nature chemical biology

Article

https://doi.org/10.1038/s41589-022-01206-0

Designer installation of a substrate recruitment domain to tailor enzyme specificity

Received: 27 March 2022

Accepted: 10 October 2022

Published online: 12 December 2022

Check for updates

Rodney Park[1], Chayanid Ongpipattanakul[2,7], Satish K. Nair[2,3,4], Albert A. Bowers[5] ✉ & Brian Kuhlman[1,6] ✉

Promiscuous enzymes that modify peptides and proteins are powerful tools for labeling biomolecules; however, directing these modifications to desired substrates can be challenging. Here, we use computational interface design to install a substrate recognition domain adjacent to the active site of a promiscuous enzyme, catechol *O*-methyltransferase. This design approach effectively decouples substrate recognition from the site of catalysis and promotes modification of peptides recognized by the recruitment domain. We determined the crystal structure of this novel multidomain enzyme, SH3-588, which shows that it closely matches our design. SH3-588 methylates directed peptides with catalytic efficiencies exceeding the wild-type enzyme by over 1,000-fold, whereas peptides lacking the directing recognition sequence do not display enhanced efficiencies. In competition experiments, the designer enzyme preferentially modifies directed substrates over undirected substrates, suggesting that we can use designed recruitment domains to direct post-translational modifications to specific sequence motifs on target proteins in complex multisubstrate environments.

Enzymes are incredibly proficient at modifying specific proteins and peptides in complex biological environments[1–3]. Specificity is often conferred by recognition surfaces or domains that are physically distinct from the active site. Src family kinases are an excellent example of this theory. They contain three domains: the kinase domain, a Src homology 2 (SH2) domain and a Src homology 3 (SH3) domain. In some cases, the kinase domain alone is sequence promiscuous and phosphorylates diverse tyrosine-containing peptides, while the SH2 and SH3 domains can enhance specificity by binding to unique sequence motifs on substrates[4–7] (Fig. 1a). Separation of substrate recognition and catalysis plays a similarly important role in natural product biosynthesis where ribosomally synthesized and post-translationally modified peptides (RiPPs) undergo a series of directed enzymatic modifications to yield mature natural products[8,9]. Recruitment occurs when a separate domain of the enzyme binds a recognition sequence in the peptide substrate and directs the core to the active site for modification (Fig. 1b). An evolutionary advantage of separating binding and catalysis is that enzymes with new specificity profiles can be readily generated by combining pre-existing promiscuous catalytic domains with pre-existing peptide-binding domains[3,10].

Protein engineers frequently take advantage of the modular nature of enzymes. This is exemplified by the many engineered proteins that have been created for genome editing[11,12]. Typically, a protein domain with high binding specificity for a unique DNA sequence is genetically

[1]Department of Biochemistry and Biophysics, University of North Carolina School of Medicine, Chapel Hill, NC, USA. [2]Department of Biochemistry, University of Illinois at Urbana–Champaign, Urbana, IL, USA. [3]Center for Biophysics and Computational Biology, University of Illinois at Urbana–Champaign, Urbana, IL, USA. [4]Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL, USA. [5]Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [6]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [7]Present address: School of Pharmacy, University of California San Francisco, San Francisco, CA, USA. ✉e-mail: abower2@email.unc.edu; bkuhlman@email.unc.edu

https://doi.org/10.1038/s41589-022-01206-0

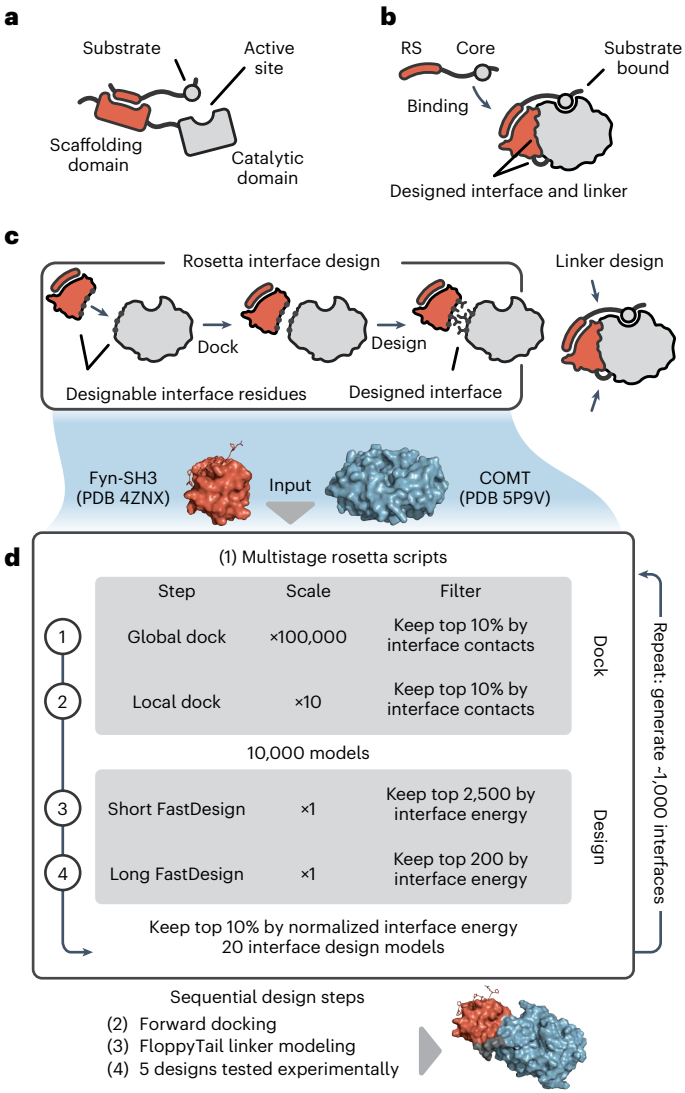


Fig. 1 | General scaffolding diagrams and Rosetta design approach. a, The flexible scaffolding approach used by our naive fusion enzyme and Src family kinases. **b**, Directed scaffolding. The diagram shows a specific interaction between a recruitment domain and promiscuous enzyme, similar to how many RiPP enzymes are constructed. RS indicates the recognition sequence on the peptide that is bound by the peptide-binding domain. **c**, Overview of the Rosetta design pipeline. Peptide-binding domains are shown in red while promiscuous enzymes are shown in gray. **d**, Multistage Rosetta scripts protocol (design protocol dock_design_sh3.xml in Supplementary Information) and follow-up design steps.

fused to a promiscuous DNA-modifying enzyme[13]. Similar strategies have been employed to target protein substrates[10,14–16]. In these studies, fusion proteins have been created by connecting a substrate recruitment domain to the catalytic domain with a simple, flexible linker[15,16]. This is a straightforward approach employed by evolution, as evidenced by the flexible linkers connecting the domains in Src family kinases[4]. However, some natural enzymes, such as many RiPP enzymes, use specifically positioned recognition domains to anchor and direct the peptide substrate to the active site[17,18]. The role of this kind of recruitment domain placement in catalysis has not been studied extensively, in part because methods have not been available for designing multidomain proteins that adopt a well-defined tertiary structure. Strategic positioning of a recruitment domain relative to an active site may influence catalysis in a few ways: (1) affinity for the substrate (the Michaelis constant (K_m)) may be enhanced by increasing the local concentration near the active site[3,5,7]; (2) the turnover number (that is, the maximum number of chemical conversions of substrate molecules per second that a single active site will execute; k_{cat}) may be perturbed if specific binding orientations are required for catalysis or binding and unbinding rates are rate limiting[15]; and (3) modification site selectivity could probably be controlled by which substrate residues are directed toward the active site.

Here, we test whether the catalytic efficiency and specificity of a promiscuous enzyme can be enhanced by installing a substrate scaffolding domain in a specific position relative to its active site (Fig. 1c). Our approach takes advantage of developments in the field of computational protein design, which now enable the engineering of predefined interfaces between proteins[19–23]. As a model system, we use catechol *O*-methyltransferase (COMT) as the promiscuous enzyme and the SH3 domain of human Fyn tyrosine kinase (Fyn-SH3) as the peptide-binding domain[24]. In nature, COMT transfers a methyl group from *S*-adenosyl methionine (SAM) to catechols such as L-dihydroxyphenylalanine (L-DOPA) or dopamine[25]. However, COMT can be repurposed to install a variety of chemical groups on DOPA-containing peptides[26]. SH3 domains bind to well-defined poly-proline motifs, and the Fyn-SH3 domain is robust to mutation, making it amenable to design[27]. We show that we can create a multidomain protein with the SH3 domain placed as designed, and that the engineered protein enhances catalytic efficiency for poly-proline-containing peptides by over 1,000-fold. Furthermore, we test and model the performance of our designed enzyme in multisubstrate environments and find that it largely prefers directed substrates.

Results

Rosetta interface design of multidomain proteins

To create a multidomain protein from Fyn-SH3 and COMT, we employed a two-step computational protocol that first redesigned the surfaces of the two proteins to create a favorable interaction and then connected their termini to produce a single-chain construct (Fig. 1c). Interface design was performed using iterative rounds of protein–protein docking and sequence optimization using the molecular modeling program Rosetta. We biased docking to favor inter-domain contacts between well-ordered secondary structural elements on the two domains, in particular along the solvent-exposed faces of helices 1 and 2 of COMT and β-strands 2–4 of Fyn-SH3. Placing these structural elements in proximity brought the domain termini near each other and directed the carboxy (C) terminus of peptides bound to the SH3 domain toward the COMT active site. To stabilize the docked conformation, we employed two-sided interface design, where contacting residues on both protein surfaces were allowed to mutate to form well-packed hydrophobic interactions and hydrogen bonds. Because both the docking and design protocols in Rosetta are stochastic, thousands of independent dock and design trajectories with stages of filtering and focusing (Fig. 1d) were used to identify low-energy models.

After sorting the models by calculated energies and removing designs that did not satisfy a set of quality metrics (see Methods), we evaluated the top ~50 designs by manual inspection and selected five designs for experimental studies. The designs averaged 11 mutations across the two domains (ranging from three to seven mutations on each domain). In a last design step, a short glycine–serine linker was added to connect the C terminus of Fyn-SH3 to COMT to create a single multidomain enzyme. All five of the selected designs favored similar domain placement and peptide proximity to the active site, with the C terminus of the poly-proline motif ~30 Å from the sulfur on SAM. The designed interfaces were predominately hydrophobic, with small hydrogen bond networks in three of the models.

Biophysical and structural characterization of the designs

All five designed enzymes expressed and were soluble. Of these, two designs, SH3-588 and SH3-003, did not aggregate and were found to be

https://doi.org/10.1038/s41589-022-01206-0

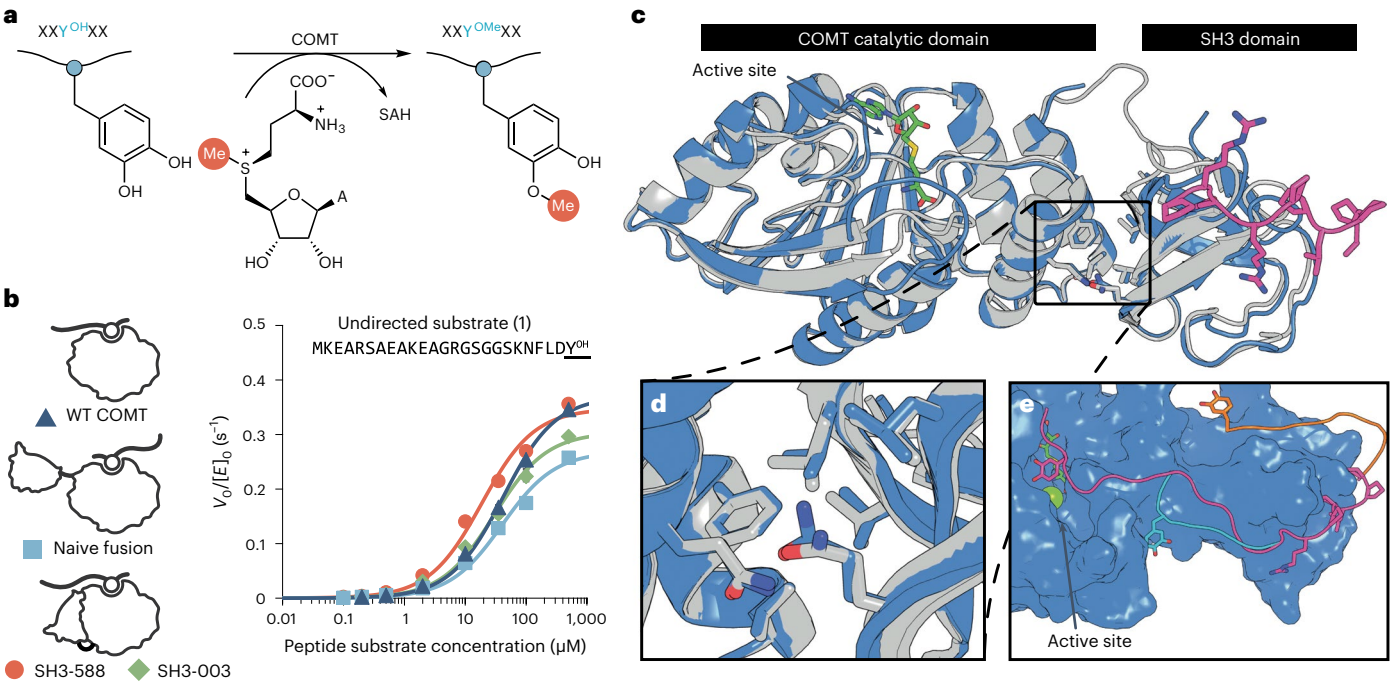


Fig. 2 | Structural characterization and background activity of the designed enzymes. a, Conversion of DOPA to methylated DOPA by COMT. **b**, Plot of rate (normalized reaction rate given by the ratio of initial reaction rate (V_0) over enzyme concentration ($[E]_0$)) versus substrate concentration for WT COMT, naive fusion, SH3-588 and SH3-003 for an undirected peptide substrate (peptide 1). All three enzymes have similar catalytic parameters with undirected substrate, indicating that catalytic activity has not been compromised by the design process. Rates are presented as points calculated from slopes of best fit from three distinct reactions ($n = 3$). The error bars represent ±s.e. of the best fit centered on the calculated rate (an example is provided in Supplementary Fig. 3a–d). **c**, Crystal structure of SH3-588 (blue; PDB code: 7UD6) structurally aligned to its design model (gray). SAH (green) was observed in the crystal structure. The APP12 peptide is modeled bound to the SH3 domain (pink) to show relative placement near the active site. **d**, Alignment of the interface residues between the SH3 and COMT domains (boxed area in **c**), showing that packing at the interface is very similar in the crystal structure and design model. **e**, Substrate peptide modeling on the crystal structure (a short peptide 7 (teal), medium peptide 2 (pink) and misdirected peptide 3 (orange)), indicating stretch of the residues from the peptide-binding domain to the active site. A magnesium ion (lime green) indicates the location of the active site but is not present in the SH3-588 crystal structure or during modeling.

monomeric by multi-angle light scattering coupled with size exclusion chromatography (Supplementary Fig. 1), while the remainder were oligomeric or precipitated during purification. For comparative studies, we additionally expressed and purified a naive fusion of the two proteins in which the C terminus of wild-type (WT) Fyn-SH3 was connected to the amino (N) terminus of WT COMT with a glycine–serine linker (Fig. 1a).

Binding studies with a tetramethylrhodamine (TAMRA)-labeled poly-proline peptide, TAMRA-APP12 (peptide S1), confirmed that the naive fusion, SH3-588 and SH3-003 all bind to the recognition sequence with similar affinities. The measured dissociation constant (K_d) values were 330, 690 and 670 nM, respectively, with no binding to WT COMT (Supplementary Fig. 2). Similar affinities were also measured using a competitive binding experiment that did not require direct labeling of the peptide with TAMRA. To verify that the design process did not disrupt catalytic activity, we measured methylation rates with a peptide substrate, MKEARSAEAKEAGRGSGGSKNFLDY^{OH} (peptide 1), that includes a C-terminal L-DOPA residue (Y^{OH}) but lacks a recognition sequence. We used a luciferase-coupled assay that monitors the production of *S*-adenosylhomocysteine (SAH), a byproduct of the reaction (Fig. 2a). Similar k_{cat} (between 0.27 and 0.37 s^{-1}) and K_m values (between 19 and 42 μM) were measured for SH3-588, SH3-003, the naive fusion and WT COMT (Table 1 and Fig. 2b). This indicates that the active site residues remain undisturbed after design and suggests that all of the enzymes engage peptide 1 directly through their active sites.

To determine whether SH3-588 adopts the designed conformation, we solved its crystal structure to a resolution of 2.6 Å. The crystal structure closely matches our design model (backbone root-mean-square deviation = 0.543 Å), with identical orientation of the catalytic and binding domains (Fig. 2c) and similar interface surface areas between the design model and crystal structure (904 and 848 $Å^2$, respectively). Many of the side chain conformations at the interface were accurately predicted (Fig. 2d), including a designed hydrogen bond between residues Gln^{46} (from the SH3 domain) and Gln^{93} (from COMT). Meanwhile, the glycine–serine linker connecting the two domains is not resolved in the structure. This, along with the well-ordered interface contacts, indicates that the designed interface predominately stabilizes the multidomain enzyme, not the linker. Critically, the peptide-binding site in the SH3 domain is accessible and alignment of the substrate peptide onto SH3-588 shows that the C terminus of the peptide is pointed toward the active site. We were unable to crystallize our second design, SH3-003.

K_m enhancement by substrate scaffolding

Molecular modeling with SH3-588 and peptide substrates suggested that at least nine residues are necessary to span between the SH3 domain and the active site (Fig. 2e); therefore, we designed a directed substrate EW**APPLPPR**NRPRRSGGSKET**Y^{OH}**SK (peptide 2, where the bold text represents the recognition sequence) with 12 residues between the poly-proline motif and the DOPA residue to ensure sufficient flexibility for DOPA to productively engage the active site. With peptide 2, we found that SH3-588 dramatically enhanced methylation activity compared with WT COMT (Table 1 and Fig. 3a). Specifically, we measured a >1,000-fold increase in the catalytic efficiency (k_{cat}/K_m) from 1,600 to 2.5×10^6 M^{-1} s^{-1} for WT COMT and SH3-588, respectively.

https://doi.org/10.1038/s41589-022-01206-0

Table 1 | Fitted parameters for k_{cat}, K_m and k_{cat}/K_m

Peptide number	Peptide sequence[a]	k_{cat} (s^{-1})				K_m (μM)				k_{cat}/K_m ($M^{-1}s^{-1}\times10^{-4}$)			
		WT COMT	Naive fusion	SH3-588	SH3-003	WT COMT	Naive fusion	SH3-588	SH3-003	WT COMT	Naive fusion	SH3-588	SH3-003
1	MKEARSAEAKEAGRGSGGSKNFLD**Y^{OH}**	0.37 (± 0.007)	0.27 (± 0.01)	0.35 (± 0.02)	0.3 (± 0.01)	42 (± 3)	40 (± 6)	19 (± 3)	30 (± 4)	0.9 (± 0.3)	0.7 (± 0.2)	1.9 (± 0.5)	1.0 (± 0.3)
2	EW**APPLPPR**NRPRRSGGSKET**Y^{OH}**SK	0.213 (± 0.002)	0.073 (± 0.007)	0.37 (± 0.02)	0.212 (± 0.005)	132 (± 4)	0.4 (± 0.2)	0.15 (± 0.04)	0.8 (± 0.1)	0.16 (± 0.05)	17 (± 4)	250 (± 40)	27 (± 5)
3	**Y^{OH}**DLFNKGSGG**APPLPPR**NRPRRS	0.34 (± 0.02)	0.28 (± 0.03)	0.22 (± 0.01)	0.27 (± 0.01)	45 (± 7)	0.2 (± 0.1)	6 (± 2)	0.18 (± 0.04)	0.8 (± 0.3)	140 (± 30)	3.8 (± 0.8)	150 (± 30)
4	**APPLPPR**NRPRRSGKNFLD**Y^{OH}**	0.371 (± 0.005)	0.103 (± 0.007)	0.31 (± 0.02)	0.18 (± 0.01)	24 (± 1)	0.18 (± 0.05)	0.03 (± 0.01)	0.07 (± 0.03)	1.6 (± 0.4)	60 (± 10)	1,200 (± 100)	270 (± 30)
5	**APPAPPR**NRPRRSGKNFLD**Y^{OH}**	0.22 (± 0.01)	0.198 (± 0.009)	0.28 (± 0.01)	0.238 (± 0.009)	15 (± 4)	6 (± 1)	1.5 (± 0.4)	1.0 (± 0.2)	1.5 (± 0.3)	3.5 (± 0.7)	19 (± 3)	23 (± 4)
6	EW**APPLPPR**NRPRRS$(GGS)_3$GKET**Y^{OH}**SK	0.246 (± 0.004)	0.166 (± 0.008)	0.45 (± 0.03)	0.319 (± 0.005)	220 (± 10)	1.7 (± 0.4)	0.18 (± 0.08)	0.75 (± 0.05)	0.11 (± 0.03)	10 (± 2)	250 (± 40)	43 (± 8)
7	EW**APPLPPR**NRPRRS**Y^{OH}**SK	0.3154 (± 0.0006)	0.35 (± 0.03)	0.32 (± 0.01)	0.25 (± 0.01)	183.1 (± 1.0)	400 (± 90)	130 (± 20)	0.9 (± 0.3)	0.17 (± 0.06)	0.09 (± 0.04)	0.25 (± 0.08)	29 (± 5)

[a]Recognition sequences and L-DOPA substrate residues (Y^{OH}) are shown in bold. Error values indicate ±s.e. of the best fit for the traditional Michaelis–Menten parameters.

This improvement was largely afforded by tighter substrate binding, with K_m decreasing from 132 μM with WT COMT to 150 nM with SH3-588. Similarly, SH3-003 reacts with peptide 2 with a k_{cat}/K_m of 2.7×10^5 M^{-1} s^{-1}, which is a 170-fold increase relative to WT COMT, largely due to it having a lower K_m for the peptide. The k_{cat} values for WT COMT, SH3-588 and SH3-003 are similar (ranging from 0.212–0.370 s^{-1}), suggesting that the DOPA side chain engages the active site in a similar manner for all enzymes.

Interestingly, the naive fusion displays a biphasic response in rate with increasing substrate concentration. The first phase fits to a K_m of 400 nM with a k_{cat} of 0.073 s^{-1}; however, at high substrate concentration, the rates match those of WT COMT (Fig. 3a, boxed rate points). The intersection of these two phases occurs near the K_m of WT COMT (132 μM). This suggests the naive fusion recruits substrate to the active site through SH3 domain binding at low substrate concentrations; however, at higher peptide concentrations near the K_m for undirected substrates, substrates bind directly to the active site.

Sensitivity to substrate directionality, affinity and length

Next, we tested the sensitivity of the enzymes to three substrate features: (1) directionality; (2) recognition sequence affinity; and (3) linker length. In the directionality experiments, we placed the DOPA residue N terminal to the poly-proline motif (as opposed to C terminal). Based on our model of SH3-003 and the structure of SH3-588, this should direct the substrate away from the active site, disfavoring catalysis. As anticipated, the catalytic efficiency of SH3-588 dropped considerably with the misdirected peptide Y^{OH}DLFNKGSGG**APPLPPR**NRPRRS (peptide 3) (Fig. 3b), with a k_{cat}/K_m of 3.8×10^4 M^{-1} s^{-1}. For comparison, **APPLPPR**NRPRRSGKNFLDY^{OH} (peptide 4), which directs the DOPA toward the active site, has a k_{cat}/K_m of 1.2×10^7 M^{-1} s^{-1} (an increase of ~300-fold over the misdirected substrate). The lower catalytic efficiency with the misdirected peptide is primarily due to an increase in K_m (6 μM versus 30 nM for peptides 3 and 4, respectively), suggesting that the DOPA residue interacts directly with the active site without simultaneous engagement of the poly-proline motif with the SH3 domain. In contrast, both the naive fusion and SH3-003 displayed similar catalytic efficiencies with peptides 3 and 4. For the naive fusion, this is consistent with a flexible connection between the SH3 domain and COMT and suggests that SH3-003 can adopt conformations different from our structural model. In line with this, the designed interface of SH3-003 contains a series of polar contacts that may not be forming as designed, allowing it to behave like the naive fusion with flexibility about its linker.

Next, we measured reaction rates between SH3-588 and three substrates with varied affinity for the SH3 domain: **APPLPPR**NRPRRSGKNFLDY^{OH} (peptide 4), **APPAPPR**NRPRRSGKNFLDY^{OH} (peptide 5) and MKEARSAEAKEAGRGSGGSKNFLDY^{OH} (peptide 1). Binding and kinetic studies showed a clear correlation between recognition sequence affinity (K_d) and K_m (Fig. 3c and Supplementary Fig. 2). For peptide 4 the K_d was 690 nM and the K_m was 30 nM, while for peptide 5 the K_d was 22 μM and the K_m was 1.5 μM. Meanwhile, peptide 1, which lacked the poly-proline motif, resulted in the weakest K_m of 19 μM. Interestingly, the K_m values for peptides 2, 4 and 5 are lower than the K_d values measured between the recognition sequence and SH3-588. As the K_m is frequently an upper limit on the K_d between a substrate and enzyme, this result indicates that full-length peptide binds to SH3-588 more tightly than the poly-proline sequence alone, suggesting cooperative binding between the peptide and enzyme. Indeed, the difference between K_m and recognition sequence K_d is greater for peptides 4 and 5 (~20-fold) than it is for peptide 2 (~4-fold), reflecting the stronger binding of a C-terminal DOPA (peptides 4 and 5) to COMT compared with an internal DOPA (peptide 2). Alternatively, the K_d for substrate binding can be larger than the K_m in cases where product release is slower than substrate release[28], but this typically involves major conformational changes associated with catalysis that slow product release. Conformational changes of this type have not been reported for COMT.

Despite the widely varied K_m values, k_{cat} remained largely similar with all three substrates and, more broadly, was consistent across the entire substrate scope tested in this study (Table 1). This highlights two features of our system: (1) substrate recruitment predominately improves substrate recognition (K_m); and (2) catalysis, as opposed to product release, is probably the rate-limiting step in the reaction. If the release of the methylated peptide were rate limiting, we would expect the observed k_{cat} to become dramatically slower as affinity for the recognition sequence is increased.

Lastly, we investigated the consequence of variable-length linkers between the recognition sequence and substrate motif (Fig. 3d, diagram). We designed three variable-length substrates: long peptide 6, medium peptide 2 and short peptide 7. Both the long and medium peptides displayed comparable, tight K_m values of 180 and 150 nM, respectively, for SH3-588 (Fig. 3d). Meanwhile, the short peptide had a K_m of 130 μM, which is close to that of WT COMT, indicating that the substrate interacts directly with the enzyme active site, not through the peptide-binding domain. These results suggest that longer linkers, including ones with unique functional sites, may be incorporated without sacrifice to activity. Meanwhile, peptides with short linkers

https://doi.org/10.1038/s41589-022-01206-0

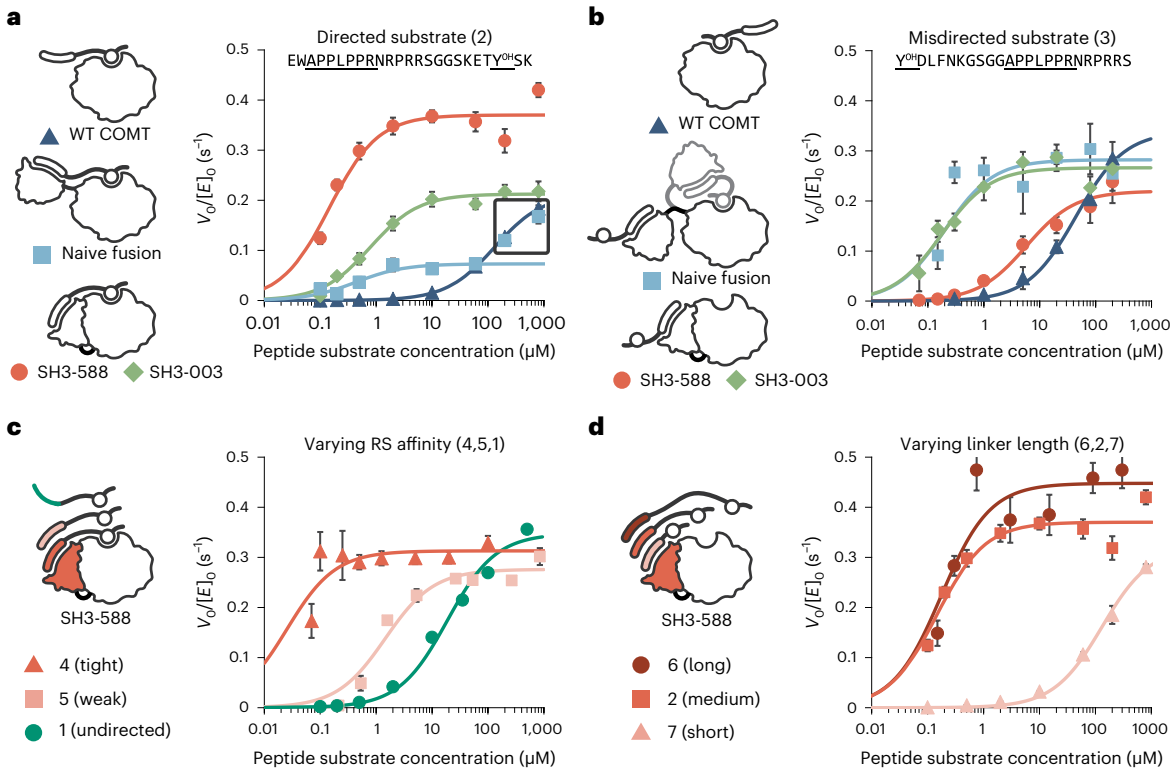


Fig. 3 | Steady-state kinetics analysis of the designed enzymes. a,**b**, Plot of rate versus substrate concentration for WT COMT, naive fusion, SH3-588 and SH3-003 for a directed substrate (peptide 2; **a**) and an N-terminal misdirected substrate (peptide 3; **b**). The boxed area in **a** indicates points that were not fit for the naive fusion and suggests direct binding of substrate to the active site. **c**,**d**, Plot of rate versus substrate concentration for SH3-588 for variable-affinity recognition sequences 4, 5 and 1 (**c**) and variable-length substrates 6, 2 and 7. Rates are presented as points calculated from slopes of best fit from three distinct reactions ($n = 3$). The error bars represent ±s.e. of the best fit centered on the calculated rate (an example is provided in Supplementary Fig. 3a–d).

disallow simultaneous binding of the recognition sequence and the DOPA residue.

We also measured the reactivity of the variable-length peptides with the naive fusion. Due to the enzyme's flexibility, we anticipated that the short peptide might be an effective substrate; however, as with SH3-588, the K_m was large (>100 μM), indicating that the peptide could not simultaneously engage the SH3 domain and the active site. As an additional control, we designed a second naive fusion (naive fusion 2) that has a longer linker (eight residues) between the SH3 domain and COMT. Naive fusion 2 has similar kinetic parameters to the first naive fusion with peptides 1–6, but with the short peptide substrate (7) naive fusion 2 has a lower K_m (600 nM; Supplementary Table 1). This result indicates that the longer linker between the SH3 domain and COMT in naive fusion 2 allows simultaneous engagement of peptide 7 with the active site and the SH3 domain. Notably, naive fusion 2 and SH3-003 have similar catalytic parameters for peptides 1–7 (Supplementary Table 1), providing further evidence that SH3-003 does not form the designed interaction and instead samples a variety of conformations.

SH3-588 preference in substrate competition experiments

Many bioconjugation enzymes are selective, discriminating between multiple potential reactive substrates and modifying a desired target. To probe the substrate selectivity of SH3-588, the naive fusion and WT COMT, we performed competition assays with two peptide substrates at equimolar concentrations: a directed peptide EW**APPLPPR**NRPRRSGGSKETY^{OH}SK (peptide 2); and an undirected peptide containing the same substrate motif but missing the poly-proline motif, MKEWRSPELKEFGRGSGGSGGKETY^{OH}SK (peptide S2; Fig. 4a). Product formation was monitored using liquid chromatography–mass spectrometry.

For all three substrate concentrations that were tested, WT COMT modified the directed and undirected peptides equally (Fig. 4b and Extended Data Fig. 1). As expected, given the lower K_m values for the directed substrate, the naive fusion and SH3-588 preferentially modified the directed peptide at substrate concentrations <10 μM. At 100 μM substrate, near the K_m for undirected peptide, the naive fusion lost selectivity and modified both peptides comparably, displaying similar yields to those of WT COMT. In contrast, SH3-588 retained preference for the target peptide by a factor of >6-fold at high substrate concentration. For all conditions that were studied, SH3-588 modified >90% of the directed peptide, while WT COMT and the naive fusion modified at most 50% of the substrate.

Substrate occlusion explains enzyme specificity

Peptides can interact with the SH3/COMT fusions by two mechanisms: directly through the active site or through recruitment by the peptide-binding domain. The competition experiment demonstrated that at high substrate concentrations the naive fusion shows nonspecific modification, while SH3-588 maintains preference for the directed substrate. A possible explanation for the SH3-588 result is that the directed peptide is simultaneously binding to the SH3 domain and the active site, thus occluding undirected peptides from reacting with the enzyme. Indeed, the K_m for many of the poly-proline directed peptides is tighter than the K_d of the recognition sequence for the SH3 domain. To further explore this concept, we constructed a mathematical model that uses differential equations and our experimental catalytic parameters to simulate competition for directed and undirected substrate at the active site.

The model uses three primary routes to simulate substrate processivity (Extended Data Fig. 2). The directed route models a

https://doi.org/10.1038/s41589-022-01206-0

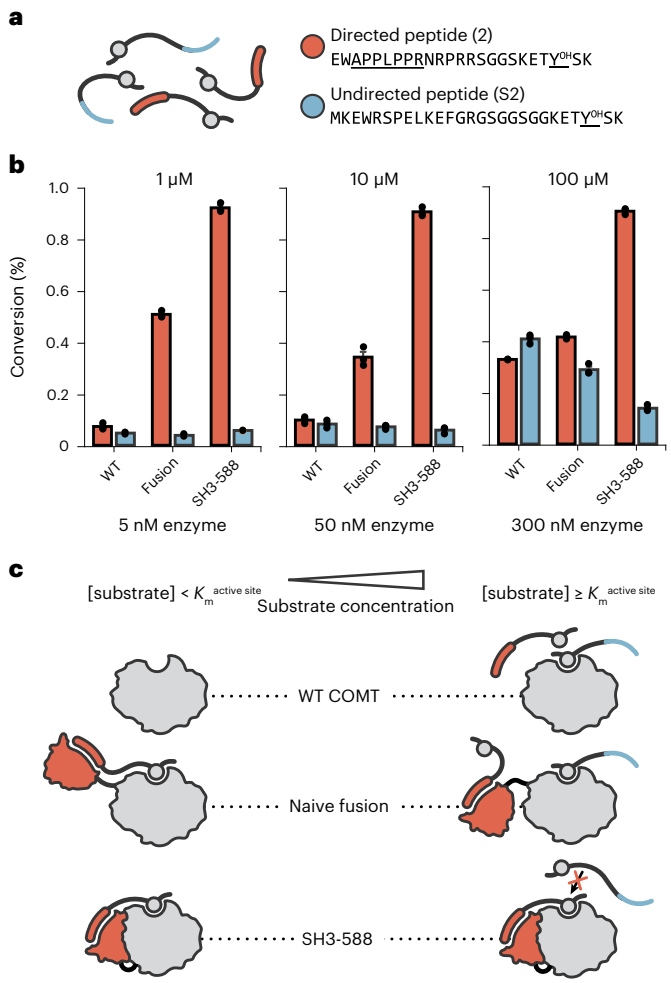


Fig. 4 | Substrate competition experiment. a, Diagram showing competing substrates (peptides 2 and S2) in a peptide reaction mixture. **b**, Percentage conversion of directed and undirected peptides in competition reactions at 1, 10 and 100 μM (2, 20 and 200 μM total peptide) by the indicated enzymes. Raw data points are represented by black dots superimposed on plots. The bars represent mean values ± s.e.m. across three distinct reactions ($n = 3$). The error bars are centered on the mean. **c**, Model of alternative modes of interaction between the enzymes and substrate at varying concentrations of substrate. $K_m^{\text{active site}}$ is the affinity (K_m) of substrate binding directly to the active site.

tethered substrate bound to the SH3 domain. The undirected route models the substrate motif interacting exclusively through the active site. A third pathway involves two peptides interacting with the enzyme: one peptide bound to the SH3 domain and a second, undirected peptide engaged at the active site (Fig. 5a). The relevance of this third pathway depends on how efficiently SH3-bound peptides occlude the active site. In our model, the equilibrium constant K_{open} is derived from the parameter fraction_open, which specifies the fraction of SH3-bound peptides that are in an open conformation with the active site accessible to undirected peptides.

We performed simulations with equimolar concentrations of directed peptide (2) and undirected peptide (S2) and predicted conversions while varying the fraction_open parameter. Because of the identical substrate sequence motifs (KETY^{OH}SK), we used the catalytic parameters of WT COMT as a surrogate for undirected peptide at the SH3-588 active site. With perfect substrate occlusion (fraction_open = 0), we predict very high specificity (~1,000-fold) for the directed peptide over the undirected peptide (Fig. 5b). This high specificity reflects the ratio of the specificity constants used in modeling SH3-588 with directed peptide 2 ($k_{cat}/K_m = 2.5 \times 10^6\ M^{-1}\ s^{-1}$) and undirected peptide S2 ($k_{cat}/K_m = 1.6 \times 10^3\ M^{-1}\ s^{-1}$). With no substrate occlusion (fraction_open = 1), we predict specificity for the directed peptide at low substrate concentrations, but this specificity is lost at high substrate concentrations (Extended Data Fig. 1). The model most closely matches the experimental results when fraction_open is set to 0.1 and 0.3 for SH3-588 and the naive fusion, respectively, suggesting that when recognition sequence is bound the DOPA residue is engaged at the active site ~90 and ~70% of the time, respectively. These fraction_open values indicate that less substrate occlusion occurs with the naive fusion compared with SH3-588, probably due to the added flexibility between the SH3 domain and COMT.

It is also conceivable that binding of the directed peptide is altering binding preferences for SAM versus SAH, which could perturb selectivity for the directed peptide; however, this is unlikely considering the ordered binding mechanism used by COMT: SAM binds first, then Mg^{2+}, followed by the L-DOPA substrate[25]. This ordered binding mechanism is inconsistent with preferential SAM binding as SAH–SAM exchange occurs in the absence of DOPA peptide binding to the active site.

Given the value for fraction_open and the affinity of DOPA for the active site, it is possible to estimate the effective concentration of DOPA when peptide is anchored to the SH3 domain (see Methods). With a binding affinity of 130 μM for DOPA (the K_m of peptide 2 for the WT COMT), an effective concentration of 1.1 mM results in 90% occupancy at the active site. The effective concentration can also be estimated by calculating the volume accessible to a tethered linker (see Methods). With 12 residues between the recognition sequence and DOPA, and assuming half of the space is occupied by enzyme, the volumetric approach predicts an effective concentration of ~2 mM, similar to the value derived from fraction_open.

SH3-588 is highly active with diverse types of substrates

As evidenced by naturally occurring kinases, physical separation of substrate recognition and catalysis provides a powerful strategy for creating systems that can modify a diverse set of substrates. Kinases can often phosphorylate peptides and proteins with varied amino acids located adjacent to the site of modification. To allow for rapid testing of a large set of sequences with SH3-588, we coupled in vitro transcription/translation (IVT) of peptide substrates with an enzymatic activity assay that used mass spectrometry to probe product formation (Extended Data Fig. 3a; see Methods). For almost all substrates tested with SH3-588, the reaction reached completion within 15 min. WT COMT, however, was less active against all substrates and showed a preference for peptides with hydrophobic residues adjacent to DOPA (Extended Data Fig. 3b; IVT peptides 1–5). Similarly, we tested variations in the recognition sequence and found that a minimal sequence, PALPAR, still results in completion by SH3-588 (IVT peptides 7–10), demonstrating plasticity and tunability in the recognition sequence.

Discussion

Our results demonstrate the feasibility and utility of designing substrate scaffolding enzymes from existing protein domains. Our use of computational protein design combines the functions of two domains to generate enhanced enzymatic activity for specially tailored substrates. Studies with our design, SH3-588, show kinetic and selectivity advantage over both WT and naive fusion enzymes. The naive fusion alone improved the efficiency (k_{cat}/K_m), but greater boosts of more than 1,000-fold were seen with our design, SH3-588, which also enhanced selectivity for poly-proline directed peptides. Notably, this selectivity was maintained at high substrate concentrations, where the naive fusion lost selectivity. Our design compares well with the most successful examples of exclusively computationally improved enzymes[29].

The crystal structure of SH3-588 suggests a few general principles for future substrate scaffolding design. First, we chose protein domains

https://doi.org/10.1038/s41589-022-01206-0

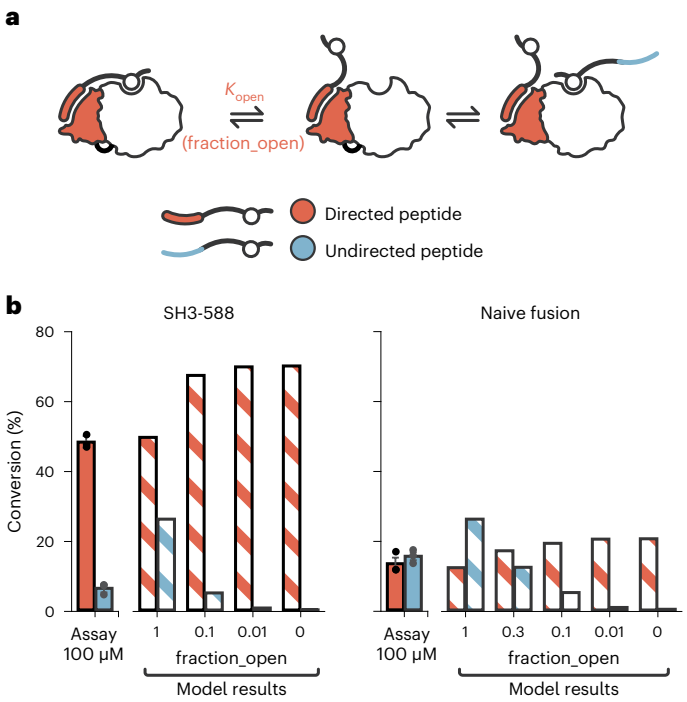


Fig. 5 | Numerical modeling of competing substrates reacting with the engineered enzymes. a, In the numerical model, directed substrates (that is, those containing the poly-proline recognition sequence) engage the SH3 domain and then exist in an equilibrium between an open and closed conformation (equilibrium constant = K_{open}). In the closed form, the DOPA residue is bound to the active site and is available for catalysis. The fraction of bound peptide in the open conformation is referred to as fraction_open, which is directly related to K_{open}. Undirected peptide can only engage the active site when the directed peptide is in the open state or when the enzyme is not bound to directed peptide. Varying fraction_open in the numerical model changes the accessibility of the active site to undirected peptides and changes the predicted specificity during a simulated reaction with directed and undirected peptide (the full mathematical model is presented in Extended Data Fig. 2). **b**, A comparison of simulated results (hashed bars) and experimental results (solid bars) for a competitive reaction in which directed and undirected peptide are both at a concentration of 100 μM. Numerical simulations performed with different values assigned to fraction_open indicate that directed peptide is in the open conformation ~10% of the time when bound to SH3-588 and ~30% of the time when bound to the naive fusion. Raw data points are represented by black dots superimposed on plots. The bars represent mean values ± s.e.m. across three distinct reactions ($n = 3$). The error bars are centered on the mean. The hashed bars represent a single computational result.

that are structurally well characterized and lend themselves well to interface design. These domains have exposed, well-ordered secondary structural elements that readily tolerate mutation and design, emphasizing the need to select stable domains. Second, correct orientation of these designable surfaces immediately puts the N and C termini in proximity for short linker design. Last, this orientation also allows unobstructed access to the active site from the poly-proline binding site. The success of the SH3-588 design is underscored, though, by some of the challenges that we faced in protein engineering. Several of the designs purified as dimers or oligomers, perhaps because of nonspecific interactions between hydrophobic residues that were added to the domain surfaces. Additionally, the SH3-003 design behaves more like a naive fusion, suggesting that the interface did not form as strongly as expected. These results indicate that some challenges remain in coupling interface design and viable folding and expression behavior. Nevertheless, we anticipate that the principles described above can be applied to other enzymes and peptide-binding domains to create new scaffolding enzymes when the domains are well chosen.

Specific placement of the substrate recruitment domain next to the active site resulted in a significant kinetic payoff over both WT and naive fusion enzymes. For example, due to improvements in K_m, both naive fusion and SH3-588 see significant improvements in efficiency, with the interface design having more than tenfold greater efficiency than the naive fusion. These improvements track with what has previously been seen in RiPP systems, which seem to be well suited for overcoming the detriments of low substrate concentrations in a complex cellular environment[8,30]. SH3-588 had additional advantages over the naive fusion in directionality and specificity. While the naive fusion modified substrate residues placed both before and after the recognition sequence, the design was less competent with DOPA residues placed N terminal to the poly-proline motif (Fig. 3b), providing an additional level of control over modification. Also, whereas specificity eroded at higher concentrations with the naive fusion, it was largely sustained with SH3-588. Substrate occlusion appears to play a strong role in this sustained specificity, as our model suggests that a change from ~30% residual accessibility for the naive fusion to only 10% for the SH3-588 design (fraction_open values of 0.3 and 0.1, respectively) can account for the observed differences. These values correlate with effective concentrations of 300 μM and 1.1 mM for the naive fusion and SH3-588, respectively, in agreement with geometric estimates[16,31].

Initially, we hypothesized that scaffolding may perturb turnover rates as k_{cat} can be sensitive to both binding kinetics[15] and orientation of substrate in the active site. Interestingly, we observed similar k_{cat} values (<5-fold differences) across all substrates and enzymes tested in the study, suggesting that the rate of the chemical reaction is an upper bound on the observed rates. Indeed, our mathematical model requires fast on ($k_{on} > 1 \times 10^6\ M^{-1}\ s^{-1}$) and off ($k_{off} > 0.4\ s^{-1}$) peptide binding rates to recapitulate our kinetic results. The SH3 domain is well suited for this application because it binds and releases peptides quickly. The Fyn-SH3 domain binds to poly-proline peptides with off rates faster than 200 s^{-1} and association rate constants > $5 \times 10^7\ M^{-1}\ s^{-1}$ (ref. [32]). Additional lowering of K_m by tightening the affinity of either the poly-proline leader sequence or the core may further improve the efficiency and selectivity by minimizing fraction_open; however, maintaining fast k_{off} (where $k_{off} > k_{cat}$) will be a critical restraint on achieving high catalytic efficiencies.

In conclusion, we have demonstrated that we can design a specific, designed inter-domain interaction that enhances enzyme affinity and specificity. We anticipate that this design approach will be generalizable and provide a structural and kinetic framework allowing the redesign of other existing, promiscuous enzymes for applications in bioconjugation or biosynthesis. The affinity enhancement we see with relatively small peptide substrates should translate advantageously to more complex protein substrates, such as antibodies or enzymes. COMT itself could be useful in labeling applications because it can install SAM analogs with orthogonal functional groups, such as alkynes[26]. Similarly, the expansion of this approach to new catalytic domains effecting different chemistries could allow new labeling strategies or even more extensive modifications, leading to new therapeutic peptide derivatives or natural product analogs (*O*-methyl DOPA itself occurs in pepticinnamin E and related antibiotics). Lastly, the improved affinity and specificity afforded by the designed interaction, especially at low substrate concentrations, should lend itself toward the targeted modification of peptide substrates in much more complex environments. For example, the chemical diversification of combinatorial libraries in peptide and protein display technologies, such as messenger RNA display, or else for synthetic and chemical biology applications in a cellular context. Indeed, naive fusion and recruitment strategies have already been used to probe subcellular interactions[33] and signaling cascades[10,14,34,35]. Clearly, there is much more to be done with these designed substrate scaffolding enzymes.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-022-01206-0.

References

1. Ho, S. H. & Tirrell, D. A. Enzymatic labeling of bacterial proteins for super-resolution imaging in live cells. *ACS Cent. Sci.* **5**, 1911–1919 (2019).
2. Chen, I., Howarth, M., Lin, W. & Ting, A. Y. Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase. *Nat. Methods* **2**, 99–104 (2005).
3. Bhattacharyya, R. P., Reményi, A., Yeh, B. J. & Lim, W. A.Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. *Annu. Rev. Biochem.* **75**, 655–680 (2006).
4. Miller, W. T.Determinants of substrate recognition in nonreceptor tyrosine kinases. *Acc. Chem. Res.* **36**, 393–400 (2003).
5. Pellicena, P., Stowell, K. R. & Miller, W. T. Enhanced phosphorylation of Src family kinase substrates containing SH2 domain binding sites. *J. Biol. Chem.* **273**, 15325–15328 (1998).
6. Scott, M. P. & Miller, W. T. A peptide model system for processive phosphorylation by Src family kinases. *Biochemistry* **39**, 14531–14537 (2000).
7. Qiu, H. & Miller, W. T. Role of the Brk SH3 domain in substrate recognition. *Oncogene* **23**, 2216–2223 (2004).
8. Ortega, M. A. & van der Donk, W. A. New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol.* **23**, 31–44 (2016).
9. Arnison, P. G. et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
10. Park, S.-H., Zarrinpar, A. & Lim, W. A. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science* **299**, 1061–1064 (2003).
11. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
12. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
13. Bolukbasi, M. F. et al. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods* **12**, 1150–1156 (2015).
14. Bashor, C. J., Helman, N. C., Yan, S. & Lim, W. A. Using engineered scaffold interactions to reshape MAP kinase pathway signaling dynamics. *Science* **319**, 1539–1543 (2008).
15. Dyla, M. & Kjaergaard, M. Intrinsically disordered linkers control tethered kinases via effective concentration. *Proc. Natl Acad. Sci. USA* **117**, 21413–21419 (2020).
16. Speltz, E. B. & Zalatan, J. G. The relationship between effective molarity and affinity governs rate enhancements in tethered kinase–substrate reactions. *Biochemistry* **59**, 2182–2193 (2020).
17. Burkhart, B. J., Hudson, G. A., Dunbar, K. L. & Mitchell, D. A.A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nat. Chem. Biol.* **11**, 564–570 (2015).
18. Grove, T. L. et al. Structural insights into thioether bond formation in the biosynthesis of sactipeptides. *J. Am. Chem. Soc.* **139**, 11734–11744 (2017).
19. Leaver-Fay, A. et al. Chapter nineteen—ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol.* **487**, 545–574 (2011).
20. Cao, L. et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science* **370**, 426–431 (2020).
21. Alford, R. F. et al. The Rosetta all-atom energy function for macromolecular modeling and design. *J. Chem. Theory Comput.* **13**, 3031–3048 (2017).
22. Karanicolas, J. et al. A de novo protein binding pair by computational design and directed evolution. *Mol. Cell* **42**, 250–260 (2011).
23. Maguire, J. B. et al. Perturbing the energy landscape for improved packing during computational protein design. *Proteins* **89**, 436–449 (2021).
24. Camara-Artigas, A., Ortiz-Salmeron, E., Andujar-Sánchez, M., Bacarizo, J. & Martin-Garcia, J. M. The role of water molecules in the binding of class I and II peptides to the SH3 domain of the Fyn tyrosine kinase. *Acta Crystallogr. F Struct. Biol. Commun.* **72**, 707–712 (2016).
25. Lotta, T. et al. Kinetics of human soluble and membrane-bound catechol *O*-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry* **34**, 4202–4210 (1995).
26. Struck, A.-W. et al. An enzyme cascade for selective modification of tyrosine residues in structurally diverse peptides and proteins. *J. Am. Chem. Soc.* **138**, 3038–3045 (2016).
27. Plaxco, K. W. et al. The folding kinetics and thermodynamics of the Fyn–SH3 domain. *Biochemistry* **37**, 2529–2537 (1998).
28. Johnson, K. A. New standards for collecting and fitting steady state kinetic data. *Beilstein J. Org. Chem.* **15**, 16–29 (2019).
29. Goldsmith, M. & Tawfik, D. S. Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol.* **47**, 140–150 (2017).
30. Tianero, M. D. et al. Metabolic model for diversity-generating biosynthesis. *Proc. Natl Acad. Sci. USA* **113**, 1772–1777 (2016).
31. Krishnamurthy, V. M., Semetey, V., Bracher, P. J., Shen, N. & Whitesides, G. M. Dependence of effective molarity on linker length for an intramolecular protein–ligand system. *J. Am. Chem. Soc.* **129**, 1312–1320 (2007).
32. Meneses, E. & Mittermaier, A. Electrostatic interactions in the binding pathway of a transient protein complex studied by NMR and isothermal titration calorimetry. *J. Biol. Chem.* **289**, 27911–27923 (2014).
33. Cho, K. F. et al. Split-TurboID enables contact-dependent proximity labeling in cells. *Proc. Natl Acad. Sci. USA* **117**, 12143–12154 (2020).
34. Rivera, V. M. et al. A humanized system for pharmacolog ic control of gene expression. *Nat. Med.* **2**, 1028–1032 (1996).
35. Yazawa, M., Sadaghiani, A. M., Hsueh, B. & Dolmetsch, R. E. Induction of protein–protein interactions in live cells using light. *Nat. Biotechnol.* **27**, 941–945 (2009).

Methods

Computational methods

Selection of the enzyme, peptide-binding domain and designable interfaces. COMT was chosen as a catalytic domain because of its well-understood chemical mechanism[25,26], high-resolution structural characterization (Protein Data Bank (PDB) ID: 5P9V) and surface-exposed active site[36]. Furthermore, peptide substrates can readily be synthesized that incorporate the unnatural amino acid substrate residue L-DOPA[26]. The Fyn-SH3 domain (PDB ID: 4ZNX) has relatively high thermostability and has been used in the past for protein folding studies[27]. Binding studies indicate that Fyn-SH3 binds its cognate APP12 peptide with a K of 260 nM[24]. Designable interface residues were chosen from large patches of residues on well-defined secondary structural elements (that is, α-helices on COMT and β-sheet residues on the SH3 domain). Input PDBs were pre-relaxed with energy minimization in Rosetta and oriented in a single PDB file such that the designable interface of COMT and the SH3 domain faced each other but were separated by nearly 70 Å.

Multistage interface design. The multistage Rosetta script design protocol is summarized in Fig. 1d. Briefly, four separate stages were used to dock (stages 1 and 2) and subsequently design (stages 3 and 4) two domains into a single multidomain enzyme (example script dock_design_sh3.xml in Supplementary Information). In the first stage, rigid body global docking brought the two designable interfaces into proximity. Small rotational and translational perturbations were used to sample the docking landscape. To enrich for a designable population of models, the 100,000 generated models were sorted by the number of designable interface contacts and the top 10,000 were kept. The second step, local docking, perturbed the conformation further by allowing small rotational and translational movements about the initial docked confirmation. Inputs to this step were used for ten independent local docking simulations and the total stage 2 model set was again filtered by number of interface contacts. Docking constraints were used through both the first and second stages to minimize the distances between the termini of the two domains (necessary for short linker design) and to minimize the gap between the active site and peptide-binding site.

Two-sided interface design was then used to stabilize the resulting docked conformations. The third stage consisted of a short round of FastDesign (repeats = 1), which iterates between fast rotamer packing across the interface and energetic minimization of side chain and backbone atoms[23]. This stage largely removed clashing side chains between the two domains to assess models for further design. Out of the 10,000 input models, 2,500 were kept for further design after sorting according to a combined energy calculation that included the total energy of the complex and a normalized interface energy. Stage 4 involved more extensive interface design (FastDesign repeats = 3) and sorting according to the previously mentioned metric, saving the top 200 designs for further analysis. A final sorting step was applied to save the top 10% of output models by interface energy. The entire multistage script was repeated until >1,000 design models were generated. Several different design scripts were run with custom distance constraints to generate multiple populations of docked models.

Forward docking and final model selection. Selected design sequences were further evaluated with protein docking (forward docking) to determine whether the designed pose was strongly preferred over alternative conformations. Designed models underwent random perturbation and docking to attempt to reclose the interface. An example forward docking script is provided in forward_dock.xml in the Supplementary Information. Forward docking produced approximately 20–50 designs from which manual selection was used to pick unique interfaces that varied in interface composition (that is, bulky versus small residues and polar versus hydrophobic residues at the interface periphery) and peptide directionality toward the active site. Ultimately, five unique design models were selected for further linker design and experimental characterization.

Glycine–serine linker design. We fused the newly installed peptide-binding domain to the COMT catalytic domain by linking the C terminus of the SH3 domain to the N terminus of COMT. Linker design was performed using FloppyTail—a Rosetta application that models a linear stretch of amino acids—to extend the C terminus of the SH3-binding domain toward the N terminus of COMT. The minimum number of glycine–serine residues was estimated from this approach and an extra two residues were added to the linker before ordering the genes as described below.

Experimental methods

Cloning and expression of enzymes. Enzyme (COMT native, naive fusion, SH3-588 and SH3-003) expression constructs were either ordered and synthesized from Twist Bioscience (Supplementary Table 4) or manually cloned using forward primer (CTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGTCTCTCCC) and reverse primer (GGATTGGAAGTACAGGTTCTCCCC). These constructs were transformed into chemically competent *Escherichia coli* and grown on LB agar plates (100 mg l^{-1} ampicillin). Colonies were selected and inoculated into 50 ml LB cultures for growth overnight (37 °C with 250 r.p.m. agitation). Approximately 8 ml overgrown culture was introduced into fresh Terrific Broth media with an additional 8 ml 50% glycerol solution. Cells were grown at 37 °C until an optical density measured at a wavelength of 600 nm (OD_{600}) of 1.0–1.2, upon which the culture temperature was dropped to 16 °C and protein production was induced via the introduction of isopropyl β-D-1-thiogalactopyranoside to 600 μM. Cells were grown overnight at 16 °C with 200 r.p.m. agitation for typically 18 to 20 h, after which they were pelleted by centrifugation (3,148g for 20 min) and frozen and stored at −80 or −20 °C.

Cell pellets were solubilized in lysis buffer (50 mM phosphate-buffered saline (PBS; pH 7.4), 300 mM NaCl and 10 mM imidazole) and sonicated for 5 min (5 s on and 5 s off at 70% amplitude) before homogenization via four passes on an EmulsiFlex homogenizer. Cell lysates were clarified by centrifugation (24,500g for 30 min) and filtered (5.0 μM nylon membrane) before loading onto an Ni-NTA gravity flow column for affinity chromatography. The column was washed first with 20 column volumes of high salt buffer (same as lysis with 1 M NaCl) and second with 20 column volumes of low salt buffer (same as lysis buffer) before elution (same as lysis with 400 mM imidazole). Affinity-purified enzymes were further purified by size exclusion chromatography and simultaneously buffer exchanged into storage buffer (50 mM PBS (pH 7.4) and 300 mM NaCl). Enzymes were concentrated to approximately 100 μM and flash frozen with 10% glycerol for long-term storage at −80 °C.

Protein crystallization and data acquisition

For protein crystallization, pMCSG28–COMT–SH3-588 was transformed into Rosetta(DE3) cells on LB agar plates (100 mg l^{-1} ampicillin) and colonies were selected for overnight growth. Cells were grown in 1 l LB media at 37 °C to an OD_{600} of 0.6. Cultures were cooled in an ice bath for 10 min before induction with isopropyl β-D-1-thiogalactopyranoside at a final concentration of 0.4 mM. Cells were allowed to grow overnight at 18 °C before harvesting. The cell pellet was resuspended in suspension buffer (500 mM NaCl, 20 mM Tris (pH 8.0) and 10% glycerol).

Cells were lysed and clarified as above, and clarified lysate was passed through a 5 ml HisTrap HP column (Cytiva Life Sciences) pre-equilibrated with suspension buffer. The column was washed with 40 column volumes of wash buffer (1 M NaCl, 20 mM Tris (pH 8.0) and 30 mM imidazole). SH3-588-His_6 was eluted on an ÄKTAprime plus system (Cytiva Life Sciences) using a linear gradient from 25 to 250 mM imidazole over 40 ml at a flow rate of 2 ml min^{-1}. The purest fractions, as judged by sodium dodecyl sulfate–polyacrylamide gel electrophoresis,

were pooled and cleaved with TEV protease during dialysis (100 mM NaCl, 20 mM Tris (pH 8.0) and 3 µM 2-mercaptoethanol). A subtractive nickel purification was performed to remove uncut material. SH3-588 was concentrated and purified on a 16/60 Superdex 200 column (Cytiva Life Sciences) equilibrated in 100 mM KCl, 20 mM Tris (pH 8.0) and 5 mM dithiothreitol.

COMT–SH3-588 was incubated with 2 mM SAH before laying drops. Crystallization conditions for COMT–SH3-588 were established using sitting-drop sparse matrix screening (using the Crystal Gryphon LCP; Art Robbins Instruments). Crystals were obtained in 0.7 M magnesium formate and 0.1 M Bis-Tris propane (pH 7.0). Initial crystals were optimized in hanging drop trays using 10–15 mg ml^{-1} protein. Crystals resembling grains of rice formed over 1 week at 24 °C. Crystals did not form if the reducing agent was absent or at lower temperatures than 24 °C.

Structure determination

Crystals were transferred to a cryoprotectant consisting of the reservoir solution supplemented with 20% glycerol before vitrification by direct immersion into liquid nitrogen. All crystallographic measurements were collected at Sector 21-ID (LS-CAT, Advanced Photon Source; Argonne National Laboratory). Data were indexed, scaled and integrated using XDS[37] as implemented in autoPROC[38]. Phases were determined by the molecular replacement method as implemented in Phaser[39] using the coordinates of PDB 1M27 and PDB 4PYN as search probes. There was one molecule in the crystallographic asymmetric unit. A high-resolution cut-off of 2.59 Å was used for the data. Although the crystallographic data statistics suggest that the quality data may extend beyond this cut-off, this was not the case as the signal-to-noise ratio dropped rapidly beyond this resolution. For example, at a slightly higher cut-off of 2.5 Å resolution, the merging *R* value (which reports on the agreement between multiple measurements of a given reflection) in the highest resolution shell was 200% and the value of I/σ(I) (a measure of signal to noise) was below 1. Moreover, the quality of the resultant maps did not improve. Hence, the chosen cut-off of 2.59 Å was deemed to be appropriate.

Following one cycle of refinement using REFMAC5 (ref. [40]), the electron density could be observed for regions of the polypeptide that were not included in the search probes, including residues at the designed interface. While densities for the bound ligand and active site were continuous and obvious, other regions of the polypeptide lacked density at the main chain and were not modeled. These include the glycine–serine linker joining the SH3 domain with the COMT domain and a loop region in the COMT domain between Ser256 and Val266. In some instances for the COMT domain, the electron density for side chains was minimal, and the rotamers were modeled based on their common position across six different structures of the isolated COMT domain. The overall B factors are higher than would be expected for other structures of similar resolution, but the regions of the active site and domain interface are clearly defined in OMIT maps. Manual and automated rebuilding using Coot[41] and Buccaneer[42], interspersed with cycles of refinement, resulted in the final model. Crystallographic statistics may be found in Supplementary Table 5. The PDB accession code for SH3-588 is 7UD6.

Peptide synthesis and purification. Peptides were synthesized by the University of North Carolina's High-Throughput Peptide Synthesis and Array Facility via the standard Fmoc solid-phase peptide synthesis technique with C-terminal amide. DOPA residues were incorporated as Fmoc protected unnatural amino acids in peptide substrates and all peptides were purified to >90% purity, as determined by high-performance liquid chromatography (Supplementary Fig. 4). All of the peptides used in this study are listed in Supplementary Table 2.

Fluorescence polarization assays. Fluorescence polarization assays were conducted where TAMRA-APP12 (**S1**) (200 nM) was incubated with varying concentrations of enzyme (5 nM to 50 µM) in phosphate buffer (50 mM; pH 7.5), NaCl (300 mM) and 0.005% Tween 20. Fluorescence polarization was assessed on a Molecular Devices SpectraMax M5 plate reader. Recognition sequence binding affinity was determined by the best fit to a single -site binding quadratic model.

Competitive binding was accomplished by incubating increasing concentrations of competitor peptide (50 nM to 500 µM) with constant fluorophore-labeled peptide (200 nM) and enzyme (2 µM) in the previously described buffer with $MgCl_2$ (3 mM) and SAH (70 µM) to facilitate binding of the DOPA side chain to the active site. The competitive binding data were fit by modeling the competitive fluorescence signal according to Hussain et al.[43]. Briefly, since the K_d of the TAMRA-labeled peptide (peptide **S1**) and its relationship to the monomer concentrations are known, the observed fluorescence polarization signal can be simulated for a given binding affinity for a competitor.

Kinetic assays. Steady-state kinetics was conducted using Promega's MTase-Glo Methyltransferase Assay. Briefly, methylation was initiated via pipette mixing of 2× enzyme (typically a 300 pM final concentration for the designed enzyme or naive fusion or 10 nM for WT COMT; this concentration changed depending on the ease of conversion of the peptide substrate) with a 2× substrate solution (typically varying from 70 nM to 1 mM) in a Tris buffer (20 mM; pH 8.0) with SAM (45 µM), NaCl (300 mM), $MgCl_2$ (3 mM), ethylenediaminetetraacetic acid (1 mM), bovine serum albumin (0.2 mg ml^{-1}), Tween 20 (0.005%) and dithiothreitol (1 mM). Reactions were incubated at 25 °C for the desired reaction period and terminated via the introduction of a reaction aliquot to 0.5% trifluoroacetic acid (to a final concentration of 0.1% trifluoroacetic acid). The sequential assay development steps were completed per Promega assay guidelines following all of the timepoint collections. The luminescence intensity was analyzed using a BMG Labtech CLARIOstar plate reader. Kinetic parameters were determined by fitting initial rates to the Michaelis–Menten equation in MATLAB (R2020a), where data processing and plotting were done with Microsoft Excel.

Substrate competition liquid chromatography–mass spectrometry assays. Competition assays were conducted similarly to the above kinetic assays. Reactions were initiated via the addition of 2× enzyme (final concentrations of 5, 50 and 300 nM) into 2× substrate (final concentrations of 1, 10 and 100 µM, respectively), where target and off-target peptides were in equimolar concentrations (total peptide concentrations of 2, 20 and 200 µM). Reactions were incubated at 25 °C for the indicated reaction period and terminated by the addition of 25 µl sample into 25 µl methanol. Reaction aliquots were cleared by centrifugation (>16,000*g* for 10 min) and stored at −20 °C until they were analyzed. Reaction aliquots were analyzed by liquid chromatography–mass spectrometry electrospray ionization (Kinetex 2.6 µm C18 column and Agilent 6520 Accurate-Mass Q-TOF ESI) where the resulting ratios of integrations of product and substrates were used to calculate product conversion for each enzyme. Ions were extracted with a 50 ppm window width for the substrate and product ions, as shown in Supplementary Table 6.

Matrix-assisted laser desorption/ionization substrate screen. DNA encoding peptide substrates of interest was ordered from Twist Bioscience (Supplementary Table 7). Peptide substrates were generated using the NEB PURExpress cell-free transcription/translation system. DNA was introduced into the custom PURExpress kit (using a custom amino acid mix that includes L-DOPA instead of tyrosine) to directly translate L-DOPA into our peptides. After incubating for 1 h at 37 °C, completed IVT reactions were introduced to a 2× enzyme solution (1 µM enzyme of interest, 2 mM SAM, 3 mM $MgCl_2$, 12.5 mM ascorbic acid, 300 mM NaCl and 50 mM PBS (pH 7.4); final concentrations). Enzyme reactions were

run for 15 min at 30 °C. To prepare samples for matrix-assisted laser desorption/ionization coupled to time-of-flight mass spectrometry (MALDI-TOF-MS; AB SCIEX TOF/TOF 5800 in positive reflector mode), enzyme reactions were zip-tipped (C18 stationary tips; Thermo Fisher Scientific) to purify peptides with 4% acetonitrile and eluted onto a MALDI plate with 80% acetonitrile with half-saturated MALDI matrix (α-cyano-4-hydroxycinnamic acid). Enzyme conversions were calculated as the product integral over the summation of the product and substrate integrals.

Numerical integration model. Numerical integration models were constructed and tested in MATLAB (R2020a) to simulate the conversion of a single substrate via a recruitment mechanism (Extended Data Fig. 2a and substrate_recruitment.m script in Supplementary Information) and to simulate a competition reaction where a directed and undirected substrate compete for the active site (Extended Data Fig. 2b and substrate_competition.m script in Supplementary Information). The directed and undirected routes used binding and kinetic rates derived from experimental values (that is, the directed rate (k_{cat}^{dir}) used k_{cat} from SH3-588, while the undirected rate (k_{cat}^{undir}) used k_{cat} from WT enzyme). In the substrate_recruitment.m model, a substrate can bind to the SH3 domain before engaging the active site or it can bind directly to the active site. For the substrate_competition.m experiments, a series of fraction_open parameters were tested (from 0 to 1) to reproduce the experimental data.

The effective concentration of tethered substrate was calculated using a geometric approach or using the fraction_open parameter derived from numerical modeling of the competition reactions. The geometric approach estimates the volume of a sphere from the calculated radius of an amino acid chain bound to the enzyme[16]. Fraction_open can be used to infer the effective concentration because it is related to the relative population of tethered substrate that is engaging the active site at any time. A fraction_open of 0 would indicate that the DOPA residue is always bound to the active site when the peptide is bound to the SH3 domain and thus would indicate a very high effective concentration. A fraction_open of 1 would indicate that the DOPA residue is not engaging the active site and thus the effective concentration is very low. Equation (1) was used to estimate the effective concentration (that is, [substrate]) from the fitted fraction_open parameter and the K_d of DOPA for the active site. K_d was set equal to the K_m value measured for peptides that did not include a poly-proline motif.

$$1 - \text{fraction open} = \frac{[\text{substrate}]}{K_d + [\text{substrate}]} \quad (1)$$

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Atomic coordinates for the designed SH3–COMT fusion, SH3-588, have been deposited to the Protein Data Bank under the accession number 7UD6. The plasmid for SH3-588 has been deposited to AddGene under plasmid number 185920 (pMCSG28-SH3-588). Source data are provided as Supplementary Data 1.

Code availability

Code used to model kinetic and substrate competition data is available in the Supplementary Information. Rosetta scripts used to design the protein interfaces are also provided in the Supplementary Information and a demo folder containing scripts and input files can also be found in the Supplementary Information.

References

36. Lerner, C. et al. Design of potent and druglike nonphenolic inhibitors for catechol *O*-methyltransferase derived from a fragment screening approach targeting the *S*-adenosyl-L-methionine pocket. *J. Med. Chem.* **59**, 10163–10175 (2016).
37. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).
38. Vonrhein, C. et al. Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 293–302 (2011).
39. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
40. Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
41. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
42. Cowtan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D Biol. Crystallogr.* **62**, 1002–1011 (2006).
43. Hussain, M., Cummins, M. C., Endo-Streeter, S., Sondek, J. & Kuhlman, B. Designer proteins that competitively inhibit $G\alpha_q$ by targeting its effector site. *J. Biol. Chem.* **297**, 101348 (2021).

Acknowledgements

We thank members of the laboratories of A.A.B. and S.L. Campbell for sharing advice and equipment. In addition, we thank D. Thieker, J. Maguire and A. Leaver-Fay for invaluable advice in creating the design protocol for the multidomain enzyme. This work was supported by NIH grant R35GM131923 (to B.K.) and is based on work supported in part by a discovery grant from the Eshelman Institute for Innovation and National Science Foundation under grant number 2204094 (to A.A.B.).

Author contributions

R.P. designed SH3-588 and conducted all of the experimental enzymatic assays. C.O. and S.K.N. crystalized and determined the structure of the designed protein and contributed to the Methods. R.P. and B.K. wrote and tested the custom MATLAB scripts for substrate occlusion. R.P., A.A.B. and B.K. wrote the manuscript and contributed intellectually to the design of the system.

Competing interests

The authors declare no competing interests.

Additional information

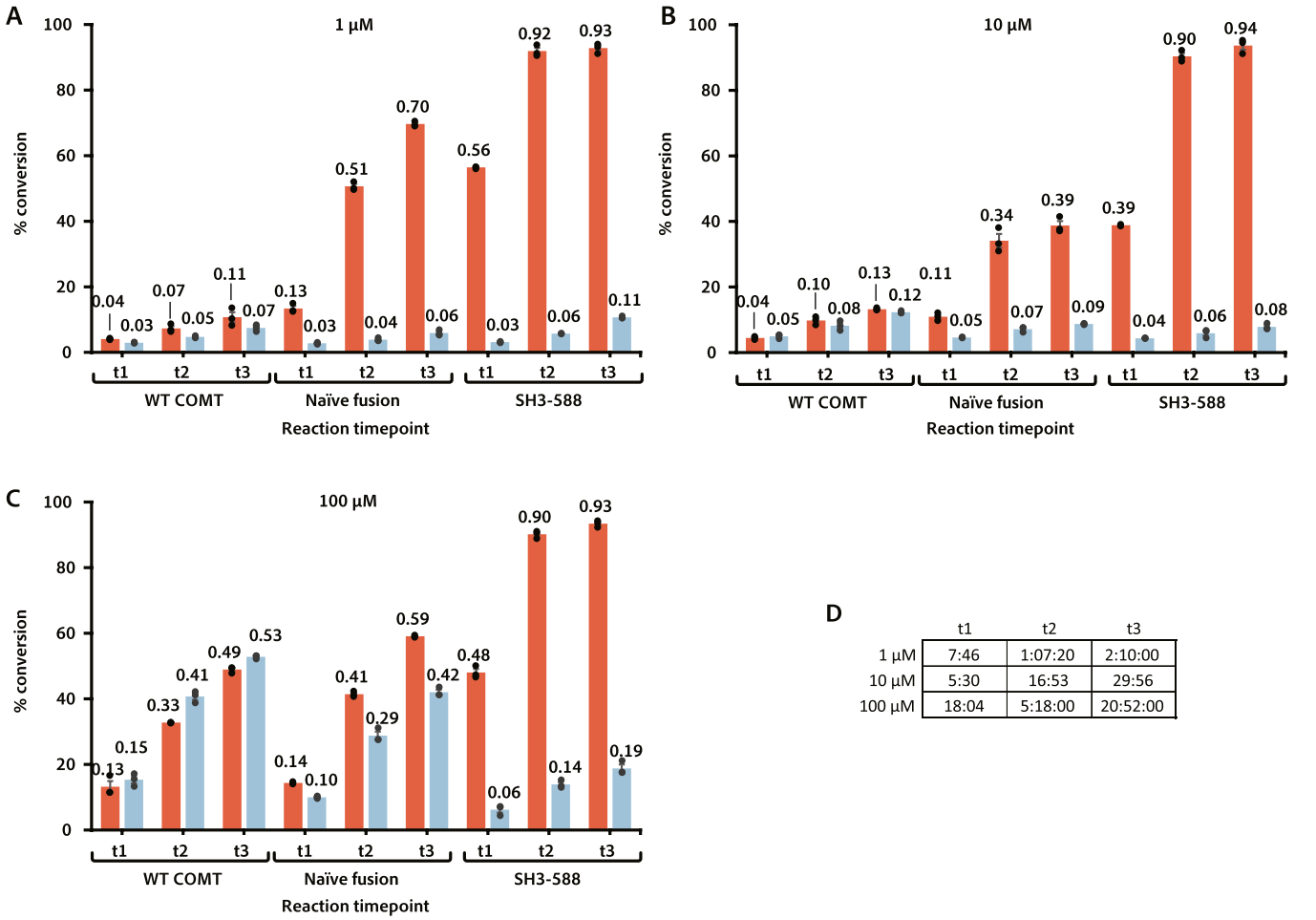
Extended data is available for this paper at https://doi.org/10.1038/s41589-022-01206-0.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-022-01206-0.

Correspondence and requests for materials should be addressed to Albert A. Bowers or Brian Kuhlman.

Peer review information *Nature Chemical Biology* thanks Andrew Buller and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

https://doi.org/10.1038/s41589-022-01206-0

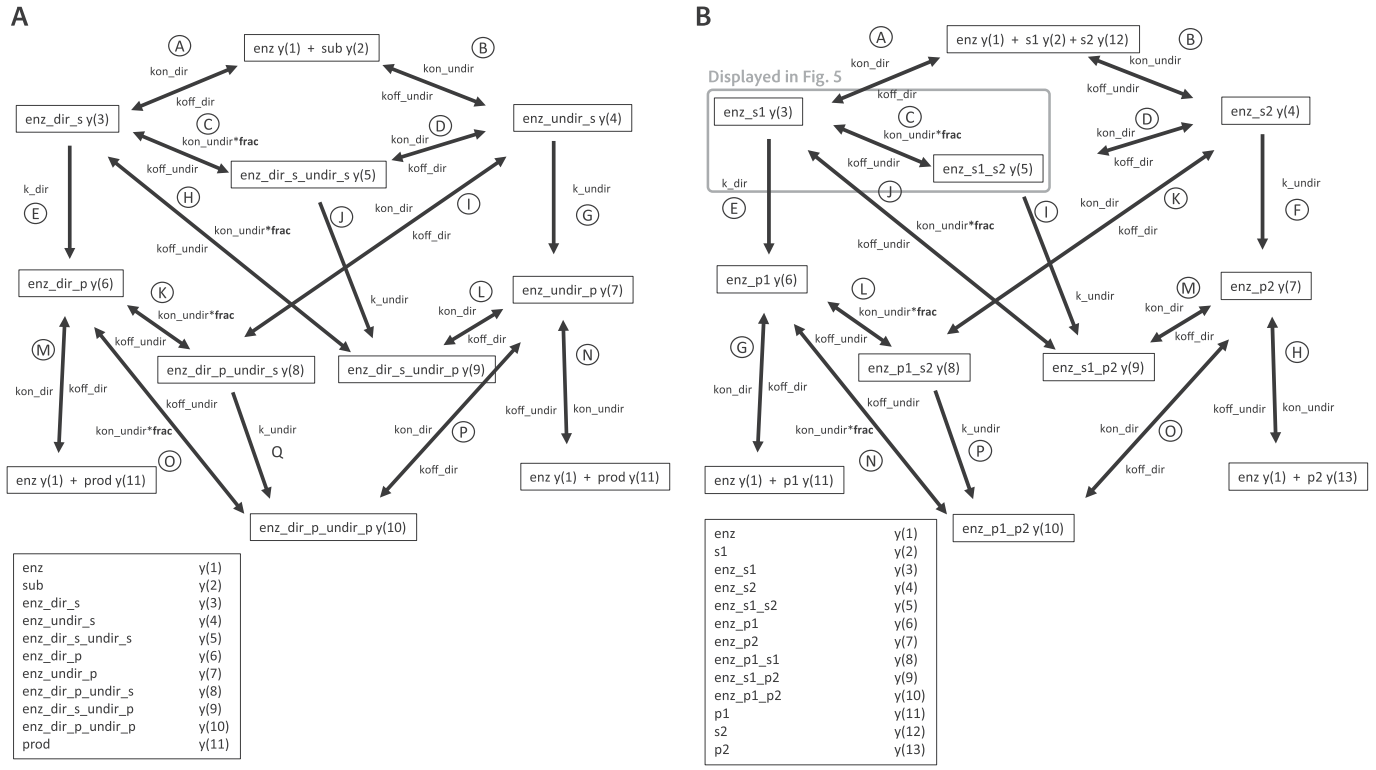


D

	t1	t2	t3
1 µM	7:46	1:07:20	2:10:00
10 µM	5:30	16:53	29:56
100 µM	18:04	5:18:00	20:52:00

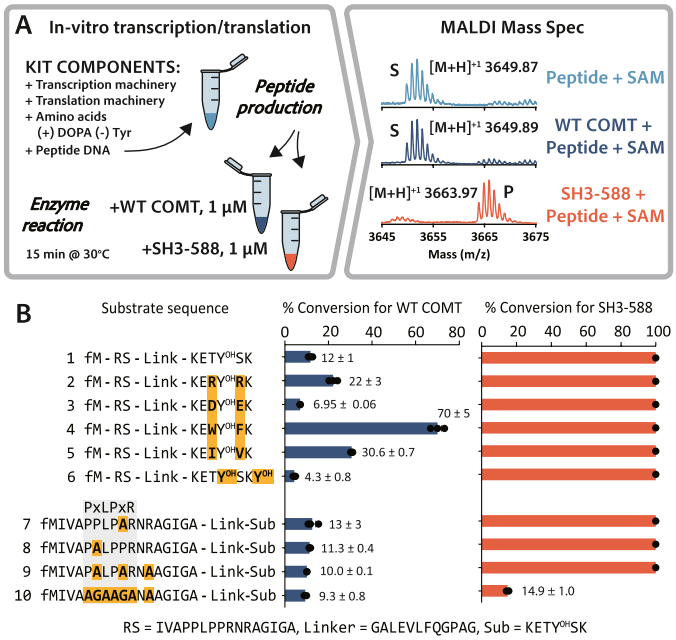
Extended Data Fig. 1 | Expanded set of timepoints for substrate competition assay. Expanded set of timepoints for substrate competition assay for target peptide (red) and off-target (light blue) at 1 (**a**), 10 (**b**), and 100 (**c**) µM peptide substrate (2, 20, and 200 µM total peptide respectively). Raw data points are represented by black dots superimposed on plots. Data bars present mean values +/− SEM across three distinct reactions (n = 3, frequently hidden under data points). Error bars are centered on the mean. (**d**) Exact timepoints collected for 1, 10, and 100 µM. Timepoint 2 is displayed in Fig. 4 in main text and t1 at 100 µM is used in competition reaction mathematical modeling.

https://doi.org/10.1038/s41589-022-01206-0



Extended Data Fig. 2 | Full computational model diagram. Full computational model diagram. (**a**) Single, directed substrate simulation, where substrate can bind to the active site and peptide binding site. (**b**) Multi-substrate (directed and undirected competition) reaction simulation diagram.

https://doi.org/10.1038/s41589-022-01206-0



Extended Data Fig. 3 | Substrate diversity assay. Substrate diversity assay. (**a**) Diagram of assay procedure. Briefly, kit components were combined with genes encoding peptides. After IVT incubation, the crude mixture was split, respective enzyme added, and reactions were run at 30 °C for 15 minutes. Samples were prepped and run for MALDI-TOF-MS analysis. (**b**) Table of IVT peptide substrates tested and their corresponding conversions. SH3-588 largely reached completion for most peptides tested (marked with 100%); remaining substrate peaks were hardly above noise (IVT peptides 1–9). IVT peptides 1–10 were analyzed by MALDI-TOF-MS. Error values indicate +/− SEM across three distinct replicates (n = 3) for IVT peptides 1–8 and two distinct replicates (n = 2) for IVT peptides 9 and 10 centered on the mean.

nature research

Corresponding author(s): Brian Kuhlman, Albert A. Bowers

Last updated by author(s): Sep 13, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For structure design, the Rosetta Scripts interface of the Rosetta macromolecular modeling suite was used to design the structures. Scripts are provided at the end of the supplemental information. For numerical integration modeling, custom MATALB (R2020a) scripts were created and are available in the supplemental information. Structure data was indexed, scaled, and integrated using XDS (Build20210205) and phases were determined by the molecular replacement method as implemented in Phaser (2.8.3). Refinement and rebuilding were accomplished with REFMAC5 (5.5), COOT (0.9), and Buccaneer (1.16.9). Full process and crystallographic statistics can be found in methods and supplementary information. Kinetic data was processed and fit using Microsoft Excel and MATLAB, respectively.
Data analysis	For kinetic analyses, data processing and statistical analyses, standard deviations and errors of fit were determined by Excel or MATLAB using their built-in functions. Crystallographic statistics were obtained after building and refining in COOT.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Atomic coordinates for the designed SH3-COMT fusion, SH3-588, have been deposited to the Protein Data Bank under the accession number 7UD6. The plasmid for SH3-588 has been deposited to AddGene under plasmid number 185920 (pMCSG28-SH3-588). Source data is provided in the supplementary information.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Assays were done in triplicate, which is standard in the field for determining kinetic parameters (Johnson, Beilstein J. Org. Chem., 2019).
Data exclusions	Typically, 12 points were fit to determine initial reaction rate; however, outlier data points were excluded from kinetics experiments in cases where negative concentrations were reported from assay wells or where reported concentrations diverged by three-fold or greater. Mass spectrometry experiments had no data exclusions.
Replication	All kinetic experiments were done in triplicate. Additionally, reaction progress was monitored via MALDI-MS and ESI-TOF-MS. Replication of kinetic data across multiple biological replicates and control designs (see Naïve fusions 1 and 2) were found to be reproducible except for peptide 7 as described.
Randomization	Randomization is not relevant to this study because no subjects were tested. Enzymes compared are tested under the same conditions.
Blinding	Blinding is not relevant to this study because no subjects were tested. Enzymes compared are tested under the same conditions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging