



Exploring modular allostery via interchangeable regulatory domains

Yifei Fan^{a,b}, Penelope J. Cross^{a,b}, Geoffrey B. Jameson^c, and Emily J. Parker^{b,d,e,1}

^aBiomolecular Interaction Centre, University of Canterbury, 8140 Christchurch, New Zealand; ^bDepartment of Chemistry, University of Canterbury, 8140 Christchurch, New Zealand; ^cMaurice Wilkins Centre, Institute of Fundamental Sciences, Massey University, 4442 Palmerston North, New Zealand; ^dMaurice Wilkins Centre, Biomolecular Interaction Centre, University of Canterbury, 8140 Christchurch, New Zealand; and ^eFerrier Research Institute, Victoria University of Wellington, 6140 Wellington, New Zealand

Edited by Perry Allen Frey, University of Wisconsin–Madison, Madison, WI, and approved February 8, 2018 (received for review October 12, 2017)

Most proteins comprise two or more domains from a limited suite of protein families. These domains are often rearranged in various combinations through gene fusion events to evolve new protein functions, including the acquisition of protein allostery through the incorporation of regulatory domains. The enzyme 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAH7PS) is the first enzyme of aromatic amino acid biosynthesis and displays a diverse range of allosteric mechanisms. DAH7PSs adopt a common architecture with a shared $(\beta/\alpha)_8$ catalytic domain which can be attached to an ACT-like or a chorismate mutase regulatory domain that operates via distinct mechanisms. These respective domains confer allosteric regulation by controlling DAH7PS function in response to ligand Tyr or prephenate. Starting with contemporary DAH7PS proteins, two protein chimeras were created, with interchanged regulatory domains. Both engineered proteins were catalytically active and delivered new functional allostery with switched ligand specificity and allosteric mechanisms delivered by their nonhomologous regulatory domains. This interchangeability of protein domains represents an efficient method not only to engineer allostery in multidomain proteins but to create a new bifunctional enzyme.

protein engineering | allostery | bifunctional enzyme | shikimate | synthetic biology

Protein allostery is central to the regulation of many biological processes, including ligand transport and metabolic function and control. Allostery occurs when ligand binding at one site communicates with a remote functional site, resulting in a change in protein function. Allosteric regulation of protein functions often involves a complex network of interactions to deliver signals between distal sites. Signal communication is achieved via diverse mechanisms ranging from large conformational changes to subtle changes in protein dynamics (1–3). The understanding of these remote communications is of great interest, particularly in the fields of drug design and protein engineering (4, 5).

Most proteins contain two or more domains (6). These domains and their interactions govern the function of a protein and are considered evolutionary units for modular assembly of new protein architectures (7–9). Biological data suggest that only a limited number of protein folds exist in nature, and protein functions evolve from mutation, duplication, and recombination of ancestral genes under selective pressure (10). Domain recombination via gene fusion events represents one of the major pathways for the evolution of allostery (11–14). This principle has been recognized in many natural systems (12).

The enzyme 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase (DAH7PS) is the first enzyme of the shikimate pathway for the biosynthesis of aromatic compounds. DAH7PS catalyzes the divalent metal-dependent condensation of two carbohydrate precursors, phosphoenolpyruvate (PEP) and D-erythrose 4-phosphate (E4P), to produce DAH7P and phosphate. This enzyme is essential for most microorganisms because of its key role in the biosynthesis of the intermediate chorismate from which the pathway branches to

allow formation of prephenate by chorismate mutase (CM) as well as a series of other important metabolites, including aromatic amino acids Phe, Tyr, and Trp (Fig. 1) (15). DAH7PS, situated at the first committed step of the pathway, is often precisely feedback regulated to control pathway flux in response to metabolic demand via a variety of allosteric strategies in different organisms. This allostery ranges from physical gating mechanisms undertaking large conformational changes to highly intricate through-protein dynamic networks (16). The levels of chorismate, prephenate, and the aromatic amino acids are important check points for feedback regulation of the DAH7PS function (16).

Most DAH7PS enzymes are tetrameric in solution, with each chain comprising a $(\beta/\alpha)_8$ catalytic barrel. This barrel is frequently decorated with additional structural elements, which are responsible for conferring allostery (16). One major class of DAH7PS enzymes (type I β) share a catalytic domain of a structurally uninterrupted classic TIM barrel $(\beta/\alpha)_8$ fold, and are either unregulated or allosterically regulated (Fig. 24). The simplest form is composed solely of the barrel without any additional domain, and hence is unregulated. This form of DAH7PS enzymes has been characterized from *Pyrococcus furiosus* and *Aeropyrum pemix* (17–19). The regulated DAH7PSs in this group display discrete domains appended to the catalytic barrel at the N or C terminus. These domains are either an ACT-like domain (20) or a CM (AroQ) domain, which possesses CM enzymatic activity. The attached ACT-like and CM domains deliver allostery on a tetrameric DAH7PS scaffold via

Significance

Protein functions are often evolved from recombination of a limited suite of protein folds. Engineering desired function into proteins via manipulation of the genes, mimicking natural evolutionary processes, represents a promising approach to synthesizing useful molecular tools. Allosteric proteins play important roles in many biological processes, often as regulatory switches or metabolic checkpoints. In this study, analysis of two natural allosteric proteins from different organisms revealed comparable yet distinct regulatory mechanisms via conformation rearrangement. Interchanging gene segments encoding for the nonhomologous regulatory elements of the two proteins results in new functional proteins with switched regulatory behavior. This modular approach represents a step toward a general scheme for construction of functional changes in proteins.

Author contributions: Y.F., P.J.C., and E.J.P. designed research; Y.F. and P.J.C. performed research; Y.F., P.J.C., G.B.J., and E.J.P. analyzed data; and Y.F., G.B.J., and E.J.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

¹To whom correspondence should be addressed. Email: emily.parker@vuw.ac.nz.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1717621115/-DCSupplemental.

Published online March 5, 2018.

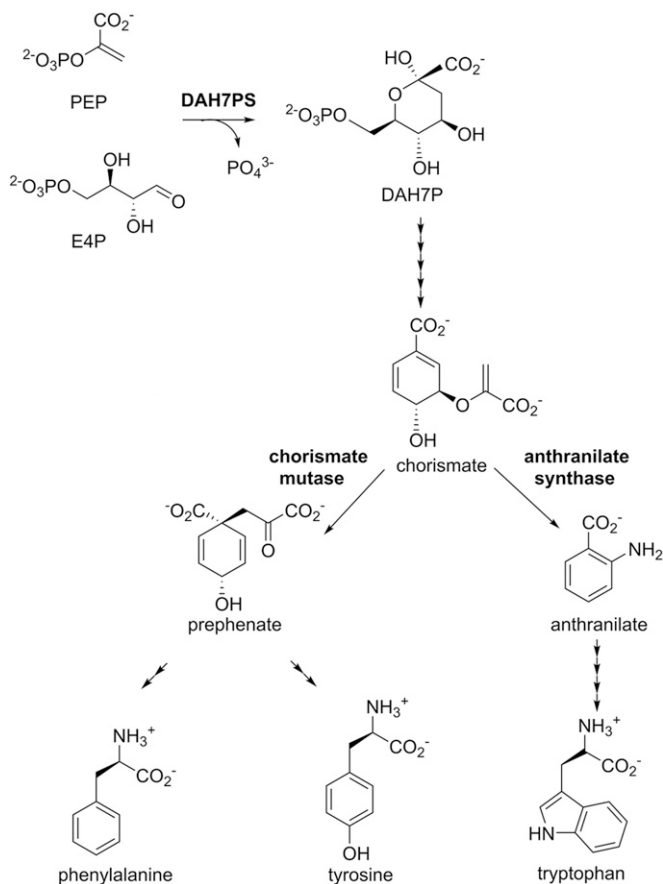


Fig. 1. The shikimate pathway leads to biosynthesis of aromatic amino acids Trp, Phe, and Tyr. DAH7PS catalyzes the first reaction. The pathway branches at chorismate, where CM catalyzes the conversion of chorismate to prephenate en route to formation of Phe and Tyr. The number of arrowheads represents the number of reactions.

physical gating of the active site associated with the binding of an allosteric ligand (Fig. 2 *B* and *C*) (21–23). This is best exemplified in the well-characterized DAH7PS from *Thermotoga maritima* (Fig. 2*B*), which undergoes a remarkable conformational change in response to the presence of Tyr, in which the ACT-like domains from opposing chains dimerize to form a binding site for the allosteric ligand and block substrate access to the active site (21). This enzyme is also inhibited by Phe, although to a reduced extent. Similarly, prephenate binding to the CM domain in DAH7PS from *Geobacillus* sp. (strain Y412MC61) is associated with the more intimate association between the CM and DAH7PS domains to form a more compact structure and limit catalysis (Fig. 2*C*) (23).

Previously, we have demonstrated the transfer of the allosteric domain of the *T. maritima* DAH7PS onto the unadorned, unregulated DAH7PS of *P. furiosus* to confer an allosteric response in the latter DAH7PS (24). We propose that the recruitment of a regulatory domain is a general strategy that, by itself, is sufficient for providing allosteric control of enzymes. To validate this hypothesis and to demonstrate that the key information associated with delivery of allostery resides in the regulatory domain, we explore here the interchangeability of the two distinct regulatory strategies employed by different allosterically controlled DAH7PS enzymes through construction of protein variants that mix and match catalytic and regulatory domains of DAH7PSs from *T. maritima* and *Geobacillus* sp. These studies illustrate the remarkable ease with which functional allostery can be acquired by gene fusion events and provide insight into the evolution of modular allostery, in

which existing ligand-binding domains or enzymes can be repurposed to provide allostery.

Results

Altered Regulatory and Catalytic Domain Combinations Deliver Functional Proteins. The wild-type parent proteins *Tma*DAH7PS and *Gsp*DAH7PS share similar homotetrameric quaternary structures and tertiary structures, with each chain composed of an N-terminal regulatory domain (respectively, ACT and CM) attached to a catalytic (β/α)₈ barrel housing the active site. The *Tma*DAH7PS and *Gsp*DAH7PS catalytic barrels share a moderate sequence identity of 56% (Fig. S1). Two protein variants were created with exchanged regulatory and catalytic domains based on the analysis of wild-type sequences and architectures. Both these domain-swapped variants delivered DAH7PS catalysis, albeit with some alterations in their catalytic efficiencies (Table 1). Compared with the wild-type *Tma*DAH7PS, *Gsp*CM-*Tma*DAH7PS, which shares the same catalytic core, exhibited impaired activity with significantly decreased k_{cat}/K_M values for both PEP and E4P substrates, implying there may be some restriction of access for substrates to the catalytic center introduced by the fused CM domains. The origin of this may lie in the fact that the CM domain is dimeric both in the absence and presence of the allosteric ligand, whereas the ACT-like domain only dimerizes upon ligand binding (21, 23). Hence, perhaps unsurprisingly, the adoption of the more restricted *Gsp*CM domain to the *Tma*DAH7PS catalytic core is accompanied by attenuation of catalysis, whereas *Tma*ACT-*Gsp*DAH7PS displayed an approximately twofold boost in catalytic efficiency compared with the wild-type *Gsp*DAH7PS. That physical constraints restrict catalysis by the DAH7PS core barrel is supported by the higher activity displayed by truncated forms of

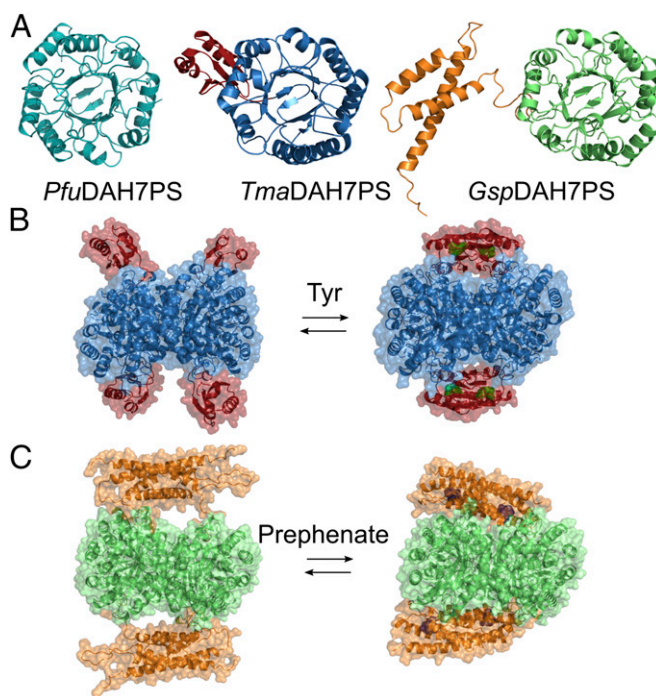


Fig. 2. Structure and allostery of DAH7PS. (A) Monomeric units of DAH7PS from *P. furiosus* (Left), *T. maritima* (Center), and *Geobacillus* sp. (Right) share an uninterrupted catalytic barrel. ACT domain in *Tma*DAH7PS is shown in red; CM domain in *Gsp*DAH7PS is shown in orange. Both *Tma*DAH7PS (B) and *Gsp*DAH7PS (C) display significant conformational changes upon ligand binding to the regulatory domain (Tyr is shown as green spheres, and prephenate is shown as purple spheres).

Table 1. Kinetics parameters for the chimeras and parent proteins

Protein	DAH7PS activity					CM activity		
	K_M^{PEP} , μM	K_M^{E4P} , μM	k_{cat} , s^{-1}	k_{cat}/K_M^{PEP} , $\mu\text{M}^{-1}\text{s}^{-1}$	k_{cat}/K_M^{E4P} , $\mu\text{M}^{-1}\text{s}^{-1}$	$K_M^{\text{chorismate}}$, μM	k_{cat} , s^{-1}	k_{cat}/K_M , $\mu\text{M}^{-1}\text{s}^{-1}$
<i>Gsp</i> CM- <i>Tma</i> DAH7PS	34 ± 2.7	19 ± 1.8	4.8 ± 0.1	0.14	0.25	190 ± 16	4.1 ± 0.1	0.02
<i>Tma</i> ACT- <i>Gsp</i> DAH7PS	57 ± 1.6	54 ± 1.4	60 ± 2	1.05	1.11	NA	NA	NA
<i>Gsp</i> DAH7PS	45 ± 4	61 ± 8	28 ± 0.9	0.62	0.46	98 ± 7	3.4 ± 0.1	0.03
<i>Tma</i> DAH7PS	8.4 ± 0.7	15 ± 1	14 ± 0.3	1.67	0.93	NA	NA	NA

NA, not applicable for the specified enzyme activity.

*Tma*DAH7PS and *Gsp*DAH7PS in which, respectively, the ACT and CM domains have been removed (21, 23).

Unlike the ACT domain, which has only a ligand-binding role, the regulatory domain of *Gsp*DAH7PS has both catalytic and regulatory functions (23). The CM catalytic activity was comparably maintained when the domain was transferred to the alternative catalytic barrel in *Gsp*CM-*Tma*DAH7PS (Table 1), although this chimera displayed an approximately twofold increase in K_M for chorismate compared with wild-type *Gsp*DAH7PS. As expected, with no CM domain, the variant *Tma*ACT-*Gsp*DAH7PS displayed only DAH7PS activity.

As with all wild-type DAH7PS enzymes characterized to date, the catalytic activities of the chimeras depend on presence of a divalent metal ion (Fig. 3A). Mn^{2+} is the most activating metal ion for the wild-type *Tma*DAH7PS, whereas Cd^{2+} delivers maximal activity for *Gsp*DAH7PS (21, 23). A range of metal ions were tested with the protein chimeras. Largely as expected, each of the DAH7PS catalytic cores reflected its inherent metal preference, with *Tma*ACT-*Gsp*DAH7PS highly favoring Cd^{2+} , whereas *Gsp*CM-*Tma*DAH7PS showed more than 90% activity in the presence of Mn^{2+} or Cd^{2+} .

Consistent with the thermophilic properties of the wild-type enzymes, both enzymes became more active at elevated temperatures, reaching their optimal activity at temperatures above 60 °C (Fig. 3B). Intriguingly, activity of *Tma*ACT-*Gsp*DAH7PS was enhanced significantly at elevated temperatures compared with the wild-type *Gsp*DAH7PS. On the other hand, *Gsp*CM-*Tma*DAH7PS appeared to be least active across all temperatures. Proteins with the *Gsp*CM domain, chimeric and wild-type, generally tend to display lower activities than proteins with the *Tma*ACT domain at the temperatures tested. This change in activity profile may relate to the structural difference in the regulatory domains and the different inherent optimal temperatures of the parent proteins.

Allosteric Inhibitor Preference Resides in the Regulatory Domain. To test the effect of potential allosteric ligands, inhibition assays were performed for both DAH7PS chimeras and compared with the response of the wild-type proteins (Fig. 4 and Fig. S2). Consistent with the presence of a regulatory domain in both chimeras, inhibition of both enzymes was observed. Presenting the ACT domain, the DAH7PS activity of *Tma*ACT-*Gsp*DAH7PS was reduced by 70% at high Tyr concentrations and by ~30% at high Phe concentrations. This difference in ligand sensitivity for Tyr and Phe is comparable with that of the wild-type *Tma*DAH7PS, for which Tyr exhibits more significant inhibitory effect than Phe, although the IC_{50} value of Tyr for the chimera (210 μM) is almost 10-fold higher than that of the wild-type protein (22.5 μM). The presence of the CM domain in *Gsp*CM-*Tma*DAH7PS delivered sensitivity toward prephenate, although the inhibition was less profound than that for the wild-type *Gsp*DAH7PS. The presence of 77 μM prephenate reduced DAH7PS activity to 50% of its uninhibited value, compared with 20 μM required for the same attenuation of wild-type protein catalytic activity. The maximum level of inhibition was also altered; the chimeric protein retained 31% activity at higher prephenate concentrations in comparison with the 4% residual activity observed for the wild-type enzyme.

Binding of allosteric ligands stabilizes the parent proteins and the chimeras (Fig. 5 and Fig. S3). In the presence of inhibitors, the melting temperatures of both chimeras increased by around 3 °C, in close agreement with the degree of stabilization for the wild-type proteins. Interestingly, the thermostability of *Tma*ACT-*Gsp*DAH7PS improved almost 20 °C compared with the wild-type *Gsp*DAH7PS. On the other hand, the stability of the *Gsp*CM-*Tma*DAH7PS appeared to be largely determined by the *Tma*DAH7PS core, with the denaturing event occurring at above 96 °C. The thermostability profiles exhibited by the chimeras largely reflect the difference in inherent thermostability of the wild-type proteins, remarkably, even when the transferred domain is only a fraction (approximately one-third) of the core DAH7PS barrel, as in *Tma*ACT-*Gsp*DAH7PS.

The Structural Changes of Allostery Are Transferred with the Regulatory Domains. The allosteric mechanisms of both parental proteins involve conformational rearrangements as observed in solution by small-angle X-ray scattering (SAXS) experiments (21, 23). To assess the allosteric mechanisms employed by the chimeras, SAXS experiments were performed. In the presence of Tyr, *Tma*ACT-*Gsp*DAH7PS became more compact, with a reduced R_g value (derived from Guinier plot) of 32.8 ± 0.2 Å compared with the R_g of the apo form 34.0 ± 0.2 Å (Fig. 6 and Table S2). The Kratky plot displayed a more defined curve in the presence of Tyr than without Tyr, consistent with the decreased flexibility of the protein. The *Gsp*CM-*Tma*DAH7PS also showed a conformational change with addition of prephenate, with R_g decreasing from 36.9 ± 0.2 Å to 34.8 ± 0.2 Å, consistent also with the more defined Kratky plot.

To compare these conformational changes observed in the protein chimeras with those observed for the wild-type proteins, the SAXS profiles were fitted with the calculated theoretical scattering from crystal structures or homology model of each parent protein (Fig. 6 and Table S2). In the absence of Tyr, scattering profiles of *Tma*ACT-*Gsp*DAH7PS presented a good fit with the open form of the *Tma*DAH7PS crystal structure, but not the closed form. Conversely, in the presence of Tyr, the scattering profile

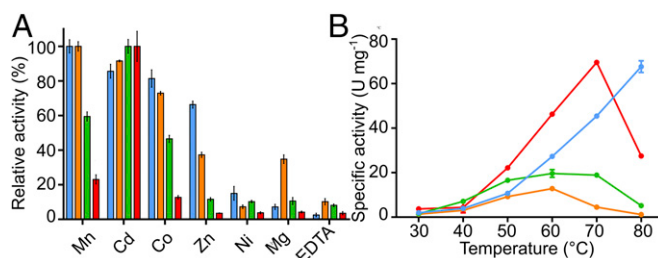


Fig. 3. Catalytic activity of the parent proteins and chimeras. (A) DAH7PS activities of *Tma*DAH7PS (blue), *Gsp*DAH7PS (green), *Tma*ACT-*Gsp*DAH7PS (red), and *Gsp*CM-*Tma*DAH7PS (orange), in the presence metal ions or EDTA. (B) The effect of temperature on specific activity of the four enzymes in the same color coding as in A.

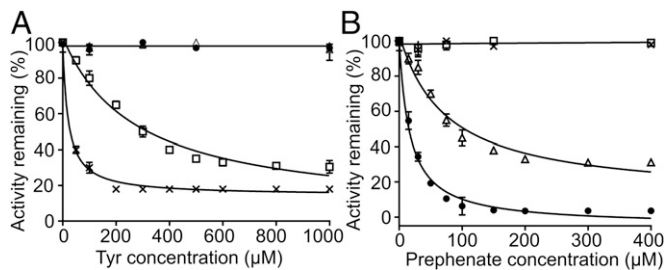


Fig. 4. Inhibition of DAH7PS activity for the parent proteins and chimeras. (A) *Tma*DAH7PS (x) and *Tma*ACT-*Gsp*DAH7PS (□) displayed sensitivity toward Tyr. (B) *Gsp*DAH7PS (●) and *Gsp*CM-*Tma*DAH7PS (△) displayed impaired catalysis with addition of prephenate.

of *Tma*ACT-*Gsp*DAH7PS gave a better fit with the closed, Tyr-bound form of *Tma*DAH7PS than with the open, apo form. In the absence of prephenate, the scattering profile of *Gsp*CM-*Tma*DAH7PS fits closely with the open-form model of the wild-type *Gsp*DAH7PS and poorly with the closed, prephenate-bound form of *Gsp*DAH7PS. In the presence of prephenate, *Gsp*CM-*Tma*DAH7PS is better fit by the closed, prephenate-bound form of *Gsp*DAH7PS than by the open, apo form.

Discussion

Design of proteins with new properties and functions is an important goal of biotechnology. The modular feature of many natural proteins suggests that common building domains and modules are likely to have the evolutionary advantages of being autonomous and portable, offering simplicity in recombination to generate new functions by allowing transfer of information through their interactions (25). Statistical coupling analysis revealed that the recombination of protein domains relies on the networks of coevolving amino acids involved in the allosteric communication and that these networks display strong connectivity, proposing the feasibility of engineering artificial allosteric systems by transfer of allosteric networks between proteins (5, 26). Although this transferability has not been extensively explored, a few studies indicate its plausibility. For example, a modulating maltose-binding protein was fused to an unrelated enzyme β -lactamase, introducing maltose regulation to the unregulated β -lactamase (27–29). In another example, ultrasensitive molecular switches were built by modular recombination of multiple SH3-peptide autoinhibitory interactions on WASP to introduce strong cooperativity with respect to the ligand SH3 (30, 31). Also, an *Escherichia coli* dihydrofolate reductase (DHFR) was coupled with a light-sensing protein from plants to generate a light-sensitive DHFR (5). These studies demonstrate that methods of mix and match between modular components can be effective ways to create substantial functional changes associated with allostery and to expand the repertoire of artificial proteins.

What we demonstrate here is the interchangeability of regulatory domains in a homooligomeric protein. By simple gene recombination of contemporary sequences, we can interchange the Tyr-binding ACT domain on one DAH7PS with a structurally nonhomologous prephenate-binding CM domain from another DAH7PS. This interchange functionally swaps allosteric regulation in both DAH7PS enzymes elicited by formation of, or changes to, a dimeric structure upon binding of the appropriate allosteric effector. What is somewhat surprising is that gene fusion to functionally link these contemporary domains requires no modifications to the core catalytic barrels. Fusion of the regulatory domain nevertheless inflicts some catalytic penalty on the barrel, and the degree of compensated activity may be associated with the nature of the extended domains. We note that the functional domain-swapped chimeras developed here are quite distinct from other examples where chimeras have been

generated by the modification of discrete structural elements on structurally homologous domains to alter allosteric ligand specificity (32–35). We have interchanged two structurally and functionally nonhomologous regulatory modules of contemporary DAH7PS scaffolds. That fully functional enzymes are generated highlights the conservation of allosteric strategy in these contemporary proteins, which is delivered by structurally diverse solutions (or gene fusions) (21, 23).

Despite the ease of functional allostery interchange, there are enabling features of the contemporary DAH7PS scaffold that are important in accommodating the recombination and for maintaining allosteric networks. Core barrel oligomerization is a prerequisite for allostery that involves dimerization of the regulatory module (17, 21, 23, 36). The DAH7PS barrel homotetramer adopts an overall conformation that supports the allosteric function, with diagonally opposite chains delivering the regulatory domain on either side of the tetramer plane (Fig. 2). This tetrameric assembly is a feature of type I β DAH7PS, which is shared by both regulated and unregulated DAH7PSs (19, 21, 23, 37) and a related enzyme, 3-deoxy-D-manno-2-octulosonate 8-phosphate synthase (38). Intriguingly, quaternary structure is delicately balanced in this DAH7PS enzyme class. Removal of the ACT-like domain from *Tma*DAH7PS results in dimerization and the (unregulated) *Pfu*DAH7PS is rendered dimeric by a single amino acid substitution (19, 21). Association of dimers into tetramers was probably followed by acquisition of terminal regulatory domains (17). Thus, it appears that the functional gene fusion to deliver allostery was facilitated serendipitously by the adoption of the appropriate homotetrameric catalytic template.

The barrels are not highly conserved overall, but there do appear to be some sequence elements that are associated with delivering allostery in both systems that are shared between both *Gsp*DAH7PS and *Tma*DAH7PS. The key hydrogen-bond contacts between either the ACT domain or the CM domain with the catalytic DAH7PS domain reside mostly at the C terminus of the diagonally adjacent barrel (Fig. S4). In the ligand-bound form of *Tma*DAH7PS, major hydrogen bonds are formed between an ACT domain and the barrel from the adjacent chain, including Asp51 and Asp309, Ser55 and Arg277, and Asp57 and Glu304. Due to the asymmetrical nature of the domain arrangement across the vertical plane in the crystal structure of

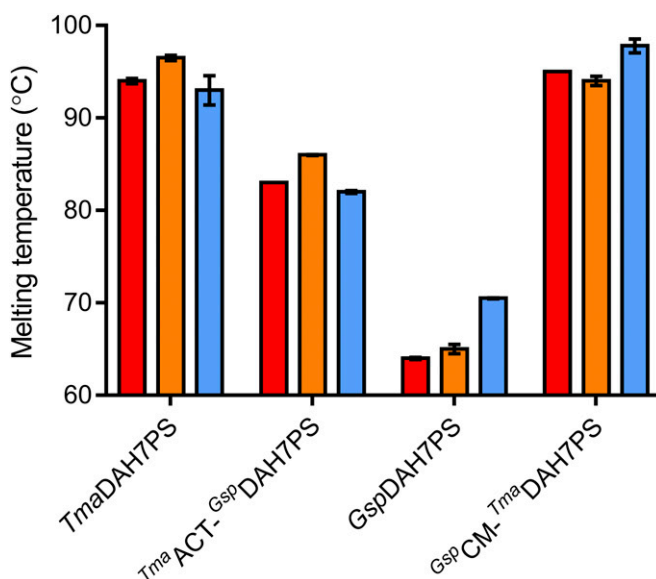


Fig. 5. Thermostability of the parent proteins and chimeras in the absence (red) and presence of Tyr (orange) or prephenate (blue).

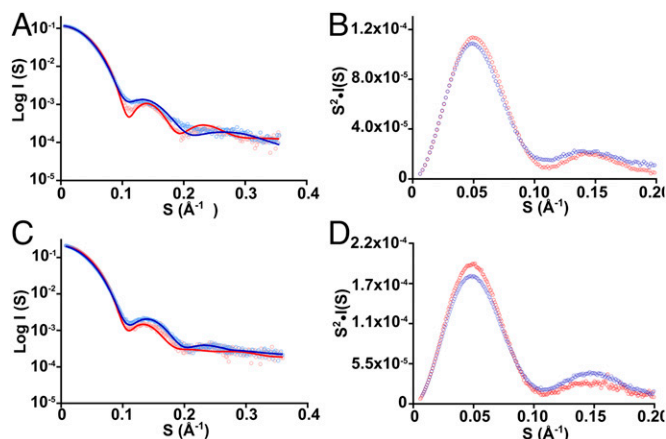


Fig. 6. SAXS profiles of the chimeras. (A) SAXS profiles of $TmaACT-GspDAH7PS$ were measured in the absence (light blue circles) and presence (pink circles) of Tyr. Theoretical scatterings were calculated from Tyr-free (blue line) and Tyr-bound (red line) $TmaDAH7PS$ crystal structures. (B) Kratky plot of $TmaACT-GspDAH7PS$ SAXS profiles in the absence (blue circles) and presence (red circles) of Tyr. (C) SAXS profiles of $GspCM-TmaDAH7PS$ were measured in the absence (light blue circles) and presence (pink circles) of prephenate. Theoretical scatterings were calculated from prephenate-free model (blue line) and prephenate-bound (red line) crystal structure of $GspDAH7PS$. (D) Kratky plot of $GspCM-TmaDAH7PS$ SAXS profiles in the absence (blue circles) and presence (red circles) of prephenate. I , scattering intensity; S , magnitude of scattering vector (where $S = 4\pi\sin\theta/\lambda$, 2θ is the scattering angle and λ is the wavelength of the radiation).

$GspDAH7PS$, each CM polypeptide chain communicates with its adjacent barrel through different hydrogen-bonding interactions. Key interactions are found between the $\alpha1-\alpha2$ loop of the CM domain and the C terminus of the barrel, including Arg52 and Asp326, as well as Tyr46 and Glu29 on one end of the barrel, plus Gly42 and Arg157 on the other end.

Sequence alignment between $TmaDAH7PS$ and $GspDAH7PS$ suggests that these key regions on the catalytic barrel involved in the communication of the allosteric information are highly similar (Fig. S1). Indeed, these regions are generally well conserved across the type I β family (Fig. S4C), which is consistent with the involvement of these contact regions in DAH7PS catalysis. All contact regions are located on the top side of the barrel close to the active site. Key residues of the $\beta2-\alpha2$ and $\beta8-\alpha8$ loops form direct hydrogen-bonding interactions with the substrates PEP and E4P, and segments of loop $\beta1-\alpha1$ and $\beta7-\alpha7$ appear to secure the correct positioning of those residues that interact with substrates. Together, the DAH7PS barrel appears to be a versatile platform for additional functions, which tend to be developed conveniently by hijacking the conserved features of the active site so that allostery can be created with ease when interdomain interactions are reestablished in the presence of ligand. Additionally, some conservation of residue character for the N-terminal regulatory domains (charged, polar, and nonpolar) may also contribute to the establishment of an allosteric network between DAH7PS and regulatory domain, despite very different secondary and tertiary structures of the ACT and CM domains. The linkers between the DAH7PS and its regulatory domain share a remarkably high degree of conservation, possibly because of the conformational restrictions for the hinging motions occurring on binding the allosteric effectors. This conservation in the flexible linker region may be indicative of its functional importance. Protein stability may have also contributed to the success of domain swapping, as proteins with enhanced stability are believed to have greater capacity for modifications (39). These underlying features and modular scaffold of the DAH7PS family appear to support the acquisition of functional variations.

In addition to the transfer of allostery, enzymatic activity of the CM domain was also transferred into the $GspCM-TmaDAH7PS$ chimera. Remarkably, thermostability was determined by the more thermophilic component of the chimera, even when that component, as in the case of the ACT domain of the DAH7PS from the hyperthermophilic *T. maritima*, was transferred onto the much larger DAH7PS domain from the less thermophilic *Geobacillus* sp.—the proverbial tail wagging the dog. The generality of this phenomenon as a mechanism for inducing thermostability into more mesophilic enzymes remains to be elucidated.

From one enzyme and two distinct regulatory domains we have four allosterically regulated enzymes, two occurring in nature and two created in the laboratory, one of which is additionally a bifunctional enzyme. The engineering of allosteric control, enzymatic activities, and thermostability by domain swapping illustrates that homologous (β/α)₈ structures can tolerate new combinations with structurally and functionally nonhomologous regulatory domains. Future investigations into the details of the corresponding allosteric network experimentally and computationally are important for the design and optimization of allosteric systems with high performance.

Materials and Methods

Design and Preparation of Protein Variants. Crystal structures of the $TmaDAH7PS$ and $GspDAH7PS$ suggest similar architecture and arrangement of the homotetramers, with the catalytic barrels connected to their respective N-terminal regulatory domains via a β -hairpin and a flexible linker region. The linker region is crucial for the interaction between regulatory and catalytic domains in both enzymes and to allow appropriate conformational change between active and inhibited states (21). The amino acid regions that incorporate regulatory domains, linker regions, and catalytic domains of $TmaDAH7PS$ and $GspDAH7PS$ are identified individually from sequence and structure alignments (Fig. S1). Two protein variants with exchanged regulatory and catalytic domains were designed: $TmaACT-GspDAH7PS$, with the regulatory domain and linker region from $TmaDAH7PS$ (residues 1 to 93) and the catalytic domain from $GspDAH7PS$ (residues 118 to 362); and the complementary chimera $GspCM-TmaDAH7PS$, with the regulatory domain and linker region from $GspDAH7PS$ (residues 1 to 117) and the catalytic domain from $TmaDAH7PS$ (residues 94 to 338). Both constructs were created by amplifying each segment from the corresponding parent wild-type gene with overlap and by fusing the amplified products (Table S1). The fused gene fragments encoding each chimeric protein were cloned using Gateway[®] technique. $TmaACT-GspDAH7PS$ was transformed in pDEST14 and expressed in BL21(DE3)* cells, and its expression and purification were performed following previously published procedures for $TmaDAH7PS$ (21). $GspCM-TmaDAH7PS$ was transformed in pDEST15 and expressed in BL21(DE3)pLysS cells using the same expression conditions, and purified using an GSTRap HP column following published procedures for the truncated $GspDAH7PS$ (23).

Enzyme Kinetics. Kinetics parameters for DAH7PS and/or CM activities of the chimeric proteins were determined following previously described procedures by measuring consumption of chorismate at 274 nm or PEP at 232 nm, with each reaction containing 0.05 μ M enzyme (23). For DAH7PS assays, the concentration of E4P was varied between 5 μ M and 400 μ M while PEP was held at 295 μ M, and the concentration of PEP was varied between 6 μ M and 350 μ M while E4P was fixed at 310 μ M. Reactions contained enzyme, PEP, and 100 μ M Mn^{2+} (for proteins containing the catalytic barrel from *T. maritima*) or Cd^{2+} (for proteins containing the catalytic barrel from *Geobacillus* sp.), and were equilibrated in 50 mM 1,3-bis[tris(hydroxymethyl)methylamino] propane (BTP) buffer at 60 $^{\circ}$ C (pH 7.4) before E4P was added to initiate the reaction. For CM assays, 9 to 400 μ M chorismate was used to initiate the reaction in the same BTP buffer at 50 $^{\circ}$ C. Metal dependency studies were performed using the same DAH7PS assay described above in the presence of saturating PEP and E4P, with a range of metal ions or EDTA at 100 μ M. For inhibition studies, assay solutions contained 283 μ M PEP, 308 μ M E4P, 100 μ M Mn^{2+} or Cd^{2+} , and 0 to 1 mM Tyr/Phe or 0 to 400 μ L prephenate in 50 mM BTP (pH 7.4 at 60 $^{\circ}$ C).

Thermal Activity and Stability Measurements. Activity profiles of the chimeric proteins and parent proteins at elevated temperatures (30 to 80 $^{\circ}$ C) were assessed using the same methods as for kinetics measurements of DAH7PS activity. Each reaction contained 0.05 μ M enzyme, 100 μ M appropriate metal

ion, 215 μM PEP, and 227 μM E4P in 50 mM BTP buffer. All buffers were at pH 7.4 at the temperatures tested. Enzyme activities are specified in U (1 U = consumption of 1 μmol substrate per minute). Specific enzyme activities are given as $\text{U}\cdot\text{mg}^{-1}$. Melting temperatures were determined by using differential scanning fluorimetry with an iCycler iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad). Protein samples ($0.1\text{ mg}\cdot\text{mL}^{-1}$) were prepared in storage size-exclusion chromatography buffer with SYPRO Orange dye in the absence or presence of ligands (1 mM Tyr/Phe or 500 μM prephenate). Controls were similarly prepared for each condition, with buffer in place of protein. The temperature was set to increase from 20 to 98 $^{\circ}\text{C}$ at a rate of 1 $^{\circ}\text{C}/\text{min}$, and fluorescence was measured in 0.2 $^{\circ}\text{C}$ increments.

SAXS. SAXS measurements were collected at the Australian Synchrotron SAXS/wide-angle X-ray scattering (WAXS) beamline using the setup previously described (21, 23). Scattering data were collected at 25 $^{\circ}\text{C}$ following elution of the protein samples (9 mg/mL) from a size-exclusion chromatography column (Superdex 200_Increase 5/150), which was pre-equilibrated with buffer containing 10 mM BTP (pH 7.4), 100 mM KCl, and 200 μM PEP, in the absence or presence of inhibitors (1 mM Tyr or 500 μM prephenate). Raw data were processed as described for wild-type proteins using Scatterbrain

(21). Scattering intensity versus magnitude of scattering vector of each protein was generated with Primus, and plots from apo- and liganded proteins were scaled (40). Theoretical scattering profiles were generated from published crystal structures of wild-type *Tma*DAH7PS 1RZM (apo) and 3PG9 (Tyr), homology model of wild-type *Gsp*DAH7PS (apo), and crystal structure 5J6F (prephenate) (21, 23, 37). They were fitted with corresponding experimental scatterings using CRY SOL (41).

Sequence Analysis. Sequence alignments of the wild-type *Pfu*DAH7PS, *Tma*DAH7PS, and *Gsp*DAH7PS were generated with Clustal Omega (42) and formatted with ESPrnt 3 (43). A total of 2,523 sequences from the DAH7PS family of interest (type I β) were obtained from Pfam (PF00793) (44) and aligned using SeaView 4 (45), and conserved residues were identified and graphed with WebLogo3 (46).

ACKNOWLEDGMENTS. Y.F. is a recipient of a University of Canterbury Doctoral Scholarship. This work was supported by grants from the Marsden Fund (UOC1105), New Zealand Synchrotron Group Ltd., and the SAXS/WAXS beamline at the Australian Synchrotron.

- Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. *Nature* 402(6761, Suppl):C47–C52.
- Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. *Nature* 508:331–339.
- Goodey NM, Benkovic SJ (2008) Allosteric regulation and catalysis emerge via a common route. *Nat Chem Biol* 4:474–482.
- Swain JF, Gierasch LM (2006) The changing landscape of protein allostery. *Curr Opin Struct Biol* 16:102–108.
- Lee J, et al. (2008) Surface sites for engineering allosteric control in proteins. *Science* 322:438–442.
- Ma B, Tsai CJ, Haliloglu T, Nussinov R (2011) Dynamic allostery: Linkers are not merely flexible. *Structure* 19:907–917.
- Doolittle RF (1995) The multiplicity of domains in proteins. *Annu Rev Biochem* 64:287–314.
- Patthy L (2003) Modular assembly of genes and the evolution of new functions. *Genetica* 118:217–231.
- Vogel C, Bashton M, Kerrison ND, Chothia C, Teichmann SA (2004) Structure, function and evolution of multidomain proteins. *Curr Opin Struct Biol* 14:208–216.
- Chothia C (1992) Proteins. One thousand families for the molecular biologist. *Nature* 357:543–544.
- Peracchi A, Mozzarelli A (2011) Exploring and exploiting allostery: Models, evolution, and drug targeting. *Biochim Biophys Acta* 1814:922–933.
- Bashton M, Chothia C (2007) The generation of new protein functions by the combination of domains. *Structure* 15:85–99.
- Ostermeier M, Benkovic SJ (2000) Evolution of protein function by domain swapping. *Adv Protein Chem* 55:29–77.
- Pasek S, Risler JL, Brézellec P (2006) Gene fusion/fission is a major contributor to evolution of multi-domain bacterial proteins. *Bioinformatics* 22:1418–1423.
- Bentley R (1990) The shikimate pathway—A metabolic tree with many branches. *Crit Rev Biochem Mol Biol* 25:307–384.
- Light SH, Anderson WF (2013) The diversity of allosteric controls at the gateway to aromatic amino acid biosynthesis. *Protein Sci* 22:395–404.
- Schofield LR, et al. (2005) Substrate ambiguity and crystal structure of *Pyrococcus furiosus* 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase: An ancestral 3-deoxyald-2-ulosonate-phosphate synthase? *Biochemistry* 44:11950–11962.
- Zhou L, et al. (2012) Structure and characterization of the 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase from *Aeropyrum pernix*. *Bioorg Chem* 40:79–86.
- Nazmi AR, Schofield LR, Dobson RCJ, Jameson GB, Parker EJ (2014) Destabilization of the homotetrameric assembly of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from the hyperthermophile *Pyrococcus furiosus* enhances enzymatic activity. *J Mol Biol* 426:656–673.
- Lang EJ, Cross PJ, Mittelstädt G, Jameson GB, Parker EJ (2014) Allosteric ACTion: The varied ACT domains regulating enzymes of amino-acid metabolism. *Curr Opin Struct Biol* 29:102–111.
- Cross PJ, Dobson RC, Patchett ML, Parker EJ (2011) Tyrosine latching of a regulatory gate affords allosteric control of aromatic amino acid biosynthesis. *J Biol Chem* 286:10216–10224.
- Light SH, Halavaty AS, Minasov G, Shuvalova L, Anderson WF (2012) Structural analysis of a 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase with an N-terminal chorismate mutase-like regulatory domain. *Protein Sci* 21:887–895.
- Nazmi AR, et al. (2016) Interdomain conformational changes provide allosteric regulation en route to chorismate. *J Biol Chem* 291:21836–21847.
- Cross PJ, Allison TM, Dobson RCJ, Jameson GB, Parker EJ (2013) Engineering allosteric control to an unregulated enzyme by transfer of a regulatory domain. *Proc Natl Acad Sci USA* 110:2111–2116.
- Pawson T, Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* 300:445–452.
- Hatley ME, Lockless SW, Gibson SK, Gilman AG, Ranganathan R (2003) Allosteric determinants in guanine nucleotide-binding proteins. *Proc Natl Acad Sci USA* 100:14445–14450.
- Ostermeier M (2005) Engineering allosteric protein switches by domain insertion. *Protein Eng Des Sel* 18:359–364.
- Betton J-M, Jacob JP, Hofnung M, Broome-Smith JK (1997) Creating a bifunctional protein by insertion of β -lactamase into the maltodextrin-binding protein. *Nat Biotechnol* 15:1276–1279.
- Guntas G, Ostermeier M (2004) Creation of an allosteric enzyme by domain insertion. *J Mol Biol* 336:263–273.
- Dueber JE, Mirsky EA, Lim WA (2007) Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol* 25:660–662.
- Dueber JE, Yeh BJ, Chak K, Lim WA (2003) Reprogramming control of an allosteric signaling switch through modular recombination. *Science* 301:1904–1908.
- Cunin R, Rani CS, Van Vliet F, Wild JR, Wales M (1999) Intramolecular signal transmission in enterobacterial aspartate transcarbamylases II. Engineering co-operativity and allosteric regulation in the aspartate transcarbamylase of *Erwinia herbicola*. *J Mol Biol* 294:1401–1411.
- Pawlyk AC, Pettigrew DW (2002) Transplanting allosteric control of enzyme activity by protein-protein interactions: Coupling a regulatory site to the conserved catalytic core. *Proc Natl Acad Sci USA* 99:11115–11120.
- Page MJ, Carrell CJ, Di Cera E (2008) Engineering protein allostery: 1.05 A resolution structure and enzymatic properties of a Na⁺-activated trypsin. *J Mol Biol* 378:666–672.
- Helmstaedt K, Heinrich G, Lipscomb WN, Braus GH (2002) Refined molecular hinge between allosteric and catalytic domain determines allosteric regulation and stability of fungal chorismate mutase. *Proc Natl Acad Sci USA* 99:6631–6636.
- Hartmann M, et al. (2003) Evolution of feedback-inhibited beta/alpha barrel isoenzymes by gene duplication and a single mutation. *Proc Natl Acad Sci USA* 100:862–867.
- Shumilin IA, Bauerle R, Wu J, Woodard RW, Kretsinger RH (2004) Crystal structure of the reaction complex of 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase from *Thermotoga maritima* refines the catalytic mechanism and indicates a new mechanism of allosteric regulation. *J Mol Biol* 341:455–466.
- Allison TM, et al. (2011) An extended $\beta\alpha 7$ substrate-binding loop is essential for efficient catalysis by 3-deoxy-D-manno-octulosonate 8-phosphate synthase. *Biochemistry* 50:9318–9327.
- Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103:5869–5874.
- Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) PRIMUS: A Windows PC-based system for small-angle scattering data analysis. *J Appl Crystallogr* 36:1277–1282.
- Svergun D, Barberato C, Koch MH (1995) CRY SOL—A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773.
- Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539.
- Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42:W320–W324.
- Finn RD, et al. (2016) The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res* 44:D279–D285.
- Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224.
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: A sequence logo generator. *Genome Res* 14:1188–1190.