third, and most importantly, many bacteria, in addition to AS, also contain the homologous ADCS complex, which catalyzes a similar reaction in folate biosynthesis. In these organisms, the TrpEx interface add-on determines the specific formation of AS and ADCS complexes, despite the highly similar core interfaces of PabB and TrpEx (Fig. 5A). This is exemplified by the properties of the stTrpEx Δ variant, where the deletion of six residues allowed for the binding of a PabA glutaminase.

The Interface Add-On in TrpEx Has Important Physiological Consequences.

The *in vivo* growth experiments with *B. subtilis* suggest that the expression of an additional, plasmid-borne *bstrpE* gene might sequester the available PabA glutaminases, thereby impairing folate biosynthesis and significantly affecting cellular fitness (Fig. 4). From a broader perspective, we have demonstrated that cross-talk between metabolic pathways can be harmful for an organism. A metabolic conflict comparable to that in Firmicutes does not exist in species that possess TrpEx and its specific, associated glutaminase TrpG. In these species, *trpG* is an integral part of the *trp* operon and is translationally coupled to *trpEx* by overlapping start and stop codons, which facilitates an equimolar synthesis of the both AS components. Consequently, the synthesis of the two AS components can be controlled independent of cellular folate levels, and a repression and attenuation mechanism is sufficient to regulate the *trp* operon (47).

Glutaminase Intermediates with Relaxed Interaction Specificity Enabled the Evolution of TrpEx Species. Phylogenetic distribution and sequence similarities of the glutaminases suggest that TrpG has evolved from a PabA ancestor. It is, however, highly unlikely that the appearance of TrpG was a compensatory response to the emergence of TrpEx. Our experimental characterization of TrpEx shows that it cannot interact with PabA. Thus, an organism

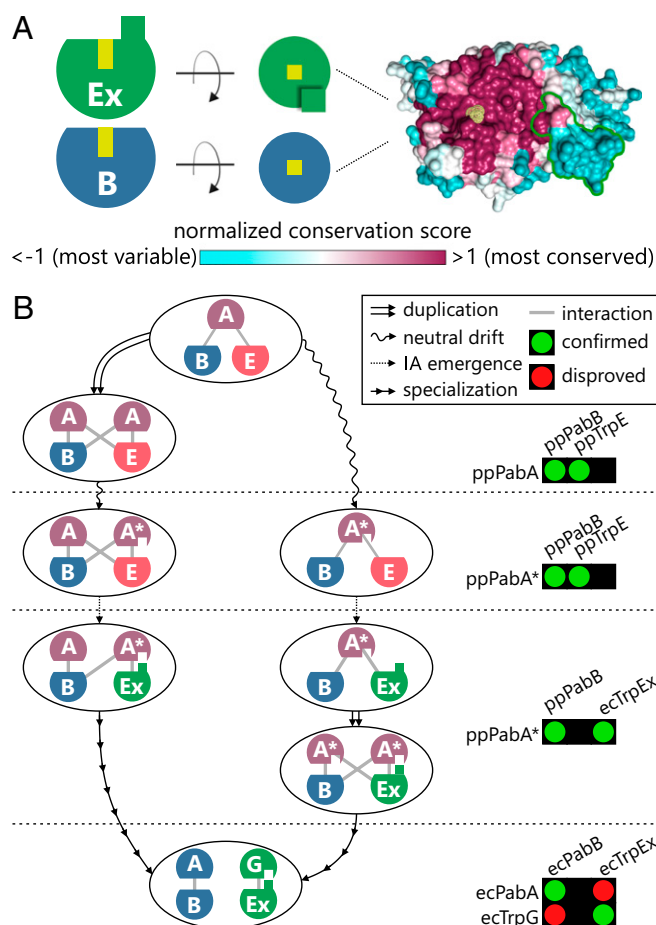


Fig. 5. Conservation of the interface region in TrpEx/PabB and possible evolutionary trajectories leading to the orthogonal complexes of γ -Proteobacteria. (A) TrpEx and PabB are depicted as circle segments with their half of the ammonia channel marked by a yellow box. The interface add-on is indicated by a green square. A top view of the interface region of TrpEx is shown in surface representation with the interface add-on and the entry to the ammonia channel marked in green and yellow, respectively. The surface is colored according to the normalized ConSurf conservation score, which was calculated from 5,849 TrpEx and PabB sequences with a pairwise identity <90%. (B) Evolutionary scenarios leading to separate AS and ADCS complexes in TrpEx species. On the right hand side, experimentally characterized interactions are shown that represent the respective evolutionary phases.

containing TrpEx, PabB, and only PabA would be nonviable due to nonfunctional tryptophan biosynthesis.

Plausible scenarios that lead from a TrpE species to a TrpEx species avoid such evolutionary dead-ends by assuming promiscuous PabA* glutaminase intermediates with relaxed interaction specificity. Promiscuity in this context refers to their ability to interact with synthases that contain an interface add-on (i.e., TrpEx), and likewise with synthases that do not contain an interface add-on (i.e., TrpE, PabB). Our designed ppPabA* variant displays such a relaxed interaction specificity. It contains five amino acid substitutions that are sufficient to establish stable and functional interactions with TrpEx. Moreover, it forms stable and functional complexes also with ppTrpE and ppPabB, which both lack the TrpEx interface add-on.

A parsimonious approach building on comparable intermediates leads to two different plausible evolutionary trajectories (Fig. 5B). In the neofunctionalization trajectory (Fig. 5B, left path), the gene of an ancestral PabA was duplicated, thus allowing for the acquisition of mutations that lead to a relaxed interaction specificity. Mutational drift and coevolution (48)

eventually led to a specialization of PabA* to the contemporary TrpG glutaminase found in γ -Proteobacteria. In the alternative subfunctionalization trajectory (Fig. 5B, right path), the presence of a promiscuous PabA* intermediate made it possible to tolerate the integration of the interface add-on into TrpE. After gene duplication, the copies coevolved with TrpEx and PabB, respectively, and subfunctionalization led to specific contemporary AS and ADCS complexes.

With the data at hand, it is not possible to decide which evolutionary trajectory might more accurately describe the evolutionary history of TrpG. For example, it is known that most duplicated genes do not stay in the gene pool of a population long enough to accumulate function- or specificity-changing mutations and are lost instead (49, 50), arguing against the neofunctionalization path. On the other hand, the promiscuity-inducing mutations in a PabA* variant must not be at the expense of catalytic efficiency or protein stability, an important point to consider with the subfunctionalization path. The phylogenetic distribution of TrpG, which exclusively occurs in Proteobacteria, and the pervasive PabA, which does not interact with TrpEx, suggest that TrpEx interaction is not an ancestral feature and argue against a PabA* predecessor. Irrespective of which trajectory reflects the actual evolutionary path, a promiscuous PabA* intermediate is required to interact with both TrpEx and PabB, ensuring that ammonia is made available for tryptophan and folate biosynthesis.

Materials and Methods

Detailed materials and methods are described in *SI Appendix, SI Materials and Methods*.

Survey of Interface Add-Ons in Heteromeric Protein Complexes. The initial dataset contained 1,739 heteromeric bacterial protein complex structures available in the PDB with subunit stoichiometries of AB, A₂B₂, A₃B₃, A₄B₄, A₆B₆, ABC, and A₂B₂C₂. From this set, 305 reference complex structures were selected, and for each subunit of this dataset, all homologs were compiled from the corresponding InterPro family. Insertions in the reference structures or their homologs from InterPro were subsequently identified by pairwise sequence alignments. Insertions were evaluated with respect to (i) containing more than eight residues to exclude short indels from the analysis, (ii) being located in a protein interface, and (iii) containing binding hot-spot residues, whose in silico mutation to alanine significantly decreased the stability of the corresponding complex structures.

Computation of SSN, Amino Acid Conservation, and Sequence Logos. SSNs of the InterPro entries IPR019999 and IPR015890 were computed according to Gerlt et al. (51) and visualized with Cytoscape 3.3. Amino acid conservation and sequence logos were computed from multiple sequence alignments (MSAs) of synthase and glutaminase sequences.

Genetic Profiling of Archaea and Bacteria to Determine the Phylogenetic Distribution of AS and ADCS Complexes. To investigate whether the presence of TrpEx or TrpE affects the distribution of PabB and the type and number of associated glutaminases, we determined the occurrence of TrpEx, TrpE, PabB, TrpG, and PabA for all species associated with the TrpEx and TrpE subclusters in the SSN of IPR015890. A grouping routine that is based on HMMs was designed (*SI Appendix, Fig. S3A*). In brief, all species that contribute sequences to the TrpEx and TrpE subclusters were individually scanned for the presence of TrpEx, TrpE, and PabB, as well as for the presence of TrpG and PabA, using BLAST. Hits were classified by comparison with five enzyme-specific HMMs (*SI Appendix, Fig. S3 B and C*). Finally, two representative datasets, *TrpEx_{repr}* and *TrpE_{repr}*, containing the cooccurring proteins for TrpEx and TrpE species, respectively, were generated.

Computation of Interface Conservation. The similarity of the interface regions between TrpEx and PabB was deduced from structure-based MSAs comprising sequences from *TrpEx_{repr}* and *TrpE_{repr}*.

Cloning and Mutagenesis. The genes of TrpEx, TrpE, PabB, TrpG, and PabA proteins were amplified from genomic DNA or whole-cell lysate in standard PCR reactions and cloned into pET21a, pET28a, or pMAL-c5T vector. The *sttrpEx_Δ* was generated by deleting codons 111–116 and inserting the codons AGC and GGC coding for serine and glycine, respectively, in pET21a-*sttrpEx*

by using a New England Biolabs Q5 Site-Directed Mutagenesis Kit. The gene of ppPabA* was optimized for expression in *E. coli* and synthesized (Life Technologies).

Expression and Purification of Proteins. If not stated otherwise, all proteins were produced by gene expression in *E. coli* BL21-Gold (DE3) cells. The ssTrpE was produced by gene expression in *E. coli* BL21-CodonPlus (DE3)-RIPL cells. Cells were grown in Luria broth medium at 20 °C or 37 °C overnight. Proteins were purified from the soluble fraction of the cell extracts by Ni²⁺-affinity and SEC.

HPLC Analysis of AA and ADC Formation. TrpEx, TrpE, and PabB were assayed in 20 mM bicine buffer (pH 8.5), 5 mM MgCl₂, 1 mM DTT, 200 mM NH₄Cl, and 500 μM CH. PabB assays additionally contained 10 μM ADC-lyase from *E. coli* (ecPabC) for conversion of the PabB product ADC to *p*-aminobenzoate (PABA).

Analysis of Complex Formation Between Different Synthases and Glutaminases. The ability of the different synthases and glutaminases to form heteromeric complexes was examined by SEC-SLS. Individual proteins and mixtures of glutaminases and synthases were assayed at 50 μM. Complex formation of selected glutaminase and synthase pairs was additionally analyzed with native mass spectrometry on a quadrupole ion mobility time-of-flight mass spectrometer equipped with a nanoelectrospray ionization source.

Steady-State Enzyme Kinetics. The TrpEx/TrpE reaction was measured at 25 °C in a fluorimetric assay monitoring AA fluorescence (excitation of 313 nm and emission of 390 nm). The PabB reaction was measured at 25 °C monitoring PABA fluorescence (excitation of 320 nm and emission of 350 nm). The PabB product ADC was converted in situ to PABA by PabC. The glutaminase activity was measured spectrophotometrically in a coupled enzymatic assay, in which formed glutamate was converted to α-ketoglutarate by glutamate dehydrogenase.

***B. subtilis* Growth Experiments.** The *bspabA*, *bspabB*, *bstrpE*, *sttrpEx*, and *smtrpEx* genes were cloned into the pDG148 vector using a ligation-independent cloning protocol. Electrocompetent *B. subtilis* SB 491 cells were generated and transformed with pDG148_*bspabA*, pDG148_*bspabB*, pDG148_*bstrpE*, pDG148_*sttrpEx*, or pDG148_*smtrpEx*. In vivo competition assays were performed on Spizizen's minimal medium agar plates.

ACKNOWLEDGMENTS. We thank Alexandra Holinski for providing the pMAL-cST expression vector and Christiane Endres, Sonja Fuchs, Sabine Laberer, and Jeannette Ueckert for expert technical assistance. M.G.P. and F.S. were supported by doctoral fellowships from the Fonds der Chemischen Industrie and the Konrad-Adenauer-Stiftung, respectively. F.B. was supported by NIH Grant R01 GM113658 (to V.H.W.).

- Yu H, et al. (2008) High-quality binary protein interaction map of the yeast interactome network. *Science* 322:104–110.
- Gao M, Skolnick J (2010) Structural space of protein-protein interfaces is degenerate, close to complete, and highly connected. *Proc Natl Acad Sci USA* 107:22517–22522.
- Garma L, Mukherjee S, Mitra P, Zhang Y (2012) How many protein-protein interaction types exist in nature? *PLoS One* 7:e38913.
- Chothia C (1992) Proteins. One thousand families for the molecular biologist. *Nature* 357:543–544.
- Zhang Y, Hubner IA, Arakaki AK, Shakhnovich E, Skolnick J (2006) On the origin and highly likely completeness of single-domain protein structures. *Proc Natl Acad Sci USA* 103:2605–2610.
- Hashimoto K, Panchenko AR (2010) Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc Natl Acad Sci USA* 107:20352–20357.
- Ofran Y, Rost B (2003) Analysing six types of protein-protein interfaces. *J Mol Biol* 325:377–387.
- Schreiber G, Keating AE (2011) Protein binding specificity versus promiscuity. *Curr Opin Struct Biol* 21:50–61.
- Zalkin H (1993) The amidotransferases. *Advances in Enzymology and Related Areas of Molecular Biology*, ed Meister A (Wiley, Hoboken, NJ), Vol 66, pp 203–309.
- Bouvier B, Grünberg R, Nilges M, Cazals F (2009) Shelling the Voronoi interface of protein-protein complexes reveals patterns of residue conservation, dynamics, and composition. *Proteins* 76:677–692.
- Guharoy M, Chakrabarti P (2005) Conservation and relative importance of residues across protein-protein interfaces. *Proc Natl Acad Sci USA* 102:15447–15452.
- Webb B, Sali A (2014) Protein structure modeling with MODELLER. *Methods Mol Biol* 1137:1–15.
- Pires DE, Ascher DB, Blundell TL (2014) mCSM: Predicting the effects of mutations in proteins using graph-based signatures. *Bioinformatics* 30:335–342.
- Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J Mol Biol* 280:1–9.
- Dosselaere F, Vanderleyden J (2001) A metabolic node in action: Chorismate-utilizing enzymes in microorganisms. *Crit Rev Microbiol* 27:75–131.
- Gollnick P, Babitzke P, Antson A, Yanofsky C (2005) Complexity in regulation of tryptophan biosynthesis in *Bacillus subtilis*. *Annu Rev Genet* 39:47–68.
- Slock J, Stahly DP, Han CY, Six EW, Crawford IP (1990) An apparent *Bacillus subtilis* folic acid biosynthetic operon containing *pab*, an amphibolic *trpG* gene, a third gene required for synthesis of para-aminobenzoic acid, and the dihydropteroate synthase gene. *J Bacteriol* 172:7211–7226.
- Söding J (2005) Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21:951–960.
- Yoon BJ (2009) Hidden Markov models and their applications in biological sequence analysis. *Curr Genomics* 10:402–415.
- Farrow JM, 3rd, Pesci EC (2007) Two distinct pathways supply anthranilate as a precursor of the *Pseudomonas* quinolone signal. *J Bacteriol* 189:3425–3433.
- Xie G, et al. (2003) Lateral gene transfer and ancient paralogy of operons containing redundant copies of tryptophan-pathway genes in *Xylella* species and in heterocystous cyanobacteria. *Genome Biol* 4:R14.
- White RH (1988) Analysis and characterization of the folates in the nonmethanogenic archaeobacteria. *J Bacteriol* 170:4608–4612.
- Worrell VE, Nagle DP, Jr (1988) Folic acid and pteroylpolylglutamate contents of archaeobacteria. *J Bacteriol* 170:4420–4423.
- Bera AK, Chen S, Smith JL, Zalkin H (1999) Interdomain signaling in glutamine phosphoribosylpyrophosphate amidotransferase. *J Biol Chem* 274:36498–36504.
- Goto Y, Zalkin H, Keim PS, Heinrichson RL (1976) Properties of anthranilate synthetase component II from *Pseudomonas putida*. *J Biol Chem* 251:941–949.
- Miles BW, Banzon JA, Raushel FM (1998) Regulatory control of the amidotransferase domain of carbamoyl phosphate synthetase. *Biochemistry* 37:16773–16779.
- List F, Bocola M, Haeger MC, Sterner R (2012) Constitutively active glutaminase variants provide insights into the activation mechanism of anthranilate synthase. *Biochemistry* 51:2812–2818.
- Aakre CD, et al. (2015) Evolving new protein-protein interaction specificity through promiscuous intermediates. *Cell* 163:594–606.
- Morollo AA, Eck MJ (2001) Structure of the cooperative allosteric anthranilate synthase from *Salmonella typhimurium*. *Nat Struct Biol* 8:243–247.
- Battistuzzi FU, Hedges SB (2009) Eubacteria. *The Timetree of Life*, eds Hedges SB, Kumar S (Oxford Univ Press, Oxford), pp 106–115.
- Babitzke P (1997) Regulation of tryptophan biosynthesis: Trp-ing the TRAP or how *Bacillus subtilis* reinvented the wheel. *Mol Microbiol* 26:1–9.
- Yakhnin H, Yakhnin AV, Babitzke P (2007) Translation control of *trpG* from transcripts originating from the folate operon promoter of *Bacillus subtilis* is influenced by translation-mediated displacement of bound TRAP, while translation control of transcripts originating from a newly identified *trpG* promoter is not. *J Bacteriol* 189:872–879.
- Ahnert SE, Marsh JA, Hernández H, Robinson CV, Teichmann SA (2015) Principles of assembly reveal a periodic table of protein complexes. *Science* 350:aaa2245.
- Capra EJ, Perchuk BS, Skerker JM, Laub MT (2012) Adaptive mutations that prevent cross-talk enable the expansion of paralogous signaling protein families. *Cell* 150:222–232.
- Fiebig A, Castro Rojas CM, Siegal-Gaskins D, Crosson S (2010) Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin-antitoxin systems. *Mol Microbiol* 77:236–251.
- Skerker JM, et al. (2008) Rewiring the specificity of two-component signal transduction systems. *Cell* 133:1043–1054.
- Ames RM, Talavera D, Williams SG, Robertson DL, Lovell SC (2016) Binding interface change and cryptic variation in the evolution of protein-protein interactions. *BMC Evol Biol* 16:40.
- Mintseris J, Weng Z (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc Natl Acad Sci USA* 102:10930–10935.
- Sinthak MD, Arjara G, Kellogg BA, Stubbe J, Drennan CL (2002) The crystal structure of class II ribonucleotide reductase reveals how an allosterically regulated monomer mimics a dimer. *Nat Struct Biol* 9:293–300.
- Li J, Mahajan A, Tsai MD (2006) Ankyrin repeat: A unique motif mediating protein-protein interactions. *Biochemistry* 45:15168–15178.
- Bardwell VJ, Treisman R (1994) The POZ domain: A conserved protein-protein interaction motif. *Genes Dev* 8:1664–1677.
- Filippakopoulos P, Knapp S (2012) The bromodomain interaction module. *FEBS Lett* 586:2692–2704.
- Cai SJ, Inouye M (2002) EnvZ-OmpR interaction and osmoregulation in *Escherichia coli*. *J Biol Chem* 277:24155–24161.
- Laub MT, Goulian M (2007) Specificity in two-component signal transduction pathways. *Annu Rev Genet* 41:121–145.
- Nooren IM, Thornton JM (2003) Diversity of protein-protein interactions. *EMBO J* 22:3486–3492.
- Pascarella S, Argos P (1992) Analysis of insertions/deletions in protein structures. *J Mol Biol* 224:461–471.
- Merino E, Jensen RA, Yanofsky C (2008) Evolution of bacterial *trp* operons and their regulation. *Curr Opin Microbiol* 11:78–86.
- Pazos F, Valencia A (2008) Protein co-evolution, co-adaptation and interactions. *EMBO J* 27:2648–2655.
- Hughes AL (1994) The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* 256:119–124.
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155.
- Gerlt JA, et al. (2015) Enzyme function initiative-enzyme similarity tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochim Biophys Acta* 1854:1019–1037.