Co-Occurrence of Enzyme Domains Guides the Discovery of an Oxazolone Synthetase

Tristan de Rond*[a], Julia E. Asay[a] and Bradley S. Moore*[a,b]

[a]Center for Marine Biotechnology and Biomedicine, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA 92093; [b]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA 92093

ORCIDs: Tristan de Rond: 0000-0002-5203-4662, Bradley S. Moore: 0000-0002-4652-1253

*To whom correspondence may be addressed.
Email: tderond@ucsd.edu, bsmoore@ucsd.edu

Author contributions: TdR and BSM designed research; TdR and JEA performed research; TdR analyzed data; TdR, JEA and BSM wrote the manuscript

The authors declare no competing interest

This article contains supporting information

Abstract

Multi-domain enzymes are cellular machines that orchestrate two or more catalytic activities to carry out metabolic transformations with increased control and speed. Here, we report the development of a new genome mining approach for the targeted discovery of new metabolic pathways based on the co-occurrence of enzyme domains (CO-ED) in a single protein. CO-ED was designed to identify unannotated multifunction proteins for accelerated functional characterization and discovery based on the premise that linked enzyme domains have evolved to function collaboratively. Guided by CO-ED, we targeted an unannotated predicted ThiF-nitroreductase di-domain enzyme found in more than 50 proteobacteria. Through heterologous expression and biochemical reconstitution of this enzyme, we discovered a series of new natural products containing the rare oxazolone heterocycle, and identified the first reported oxazolone synthetase in biology. This proof-of-principle experiment validates CO-ED-guided genome mining as a new method with potential broad utility for both the discovery of novel enzymatic transformations and the functional gene annotation of multi-domain enzymes.

Significance

Billions of years of evolution has provided immense genetic – and hence, biochemical – diversity, much of which is still un-annotated. Genome mining seeks to explore this “metabolic dark matter” for new pharmaceuticals, biocatalysts and biochemical insight. Historically, these efforts have been guided by sequence homology to known biosynthetic genes, which often leads to re-discovery of known enzymatic transformations. Here, in contrast, we prioritize enzymes harboring multiple catalytic domains in combinations that have not been investigated before. Our approach, “CO-ED”, led to the discovery of new natural products and the identification of an enzyme capable of catalyzing the formation of an oxazolone heterocycle, a versatile chemical feedstock. CO-ED has the potential to guide the functional characterization of thousands of unannotated multi-domain enzymes.

Introduction

Our knowledge of nature's diversity of enzymatic transformations is crucial to advancing research in a multitude of disciplines. For instance, our ability to predict metabolic capacity from genome sequences enables new insights in human health and ecology(1, 2), while the development of bioprocesses to produce chemicals in an environmentally benign fashion relies on the availability of a well-stocked biocatalytic toolbox(3). Billions of years of evolution has resulted in immense natural genetic diversity, which we are rapidly starting to uncover using modern sequencing technologies. However, the functional assignment of the enzymes encoded by this sequence diversity is lagging due to its labor-intensity(4–6). Many enzymes catalyzing chemical transformations previously not known to occur in nature are still being discovered(7), and there is no end in sight of un-annotated and mis-annotated enzymes in genomic databases(8). The search for novel enzymes in this genomic wilderness can fulfill the dual purpose of functional gene annotation and biocatalyst discovery.

The computational searches for biosynthetic enzymes that underlie "genome mining" campaigns are most commonly based on gene homology to known core biosynthetic genes of well-established classes of medicinal natural products such as polyketides, non-ribosomal peptides and terpenoids(9). This approach, while likely to identify gene clusters that produce bio-active products, is prone to the re-discovery of known enzymology. Therefore, we set out to develop a genome mining strategy that could guide enzyme discovery in a less biased manner.

One notable feature amongst specialized metabolism is an overabundance of enzymes harboring multiple catalytic domains (Figure 1a). Multi-domain enzymes are thought to arise through gene fusion of two or more single-domain enzymes, affording some evolutionary advantages such as coupled temporal and spatial regulation, a fixed active site stoichiometry, and channeling of reactive intermediates(10, 11). Over time, a multi-domain enzyme may evolve to orchestrate the reactivity of its constituent active sites to such an extent that its overall transformation is greater than that effected by the sum of its parts – both in terms of catalytic efficiency and biochemical novelty. Besides the classic examples of assembly-line biosynthesis performed by fatty acid synthase, polyketide synthase, and nonribosomal peptide synthetase (NRPS) megaproteins(12) – some of which consist of dozens of domains – multi-domain enzymes also catalyze functional group modifications, such as epimerization (13, 14) and *N*-nitrosation(15, 16) reactions established recently.

Here, we present Co-Occurrence of Enzyme Domains (CO-ED): a genome mining paradigm where we leverage the evolved co-occurrence of enzymatic domains in proteins to inform enzyme discovery. Through CO-ED, we discovered a series of oxazolone-containing natural products and the first reported oxazolone synthetase, a novel bifunctional cyclodehydratase–oxidoreductase.

Results

CO-ED–guided identification of a di-domain enzyme candidate

To validate the CO-ED workflow, we first applied it to the genome of the model organism *Escherichia coli* K12. The first stage of CO-ED detects a non-redundant set of enzymatic Pfam(17) domains using HMMER(18) and displays the results as a network where nodes represent domains and edges denote the co-occurrence of two domains in the same protein (Figs. 1b,c, S1, Dataset S1). The *E. coli* network reveals 19 co-occurring pairs of domains, including some well-studied multi-functional enzymes including penicillin-binding proteins(19), the aldehyde dehydrogenase–alcohol dehydrogenase AdhE(20),

the UDP-L-Ara4FN biosynthesis enzyme ArnA(21), the amino acid biosynthesis enzymes TrpCF(22), TrpGD(23) and HisIE(24), and the enterobactin NRPS EntF(25). CO-ED's second stage colors edges representing domain pairs for which enzymes have been annotated in BRENDA(26), MIBiG(27), or Uniprot(28). In Fig. 1b, every edge is colored, suggesting that the function of every multi-domain enzyme in *E. coli* detected by CO-ED has already been established.

Next, we subjected the genomes of the model bacteria referred to in Fig. 1a – *Streptomyces coelicolor* A3(2), *Salinispora tropica* CNB-440, and *Pseudomonas fluorescens* Pf-5, each known to produce a diverse suite of specialized metabolites – to CO-ED analysis. A large proportion of the edges in the resultant networks (Figs. S2-4, Dataset S1) were associated with the multi-domain biosynthetic machinery involved in the construction of natural product molecules. Moreover, unlike the case of *E. coli*, CO-ED identified over a dozen of un-annotated domain pairs in each of these species. These observations suggest that CO-ED may help identify new classes of multi-domain enzymes, and thus novel enzymatic transformations.

We next applied CO-ED analysis to the genome of the biosynthetically talented(29), non-model marine gammaproteobacterium *Pseudoalteromonas rubra* DSM 6842, chosen because of our experience heterologously expressing enzymes from this and closely-related species(30–32). The *P. rubra* CO-ED network (Figs. 1d, S5, Dataset S1) revealed numerous assembly-line enzymes, including those involved in prodiginine biosynthesis, a putative bifunctional RibBA(33), as well as some of the same bifunctional primary metabolic enzymes observed in *E. coli*. However, unlike the *E. coli* network, twelve domain pairs are un-annotated, suggesting CO-ED–guided genome mining has the potential to discover novel enzymatic transformations in *P. rubra*.

A *P. rubra* protein harboring a ThiF-nitroreductase domain pair (highlighted in bottom right of Fig. 1d) caught our attention, because each of these domains is known to catalyze a diverse set of chemical transformations (Fig. 2a), yet their combination is unprecedented. Enzymes of the ThiF protein family catalyze various vital reactions that proceed through carboxylate adenylation, including the activation of ubiquitin and ubiquitin-like proteins by E1 enzymes(34), the cyclization of N^6-threonylcarbamoyladenosine in tRNA maturation(35), and play a role in the incorporation of sulfur into thiamin, the molybdenum cofactor and some natural products(36, 37). To date, all characterized substrates of ThiF family enzymes have been macromolecules. The nitroreductase family consists of flavoenzymes that catalyze a remarkable variety of redox reactions(38), such as the reductive deiodination of thyroid hormones(39), the dehydrogenation of oxazolines and thiazolines(40, 41), the fragmentation of flavin mononucleotide to form the vitamin B12 lower ligand 5,6-dimethylbenzimidazole(42), enol methylene reduction in cofactor F420 biosynthesis(43–45), diketopiperazine desaturation in albonoursin biosynthesis(46), and the reductive detoxification of various nitro functional groups(47).

Querying the Uniprot database with the di-domain enzyme from *P. rubra* revealed 56 proteins originating from alpha-, beta-, and gammaproteobacteria (Figs. 2b, S6, Table S1) isolated from a variety of (predominantly aquatic) environments (Fig. 2c). In those cases where the genomic context of the gene, which we named *oxzB*, was known, it is typically accompanied by a *N*-acyltransferase homolog (*oxzA*). In *P. rubra* and many other species, *oxzA* is immediately upstream of *oxzB* on the same strand, forming an apparent two-gene *oxzAB* operon (Fig. 2d).

Heterologous expression of *oxzAB* leads to production of novel oxazolone metabolites

To investigate the function of the *oxzAB* genes, we PCR amplified the gene pairs from *P. rubra*, *Rheinheimera pacifica*, *Colwellia chukchiensis* (all gammaproteobacteria), *Skermanella aerolata* (alphaproteobacterium), and *Undibacterium pigrum* (betaproteobacterium), and heterologously expressed them in *E. coli*. Colonies on solid media and cell pellets after growth in liquid media both turned visibly yellow (Fig. S7). The color would briefly intensify when incubated above pH 9, suggesting the presence of a phenolic, base-labile product (Figs. S7, S8). HPLC-UV-MS analysis of extracts of the pellets revealed several chromatographic peaks with absorbance in the 300-400 nm range which were not produced by *E. coli* natively (Fig. 3a). The product profiles of heterologously-expressed *oxzAB* genes originating from different species shared many constituents, and the inclusion of genes found up- or downstream from *oxzAB* did not substantially change the product profiles.

Through a combination of mass spectrometry and spectroscopic techniques, with notable insight provided by 1,1-ADEQUATE NMR, ^{1}H-^{15}N HMBC NMR, and the identity of degradation products formed in basic methanol(48, 49), we were able to assign the products of *oxzAB* as a series of oxazolones (Figs. 3b, S9, S10). The major products were heptyl- nonyl-, undecyl-, and ω-6-undecenyltyrazolone (Tyrz, R = C7, C9, C11 and C11:1 ω-6), which were found to occur in an equilibrium of (*E*) and (*Z*) isomers with the latter being predominant, and (*Z*)-heptylphenazolone (Phez, R = 7). None of these products, the oxazolone cores of which appear to be derived from tyrosine and phenylalanine, have been reported before in the literature. However, almazolone (Fig. 3c), an oxazolone that appears to be derived from tryptophan, was isolated from a red alga(50) and its structure confirmed through synthesis. Almazolone also exists as an (*E*)/(*Z*) equilibrium, forms degradation products analogous to those formed by the tyrazolones and phenazolones, and has similar spectroscopic properties, lending credence to our structural assignments.

In vitro reconstitution of oxazolone biosynthesis by OxzAB

Considering the enzymatic domains contained within *oxzAB*, we hypothesized oxazolone biosynthesis to proceed as follows: OxzA forms an *N*-acyl amino acid using an acyl-CoA derived from the natural fatty acyl-CoA beta oxidation pool. In order to form the oxazolone, OxzB then catalyzes cyclization and oxidation of the *N*-acyl amino acid, in either order (Fig. S11). The ThiF family enzyme TcdA is known to catalyze the cyclization of an *N*-acyl amino acid, and the nitroreductase family enzyme AlbA has been shown to oxidize amino acid-derived substrates (Fig. 2a). It is thus conceivable that OxzB, which harbors both these domains, could act to form the oxazolone heterocycle.

To test our biosynthetic hypothesis, we heterologously expressed and purified *P. rubra* OxzA and OxzB in *E. coli* as N-terminal His6 and MBP fusion proteins, respectively, and assayed their activity by HPLC (Fig. 4). When incubated with tyrosine and decanoyl-CoA, OxzA catalyzed the formation of *N*-decanoyltyrosine. OxzB was able to catalyze the formation of nonyltyrazolone both from synthetic *N*-decanoyltyrosine as well as when combined with OxzA, tyrosine and decanoyl-CoA. OxzB activity was only observed in the presence of ATP, as is typical for ThiF family enzymes. (*Z*)-nonyltyrazolone, the favored isomer of the (*E*)/(*Z*) equilibrium, was the major product upon extended incubation of the enzymatic reactions, however, shorter incubation times revealed a proportionally greater amount of (*E*)-tyrazolone (Fig. S12), suggesting that the immediate product of OxzB is (*E*)-nonyltyrazolone.

Analysis of oxazolone production in the native hosts

To determine if the oxazolones are true natural products or merely artifacts of heterologous expression, we attempted to detect these molecules in the five bacteria whose *oxzAB* gene pairs we tested heterologously above. Under standard culturing conditions, oxazolones could be reliably detected only in *C. chukchiensis*. Encouraged by recent reports of chemical elicitation methods to activate silent gene clusters(51), we evaluated a series of antibiotics to induce oxazolone production in *P. rubra* or *C. chulkchiensis* at sublethal concentrations (Figs. 5, S13). Gratifyingly, we could elicit oxazolone production in *P. rubra* with erythromycin or chloramphenicol, both known protein synthesis inhibitors(52), with oxazolone titers similar to those of uninduced *C. chukchiensis*. Oxazolone production in *C. chukchiensis* could be further induced by several antibiotics, only some of which inhibit protein synthesis. We were unable to detect tyrazolones or phenazolones in extracts of the other three species whose *oxzAB* gene pair we validated heterologously. Some of the analytes detected in *P. rubra* and *C. chukchiensis* had identical masses to, but did not exactly match the retention times of the oxazolones produced heterologously (Fig. S13), perhaps suggesting the presence of oxazolones with alkyl chains derived from branched-chain fatty acyl-CoAs(53) in these organisms.

Discussion

CO-ED as a method to prioritize multi-domain enzyme for genome mining

Using a CO-ED–guided genome mining strategy, we discovered a series of novel oxazolone natural products as well as a unique oxazolone synthetase harboring ThiF and nitroreductase domains. This proof-of-principle experiment showcases the promise of CO-ED in cataloging orphan multi-domain enzymes for targeted functional discovery. Our approach to the prioritization of enzymes for genome mining considers un-annotated enzyme domain co-occurrences in an unbiased fashion. Previously, the analysis of co-occurring protein domains has been applied to evolutionary questions(54, 55), to help identify antifungal drug targets(56), and in functional gene annotation (where it is sometimes called the 'Rosetta Stone' method)(57, 58). Parallel to this, efforts have been made to connect protein domains to enzymatic reactions(59–63). CO-ED represents, to our knowledge, the first time these concepts have been united to guide the discovery of enzymes catalyzing novel transformations.

CO-ED analysis of all proteins in the Uniprot database (Fig. S14 and Datatset S1) shows that while the most abundantly distributed domain pairs have enzymes annotated in BRENDA, MIBiG, or Uniprot, there is a still a plethora of less-widespread multi-domain enzymes whose functions are unassigned: Of the 252 enzymatic domain pairs found in at least 1000 proteins, 78.6% are annotated, whereas 82.8% of the 3381 domain pairs found in at least 10 proteins are un-annotated. These annotation rates may be an under-estimate because of a lag in of new results in the databases, or an over-estimate if the same enzymatic domain pair catalyzes different transformations in different proteins. Nevertheless, there are many more multi-domain enzymes awaiting examination for new fundamental biochemical insight and as potential biocatalysts.

A newly-discovered oxazolone synthetase

Oxazolones are versatile synthetic intermediates(48), but no biocatalytic methods for their production have been reported. Besides OxzB, the only other known oxazolone-forming enzyme is MbnBC, which is involved in the biosynthesis of the chalkophore methanobactin(64). MbnBC catalyzes a 4-electron oxidative rearrangement – a fundamentally different reaction from OxzB – and acts on a peptide substrate, which is not conducive to the development of this enzyme as a biocatalyst. Considering that

P. rubra OxzB is a soluble enzyme that acts on small molecule substrates, it has the potential to be a valuable biocatalyst.

We have characterized the product profiles of five *oxzAB* gene pairs, however dozens more are present in sequence databases, ready to be explored for further biocatalytic versatility. Given the structural similarity of almazolone to the oxazolones described here, we expect a yet-undiscovered OxzB-like enzyme selective for a tryptophan derivative to be involved in almazolone biosynthesis. Some *oxzB* genes are clustered with predicted glycosyltransferases, polyketide synthases and nonribosomal peptide synthetases, suggesting they may be part of biosynthetic gene clusters coding for more complex oxazolone-containing natural products.

The reaction catalyzed by OxzB can formally be understood to consist of a cyclization, presumably catalyzed by the ThiF-like domain, and an oxidation, catalyzed by the nitroreductase domain. It is currently unclear in which order these reactions occur, and whether they occur in the same or separate active sites. OxzB is also the first characterized ThiF family enzyme that does not act on a macromolecular substrate but on a small molecule(65). We plan to gain a deeper understanding of OxzB through detailed mechanistic and structural studies.

The biological role of oxazolones

We have shown that *P. rubra* and *C. chukchiensis* can produce oxazolone natural products, but it is unclear what the biological role of these molecules is. The oxazolone-containing methanobactins are chalkophores, but this function is mediated by thiolates that are not present in the tyrazolones or phenazolones. The observation that production of these molecules can be induced by antibiotics suggests that they may act as signals in response to stress. Like many signaling molecules, oxazolones are somewhat labile in water; at pH 8 (e.g., seawater) they hydrolyze with a half-life on the order of hours. We have thus far observed no obvious phenotypes when bacteria (including the five species whose *oxzAB* genes we studied heterologously) are exposed to these molecules, both in the presence and absence of antibiotics. As we continue to explore the distribution and repertoire of oxazolone biosynthetic genes, we hope to learn more about the biological functions of these newly discovered bacterial natural products.

Methods and Code Availability

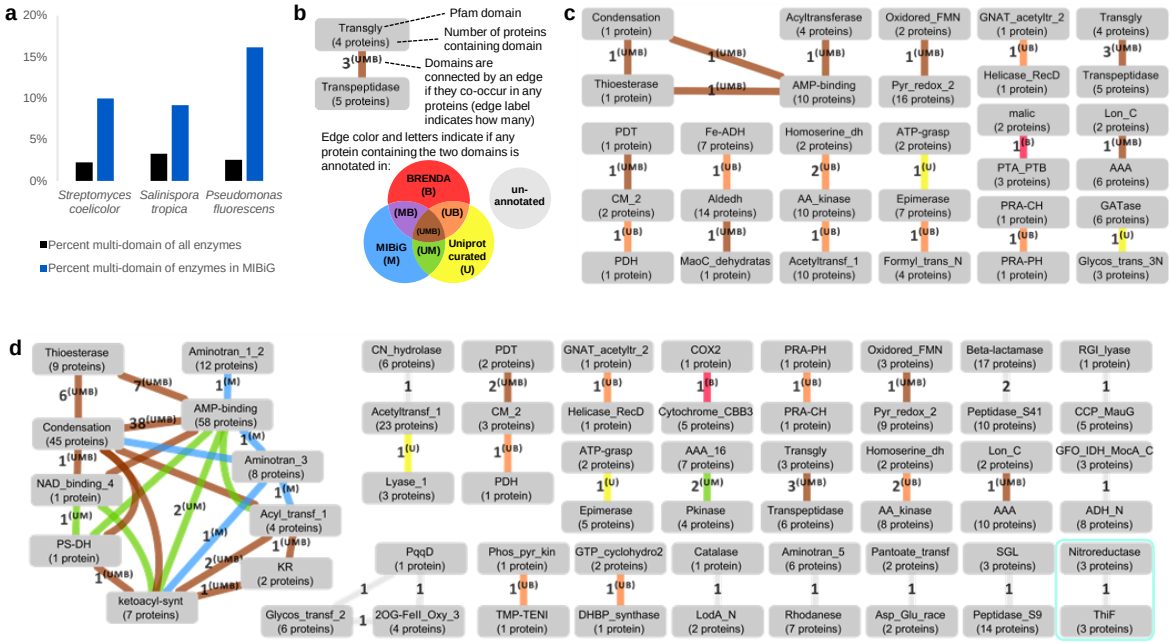
Methods, spectral data and Uniprot accession numbers for the proteins investigated are all available as part of Supplementary Information. Jupyter notebooks containing Python code for CO-ED analysis and to generate the statistics shown in Fig. 1a are publicly available at https://github.com/tderond/CO-ED. Pre-generated CO-ED networks are available as Dataset S1

Acknowledgements

We would like to thank our UCSD colleagues Jie Li and Brendan Duggan for helpful discussions. This work was supported by National Institutes of Health awards F32GM129960 to TdR and R01GM085770 to BSM, as well as the American Society for Pharmacognosy Undergraduate Research Award and the UC San Diego “Eureka” Undergraduate Research Scholarship to JEA.

Main text figures

Figure 1 (two-column)



Rationale for and application of CO-ED analysis to guide enzyme discovery

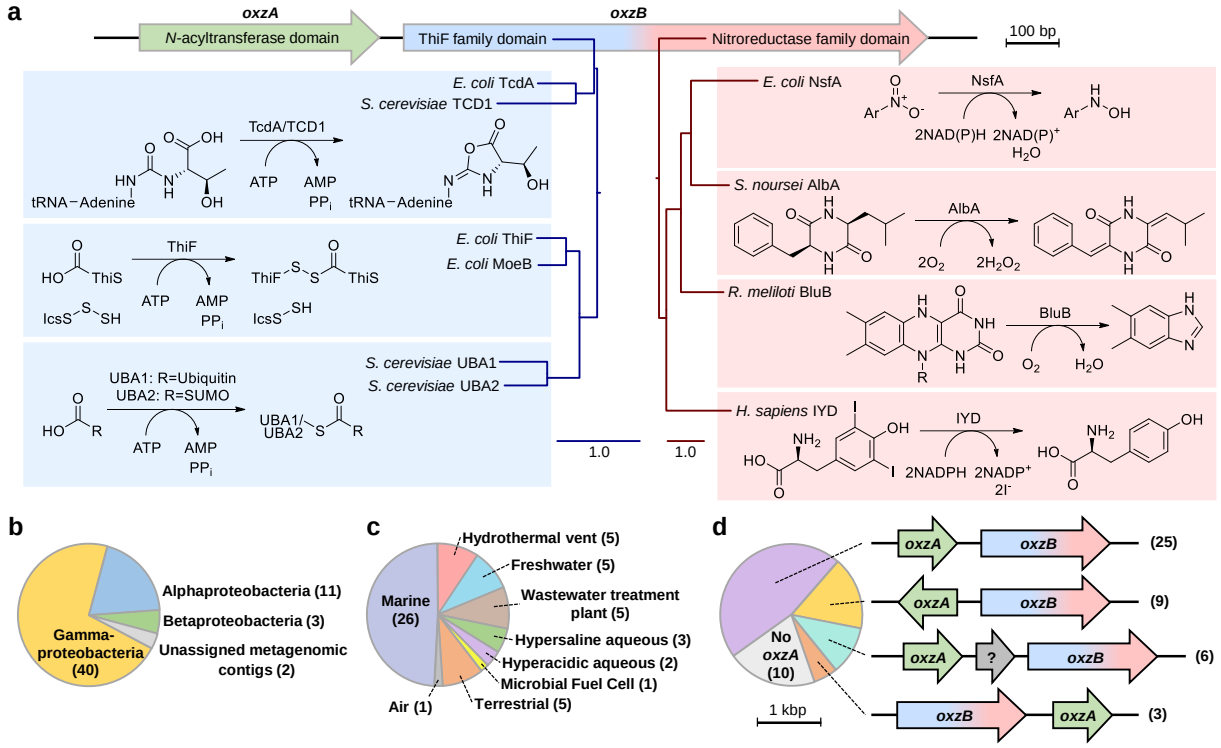
a: Multi-domain enzymes are over-represented among enzymes involved in secondary metabolism - i.e., those contained in MIBiG(27) - compared to all enzymes, in three organisms well-represented in MIBiG: *S. coelicolor* A3(2) (NCBI ASM20383v1(66), taxid 100226), *S. tropica* CNB-440 (ASM1642v1(67), taxid 369723), and *P. fluorescens* Pf-5 (ASM1226v1(68), taxid 220664)

b: Legend for the CO-ED networks shown in **c** and **d**. A curated non-redundant set of Pfam(17) domains found in enzymes were considered in the analysis. Edges are colored if any protein (from any organism, not just the proteins being analyzed) harboring both the domains the edge connects is annotated in the database. For MIBiG(27) (mibig.secondarymetabolites.org) and BRENDA(26) (www.brenda-enzymes.org), all proteins contained in the database were considered. For Uniprot(28) (www.uniprot.org), all proteins indicated to be manually curated to perform a catalytic function, as well as proteins in pathway.txt were considered. The diagrams were rendered in Cytoscape(69).

c: CO-ED network for *E. coli* K12 (NCBI ASM584v2(70)). All edges are colored, suggesting *E. coli* is of low priority for CO-ED–guided genome mining. Only the connected components are shown.

d: CO-ED network for *Pseudoalteromonas rubra* DSM 6842 (NCBI "Prub_2.0"(71)). Many edges are un-annotated (gray), suggesting there is much potential for exploring multidomain enzymes for novel biosynthetic capacity in this organism. The ThiF-Nitroreductase domain pair investigated in this study is highlighted. Only the connected components are shown.

Figure 2 (two-column)



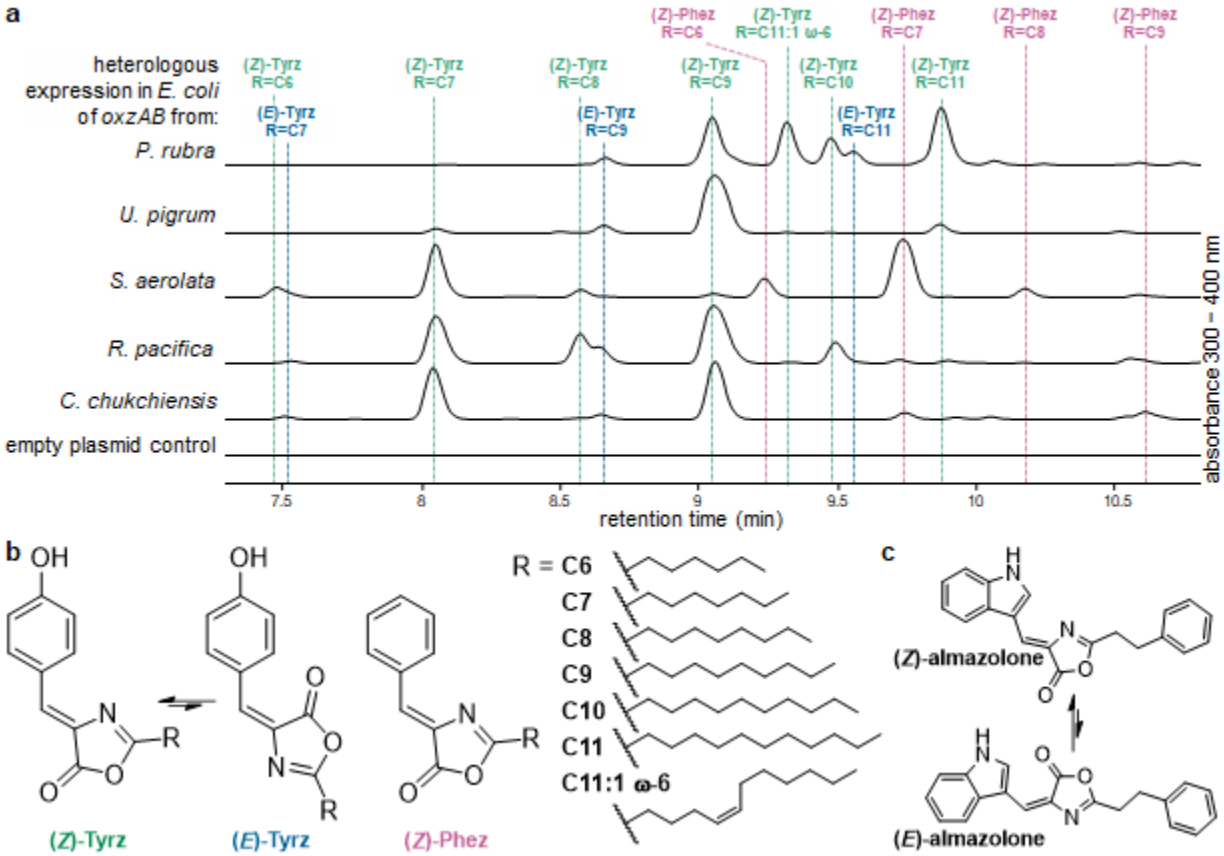
Survey of transformations known to be catalyzed by the individual domains of a di-domain enzyme identified by CO-ED, and properties of organisms harboring homologs of this enzyme.

a: OxzB exhibits homology to both the ThiF and nitroreductase enzyme families. Shown is a selection of characterized members of these enzyme families along with the transformations they catalyze, along with midpoint-rooted gene trees showing the relationships between each other and OxzB. ThiF family enzymes are known to catalyze ATP-dependent carboxylate activating reactions, while nitroreductase family enzymes catalyze a variety of redox reactions. Inferred phylogenies were generated from protein sequences using neighbor joining on the MAFFT web server(72) and midpoint-rooted. Scale bars designate 1 substitution per site.

b,c: Phylogenetic distribution (**c**) and habitat (**d**) of the host organisms harboring genes encoding OxzB proteins represented in the Uniprot database.

d: Genomic context of *oxzB* genes (wherever known), as determined using the EFI Genome Neighborhood Tool(63). Most *oxzB* genes are accompanied by *oxzA*, which codes for an *N*-acyltransferase. The arrow labeled "?" represents genes unrelated to *oxzA* or *oxzB*.

Figure 3 (two-column)

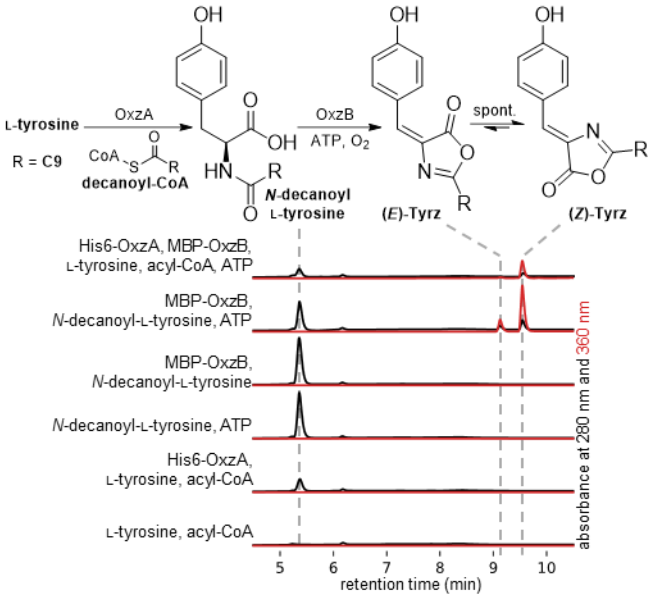


Heterologous expression of *oxzAB* results in the production of a series of oxazolones.

a: Reversed phase HPLC profiles of extracts of *E. coli* expressing *oxzAB* from various proteobacteria. An absorbance range of 300-400 nm was chosen to detect all oxazolones reported in this manuscript (under our chromatography conditions (H_2O/acetonitrile with 0.1% formic acid), the λ_{max} of the (*E*)-tyrazolones is 360 nm, that of the (*Z*)-tyrazolones is 358 nm, and that of the (*Z*)-phenazolones is 330 nm).

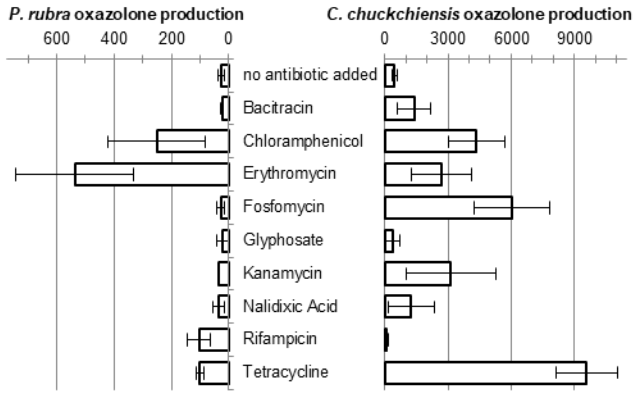
b: Structures of the tyrazolones and phenazolones, which are found as a series with varying alkyl tails. (*Z*)-Tyrz, R = C9, C10, C11 and C11:1 ω-6 and (*Z*)-Phez, R = C7 were characterized by NMR, while all others were inferred by exact mass and MS/MS fragmentation patterns. The tyrazolones exist as (*E*)/(*Z*) isomers in equilibrium. c) Almazolone, an oxazolone natural product previous isolated from an alga, also occurs as an (*E*)/(*Z*) equilibrium.

Figure 4 (one-column)



In vitro characterization of oxazolone biosynthesis by *P. rubra* OxzA and OxzB by HPLC. Black chromatograms show absorbance at 280 nm (*N*-decanoyl-L-tyrosine); red chromatograms show absorbance at 360 nm (tyrazolones) and are scaled down 4.5-fold in relation to the 280 nm signal.

Figure 5 (one-column)



Induction of oxazolone production in *P. rubra* and *C. chukchiensis* by various antibiotics. Bacteria were grown as a lawn on Marine Agar 2216 with a drop of antibiotic. A consistent amount of biomass adjacent to the zone of inhibition was harvested, extracted and analyzed by HPLC. Bars indicate summed peak areas between 300 and 400 nm. Error bars represent standard deviation of three biological replicates.

References

1. O'Brien EJ, Monk JM, Palsson BO (2015) Using genome-scale models to predict biological capabilities. *Cell* 161(5):971–987.

2. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY (2019) Current status and applications of genome-scale metabolic models. *Genome Biol* 20(1):121.

3. Bornscheuer UT (2018) The fourth wave of biocatalysis is approaching. *Philos Transact A Math Phys Eng Sci* 376(2110).

4. Gerlt JA, et al. (2011) The enzyme function initiative. *Biochemistry* 50(46):9950–9962.

5. Hanson AD, Pribat A, Waller JC, de Crécy-Lagard V (2009) Unknown proteins and orphan enzymes: the missing half of the engineering parts list – and how to find it. *Biochem J* 425(1):1–11.

6. Schnoes AM, Brown SD, Dodevski I, Babbitt PC (2009) Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comput Biol* 5(12):e1000605.

7. Scott TA, Piel J (2019) The hidden enzymology of bacterial natural product biosynthesis. *Nat Rev Chem*.

8. Ellens KW, et al. (2017) Confronting the catalytic dark matter encoded by sequenced genomes. *Nucleic Acids Res* 45(20):11495–11514.

9. Medema MH, Fischbach MA (2015) Computational approaches to natural product discovery. *Nat Chem Biol* 11(9):639–648.

10. Michael AJ (2017) Evolution of biosynthetic diversity. *Biochem J* 474(14):2277–2299.

11. Hagel JM, Facchini PJ (2017) Tying the knot: occurrence and possible significance of gene fusions in plant metabolism and beyond. *J Exp Bot* 68(15):4029–4043.

12. Weissman KJ (2015) The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* 11(9):660–670.

13. Farrow SC, Hagel JM, Beaudoin GAW, Burns DC, Facchini PJ (2015) Stereochemical inversion of (S)-reticuline by a cytochrome P450 fusion in opium poppy. *Nat Chem Biol* 11(9):728–732.

14. Winzer T, et al. (2015) Plant science. Morphinan biosynthesis in opium poppy requires a P450-oxidoreductase fusion protein. *Science* 349(6245):309–312.

15. Ng TL, Rohac R, Mitchell AJ, Boal AK, Balskus EP (2019) An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* 566(7742):94–99.

16. He H, Henderson AC, Du Y-L, Ryan KS (2019) A two-enzyme pathway links L-arginine to nitric oxide in N-nitroso biosynthesis. *J Am Chem Soc* 141(9):4026–4033.

17. El-Gebali S, et al. (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* 47(D1):D427–D432.

18. Eddy SR (2011) Accelerated profile HMM searches. *PLoS Comput Biol* 7(10):e1002195.

19. Ishino F, Mitsui K, Tamaki S, Matsuhashi M (1980) Dual enzyme activities of cell wall peptidoglycan synthesis, peptidoglycan transglycosylase and penicillin-sensitive transpeptidase, in purified preparations of Escherichia coli penicillin-binding protein 1A. *Biochem Biophys Res Commun* 97(1):287–293.

20. Goodlove PE, Cunningham PR, Parker J, Clark DP (1989) Cloning and sequence analysis of the fermentative alcohol-dehydrogenase-encoding gene of Escherichia coli. *Gene* 85(1):209–214.

21. Williams GJ, Breazeale SD, Raetz CRH, Naismith JH (2005) Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis. *J Biol Chem* 280(24):23000–23008.

22. Kaufmann M, et al. (1991) Limited proteolysis of the beta 2-dimer of tryptophan synthase yields an enzymatically active derivative that binds alpha-subunits. *Biochemistry* 30(17):4173–4179.

23. Li SL, Hanlon J, Yanofsky C (1974) Separation of anthranilate synthetase components I and II of Escherichia coli, Salmonella typhimurium, and Serratia marcescens and determination of their amino-terminal sequences by automatic Edman degradation. *Biochemistry* 13(8):1736–1744.

24. Carlomagno MS, Chiariotti L, Alifano P, Nappo AG, Bruni CB (1988) Structure and function of the Salmonella typhimurium and Escherichia coli K-12 histidine operons. *J Mol Biol* 203(3):585–606.

25. Rusnak F, Sakaitani M, Drueckhammer D, Reichert J, Walsh CT (1991) Biosynthesis of the Escherichia coli siderophore enterobactin: sequence of the entF gene, expression and purification of EntF, and analysis of covalent phosphopantetheine. *Biochemistry* 30(11):2916–2927.

26. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D (2019) BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res* 47(D1):D542–D549.

27. Kautsar SA, et al. (2020) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res* 48(D1):D454–D458.

28. The UniProt Consortium (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res* 46(5):2699.

29. Offret C, et al. (2016) Spotlight on antimicrobial metabolites from the marine bacteria Pseudoalteromonas: chemodiversity and ecological significance. *Mar Drugs* 14(7).

30. De Rond T, et al. (2017) Oxidative cyclization of prodigiosin by an alkylglycerol monooxygenase-like enzyme. *Nat Chem Biol* 13(11):1155–1157.

31. Agarwal V, et al. (2014) Biosynthesis of polybrominated aromatic organic compounds by marine bacteria. *Nat Chem Biol* 10(8):640–647.

32. Ross AC, Gulland LES, Dorrestein PC, Moore BS (2015) Targeted capture and heterologous expression of the Pseudoalteromonas alterochromide gene cluster in Escherichia coli represents a promising natural product exploratory platform. *ACS Synth Biol* 4(4):414–420.

33. Herz S, Eberhardt S, Bacher A (2000) Biosynthesis of riboflavin in plants. The ribA gene of Arabidopsis thaliana specifies a bifunctional GTP cyclohydrolase II/3,4-dihydroxy-2-butanone 4-phosphate synthase. *Phytochemistry* 53(7):723–731.

34. Schulman BA, Harper JW (2009) Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* 10(5):319–331.

35. Miyauchi K, Kimura S, Suzuki T (2013) A cyclic form of N6-threonylcarbamoyladenosine as a widely distributed tRNA hypermodification. *Nat Chem Biol* 9(2):105–111.

36. Xi J, Ge Y, Kinsland C, McLafferty FW, Begley TP (2001) Biosynthesis of the thiazole moiety of thiamin in Escherichia coli: identification of an acyldisulfide-linked protein--protein conjugate that is functionally analogous to the ubiquitin/E1 complex. *Proc Natl Acad Sci U S A* 98(15):8513–8518.

37. Godert AM, Jin M, McLafferty FW, Begley TP (2007) Biosynthesis of the thioquinolobactin siderophore: an interesting variation on sulfur transfer. *J Bacteriol* 189(7):2941–2944.

38. Akiva E, Copp JN, Tokuriki N, Babbitt PC (2017) Evolutionary and molecular foundations of multiple contemporary functions of the nitroreductase superfamily. *Proc Natl Acad Sci U S A* 114(45):E9549–E9558.

39. Mondal S, Raja K, Schweizer U, Mugesh G (2016) Chemistry and biology in the biosynthesis and action of thyroid hormones. *Angew Chem Int Ed Engl* 55(27):7606–7630.

40. Schneider TL, Shen B, Walsh CT (2003) Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* 42(32):9722–9730.

41. Melby JO, Li X, Mitchell DA (2014) Orchestration of enzymatic processing by thiazole/oxazole-modified microcin dehydrogenases. *Biochemistry* 53(2):413–422.

42. Taga ME, Larsen NA, Howard-Jones AR, Walsh CT, Walker GC (2007) BluB cannibalizes flavin to form the lower ligand of vitamin B12. *Nature* 446(7134):449–453.

43. Bashiri G, et al. (2016) Elongation of the poly-γ-glutamate Tail of F420 requires both domains of the F420:γ-glutamyl ligase (FbiB) of mycobacterium tuberculosis. *J Biol Chem* 291(13):6882–6894.

44. Bashiri G, et al. (2019) A revised biosynthetic pathway for the cofactor F420 in prokaryotes. *Nat Commun* 10(1):1558.

45. Braga D, Hasan M, Kröber T, Last D, Lackner G (2020) Biosynthesis of the redox coenzyme F420 in Thermomicrobia involves reduction by standalone nitroreductase superfamily enzymes. *Appl Environ Microbiol*.

46. Gondry M, et al. (2001) Cyclic dipeptide oxidase from Streptomyces noursei. Isolation, purification and partial characterization of a novel, amino acyl alpha,beta-dehydrogenase. *Eur J Biochem* 268(6):1712–1721.

47. Zenno S, Saigo K, Kanoh H, Inouye S (1994) Identification of the gene encoding the major NAD(P)H-flavin oxidoreductase of the bioluminescent bacterium Vibrio fischeri ATCC 7744. *J Bacteriol* 176(12):3536–3543.

48. Fisk JS, Mosey RA, Tepe JJ (2007) The diverse chemistry of oxazol-5-(4H)-ones. *Chem Soc Rev* 36(9):1432–1440.

49. Deering RW, et al. (2016) N-Acyl dehydrotyrosines, tyrosinase inhibitors from the marine bacterium Thalassotalea sp. PP2-459. *J Nat Prod* 79(2):447–450.

50. Guella G, N'Diaye I, Fofana M, Mancini I (2006) Isolation, synthesis and photochemical properties of almazolone, a new indole alkaloid from a red alga of Senegal. *Tetrahedron* 62(6):1165–1170.

51. Seyedsayamdost MR (2014) High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci U S A* 111(20):7266–7271.

52. Walsh C, Wencewicz T (2016) *Antibiotics: challenges, mechanisms, opportunities* (American Society of Microbiology, Washington, DC). 2nd Ed.

53. Kaneda T (1991) Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev* 55(2):288–302.

54. Wuchty S, Almaas E (2005) Evolutionary cores of domain co-occurrence networks. *BMC Evol Biol* 5:24.

55. Wang Z, et al. (2011) A protein domain co-occurrence network approach for predicting protein function and inferring species phylogeny. *PLoS ONE* 6(3):e17906.

56. Barrera A, Alastruey-Izquierdo A, Martín MJ, Cuesta I, Vizcaíno JA (2014) Analysis of the protein domain and domain architecture content in fungi and its application in the search of new antifungal targets. *PLoS Comput Biol* 10(7):e1003733.

57. Suhre K (2007) Inference of gene function based on gene fusion events: the rosetta-stone method. *Methods Mol Biol* 396:31–41.

58. Promponas VJ, Ouzounis CA, Iliopoulos I (2014) Experimental evidence validating the computational inference of functional associations from gene fusion events: a critical survey. *Brief Bioinformatics* 15(3):443–454.

59. Alcántara R, et al. (2013) The EBI enzyme portal. *Nucleic Acids Res* 41(Database issue):D773–80.

60. Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44–57.

61. George RA, Spriggs RV, Thornton JM, Al-Lazikani B, Swindells MB (2004) SCOPEC: a database of protein catalytic domains. *Bioinformatics* 20 Suppl 1:i130–6.

62. Alborzi SZ, Devignes M-D, Ritchie DW (2017) ECDomainMiner: discovering hidden associations between enzyme commission numbers and Pfam domains. *BMC Bioinformatics* 18(1):107.

63. Gerlt JA (2017) Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* 56(33):4293–4308.

64. Kenney GE, et al. (2018) The biosynthesis of methanobactin. *Science* 359(6382):1411–1416.

65. Cappadocia L, Lima CD (2018) Ubiquitin-like protein conjugation: structures, chemistry, and mechanism. *Chem Rev* 118(3):889–918.

66. Bentley SD, et al. (2002) Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature* 417(6885):141–147.

67. Udwary DW, et al. (2007) Genome sequencing reveals complex secondary metabolome in the marine actinomycete Salinispora tropica. *Proc Natl Acad Sci U S A* 104(25):10376–10381.

68. Paulsen IT, et al. (2005) Complete genome sequence of the plant commensal Pseudomonas fluorescens Pf-5. *Nat Biotechnol* 23(7):873–878.

69. Shannon P, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13(11):2498–2504.

70. Blattner FR, et al. (1997) The complete genome sequence of Escherichia coli K-12. *Science* 277(5331):1453–1462.

71. Xie B-B, et al. (2012) Genome sequence of the cycloprodigiosin-producing bacterial strain Pseudoalteromonas rubra ATCC 29570(T). *J Bacteriol* 194(6):1637–1638.

72. Katoh K, Rozewicki J, Yamada KD (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinformatics*:bbx108.