bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Massively parallel, computationally-guided design of a pro-enzyme

Brahm J. Yachnin[1,2], Laura R. Azouz[1,2], Ralph E. White III[3], Conceição A. S. A. Minetti[1], David P. Remeta[1], Victor M. Tan[2,6], Justin M. Drake[3,4,5], Sagar D. Khare[1,2,*]

[1]Department of Chemistry & Chemical Biology and the [2]Institute of Quantitative Biomedicine, Rutgers, The State University of New Jersey

[3]Department of Pharmacology, [4]Urology, and [5]Member, Masonic Cancer Center, University of Minnesota

[6]Department of Pharmacology, Robert Wood Johnson Medical School, Piscataway, New Jersey, USA.

[*]To whom correspondence should be addressed:

Center for Integrative Proteomics Research, Room 208M

Rutgers, The State University of New Jersey

174 Frelinghuysen Road

Piscataway, NJ 08854-8076

(848) 445-5143

khare@chem.rutgers.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

The ability to confine the activity of a protein to a specific microenvironment by design would have broad-ranging applications, such as enabling cell type-specific therapeutic action by enzymes while avoiding off-target effects. While many natural enzymes are synthesized as pro-enzymes that are activated by proteolysis, it has been a difficult challenge to effectively re-design any chosen enzyme to be similarly stimulus-responsive. Here, we develop a massively parallel computational design, screening, and next-generation sequencing-based approach for pro-enzyme design. As a model system, we employ CPG2, a clinically approved enzyme that has applications in both the treatment of cancer and controlling methotrexate toxicity. Our designed pro-enzymes are inhibited up to fivefold, and their activity is restored following incubation with specific proteases expressed by various tumor cell types. Pro-enzymes exhibit significantly lower activity relative to the fully activated enzyme when evaluated in cell culture. Structural and thermodynamic characterization of CPG2 pro-enzymes provides insights into the mechanisms associated with pro-domain inhibition. The described approach is general and should enable the design of a variety of pro-proteins for precise spatial regulation of their functions.

Introduction

Autoinhibited forms of enzymes, commonly designated as pro-enzymes or zymogens, are used ubiquitously for the precise regulation of key biological processes.[1] For example, the activities of blood clotting enzymes, digestive proteases, regulatory proteases, apoptotic proteases, and complement system proteins can be triggered by relevant environmental stimuli. Autoinhibiton is typically achieved by extending the active protein with an inhibitory pro-domain, which acts to

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

inhibit enzymatic activity until it is cleaved by an activating protease that recognizes a specific cleavage sequence in the linker connecting the pro-domain to the active protein. To enable precise control of the tissue microenvironment where an exogenously administered therapeutic enzyme is active, a general approach for redesigning enzymes into activatable, zymogen variants is valuable. There have been successful efforts to design pro-enzymes,[2–5] but these have relied on unique structural features of the enzymes under consideration: either the ability to build an inhibiting linker between the natural N- and C-termini of the enzyme,[2,3,5] or the ability to obtain an "alternate frame folding"-based design.[4] Furthermore, much like most natural pro-enzymes, these designed zymogen forms have been limited to those enzymes that target macromolecular substrates,[2–5] rather than small molecule substrates that are more difficult to sterically impede and thus represent a significantly greater design challenge. The development of a generalizable, structure-guided approach for zymogen design, particularly one that could be applied to enzymes that target small molecule substrates, would allow spatial control over a broad range of enzymes.

In order to achieve an effective means of localizing the activity of any chosen enzyme to a target microenvironment, we envisioned the addition of an inhibitory pro-domain that would be removable by specific proteases which are markers of the target tissue (Figure 1a). For example, matrix metalloproteases (MMPs) are highly expressed in tumors, and would facilitate targeting of enzymatic activity to a tumor site. Tumor-specific MMP activity has been employed previously to achieve site-selective activation of pro-drugs[6,7] and imaging contrast agents.[8] We selected the FDA-approved enzyme, carboxypeptidase G2 (CPG2), as the target for zymogenization. CPG2 is a well-studied and functionally versatile enzyme with a variety of therapeutically beneficial activities. Specifically, CPG2 efficiently inactivates the anti-cancer

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

drug, methotrexate to reduce systemic toxicity (Figure 1b),[9–12] for which it has received FDA approval. Moreover, CPG2 activates non-toxic nitrogen mustard prodrugs, such as CMDA (*N*-{4-[(2-chloroethyl)(2-mesyloxyethyl)amino]benzoyl}-L-glutamic acid), to their active, toxic forms in directed enzyme prodrug therapy (DEPT) (Figure 1c).[13,14] Extensive prodrug development,[15–17] animal studies,[18–20] and clinical trials[13,14,21] of the CPG2 DEPT system have been performed, corroborating its potential therapeutic value. Nonetheless, challenges in limiting the off-target activity of the circulating enzyme have precluded its widespread clinical adoption, and speak to the utility of a tissue-activatable form of DEPT enzymes. While we previously explored such a strategy by designing a light-activatable enzyme,[22] a protocol that exploits the natural properties of tumors to express elevated levels of highly specific proteases would be more desirable from a clinical standpoint. Designed activatable forms of CPG2 which respond to tissue-specific proteases could therefore be broadly useful therapeutic reagents for both delivering and ablating toxicity in a tissue-specific manner.

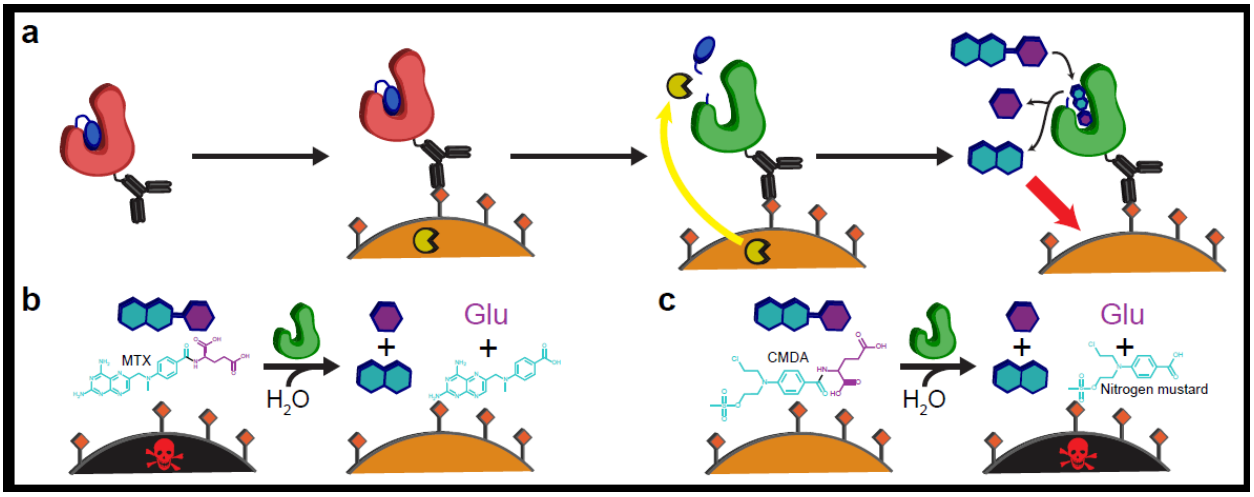

Figure 1. Utility of a protease-activatable form of CPG2. (a) Mechanism of activation of antibody-linked Pro-CPG2. The antibody (black) directs the inactive enzyme (red) to a protease-expressing cell (cell is orange, protease is yellow, antibody's targets are orange diamonds). The protease (yellow) cleaves the linker that connects the pro-domain (blue) to the enzyme, activating it (green). The enzyme can then process small molecules (teal and purple) which can have an effect (red arrow) on the protease-expressing cell. **(b)** Activated CPG2 can cleave and detoxify the drug, methotrexate, rescuing the protease expressing cells (black to orange). **(c)** Activated CPG2 can cleave and active the nitrogen mustard prodrug CMDA, killing the protease expressing cells (orange to black).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Here, we describe a computationally-guided protein design approach to develop protease-activatable zymogens, and employ this strategy to obtain pro-enzyme forms of CPG2. Combining next-generation sequencing (NGS) and a computational design approach implemented in the Rosetta software suite, we have implemented a massively parallel enzyme design strategy to generate and evaluate thousands of diverse computationally designed pro-CPG2 variants in parallel. Using this approach, we designed the first zymogen forms of enzymes that target small molecule substrates. We obtained pro-CPG2 enzymes that are inhibited up to 80% of the uninhibited variant. Their activity can be recovered *via* proteolytic cleavage of the pro-domain-enzyme linker. By varying the protease cleavage site in the linker, two structurally diverse proteases can be used to activate CPG2, thereby highlighting the modular nature of our design strategy. We apply small-angle X-ray scattering (SAXS) to uncover the structural basis of inhibition, demonstrating for the first time that key domain movements are critical for CPG2 catalytic activity. The designed Pro-CPG2s provide significant sensitization of the PC-3 human prostate cancer cell line for methotrexate. These CPG2 variants are promising leads for further therapeutic development and the resultant methodology should enable facile zymogenization of other enzymes.

Results

Overview of design strategy. In natural zymogens, pro-domains are fused by a protease-cleavable linker to one of the protein termini. Inspection of the crystal structure,[23] reveals that in CPG2, the N- and C-termini are ~33 and 34 Å distant from the active site, respectively. We previously designed active circular permutations of CPG2 ($CPG2_{CP\text{-}N89}$) with termini proximate to the active site, creating the possibility of attaching a terminal pro-domain that can access and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

effectively inhibit the active site.[24] The CPG2$_{CP\text{-}N89}$ variant (in which the polypeptide chain is re-organized such that the N-terminal residue is wild-type residue 89) served as the starting point for our design efforts. As CPG2 is a zinc metallo-enzyme, we reasoned that using a zinc-coordinating sidechain, such as glutamate, to guide pro-domain placement would result in enhanced auto-inhibition by rendering the catalytic site inaccessible while the pro-domain is bound (Figure 2).

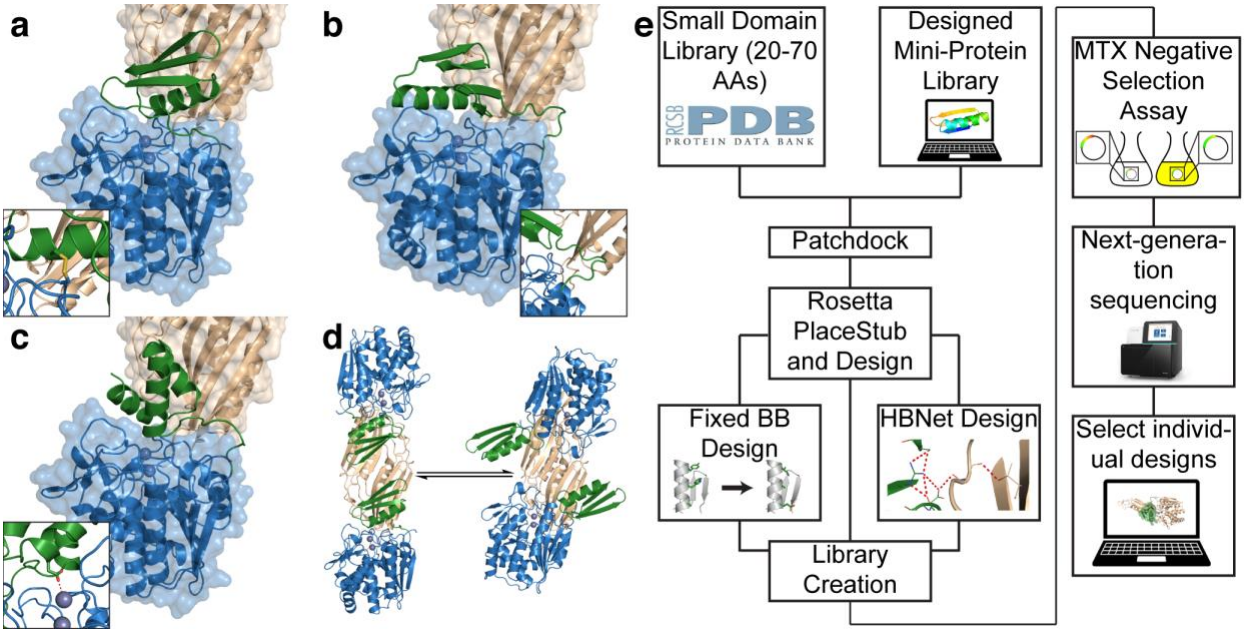

Figure 2: Massively-parallel design of Pro-CPG2. (abc) The three Pro-CPG2 designs selected for detailed study: **(a)** Pro-CPG2-1, **(b)** Pro-CPG2-2, **(c)** Pro-CPG2-3. The pro-domain (green), catalytic domain (blue), and dimerization domain (wheat) are highlighted. **Inset:** Important design characteristics are shown, including **(a)** close contacts between the pro-domain and the enzyme, sufficient to allow design of a disulfide staples, **(b)** proximity between the C-terminus of the circular permutant and the N-terminus of the pro-domain, and **(c)** an interaction between a glutamate residue on the pro-domain and one of the active site zinc ions (grey spheres). **(d)** The conformational equilibrium required for CPG2 activity, highlighting the role of the pro-domain (green). **(e)** The design strategy is depicted as a flow chart.

Helical pro-domain design. As an initial proof-of-concept for the auto-inhibition of CPG2, we designed series of pro-enzymes featuring a short helical pro-domain. We docked a 12- or 17-residue helix into the active site of the CPG2$_{CP\text{-}N89}$ circular permutant[24] while allowing design on both sides of the interface. The resulting poses were then subjected to loop closure, fusing the

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

helix to either terminus using a linker containing the TEV protease cleavage sequence. We identified multiple designs in which a Zn^{2+}-coordinating residue could be added such that it would interact with the catalytic dinuclear zinc site. Eight designs were selected for activity characterization with either the original TEV cleavage sequence or the MMP-2 cleavage sequence, and two designs (Figure S1b, G and H) were experimentally determined to be modestly auto-inhibited (Figure S1). Incubation of pro-enzyme with the relevant protease resulted in an enzyme that regained activity similar to the uninhibited scaffold. These findings validate the design of a true pro-enzyme that is both inhibited and re-activatable, although the degree of inhibition was modest. To obtain insight into the structural basis for the observed inhibition, we solved the crystal structure of one of the designed pro-enzymes (Table S1). We determined that this variant maintains the overall tertiary structure of the enzyme (Figure S2), indicating that the circular permutation does not perturb the structure or function of CPG2, as expected from our previous results.[24] We could not locate density for the appended helical pro-domain in the structure, a finding that is consistent with the weak inhibition observed for this design.

Massively-parallel pro-CPG2 computational design. Building on the modest proof-of-principle success achieved with the helical pro-CPG2 designs, we pursued a massively-parallel design approach to develop more strongly inhibiting pro-domains. Three characteristics present in these initial designs were used to guide our approach. First, the interaction between the catalytic zinc ion and a glutamate residue on the pro-domain was considered a “hotspot” interaction. This hotspot residue was incorporated in designs by directing pro-domain placement using a library of free glutamate conformations (inverse rotamers) that coordinate the zinc ion;

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

backbone coordinates of these glutamate conformations were used to position the pro-domain. Second, only C-terminal pro-domains were considered, as these proved to be more effective in our initial designs (Figure S1b). Finally, the K177A mutation was retained in all designs given its modest impact on activity while increasing degree of helical pro-domain inhibition. The latter arises presumably as a consequence of decreased steric restrictions upon pro-domain binding. A set of 3157 pro-domains was generated from small domains curated from the PDB and from a large set of previously reported *de novo* designed "miniproteins."[25] The 4444 designs produced using this protocol exhibited a wide range of folds and sequences originating from both the PDB small domain set and the *de novo* designed miniprotein set (Figure 2e).

To minimally perturb the activity of the post-cleavage form of the enzyme, we used a one-sided design approach, in which only the pro-domain sequence could be altered while the sequence of the CPG2 "scaffold" region remained fixed. As the interaction surface for the pro-domain is not a natural protein-protein interaction site, we anticipated that the interface might include fewer hydrophobic residues than is typical of natural and designed protein-protein interactions. Therefore, we also designed explicit hydrogen bond networks between the pro-domain and the enzyme to maximize the effectiveness of non-hydrophobic interaction surfaces.

We used a number of criteria (detailed description in Supplementary Results S1, Figure S3, and Table S2) to select ~7500 models from a pool of ~15,000 original designs for screening using a high-throughput assay. As controls, we included four helical pro-domain designs (one ineffective and three moderately effective autoinhibitory designs) that allow comparison of the relative

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

activities determined in our high-throughput assay to helical pro-domains with known inhibitory effects.

Massively parallel screening of the pro-CPG2 design pool. We previously reported that periplasmically-expressed CPG2 is able to protect *E. coli* CAG12184 cells from methotrexate toxicity.[24,26–28] We employed this screen to discriminate between inhibiting and non-inhibiting pro-CPG2 designs from the set of 7469 computationally-generated designs. In summary, cells expressing designs with poorly inhibiting pro-domains (i.e. full CPG2 activity) detoxify methotrexate effectively and survive, while cells expressing designs with strongly inhibiting pro-domains (i.e. low CPG2 activity) are unable to detoxify methotrexate and would be killed. By profiling the composition of designs selected at various methotrexate concentrations (0-100 μM) using NGS, we generated a relative ranking of pro-domain efficacy. By including both moderately effective and ineffective control pro-domains identified from our original helical pro-CPG2 designs, the NGS results were analyzed to identify designs that are expected to be at least as strongly inhibited as the helical pro-CPG2s (Figure 3 and Table S2). Between 400,000 and 1.3 million NGS reads were obtained for each individual test culture corresponding to methotrexate concentrations of 0-100 μM, suggesting adequate coverage of the 7500 unique sequences contained in the library.

To confirm the ability of this high throughput assay to effectively discriminate between inhibiting and non-inhibiting designs, we selected twelve designs predicted to be inhibiting and four designs predicted to be non-inhibiting (Figure 3 and Figure S4). These designs encompassed varied pro-domain sequences and structures, indicating that the selection is not

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

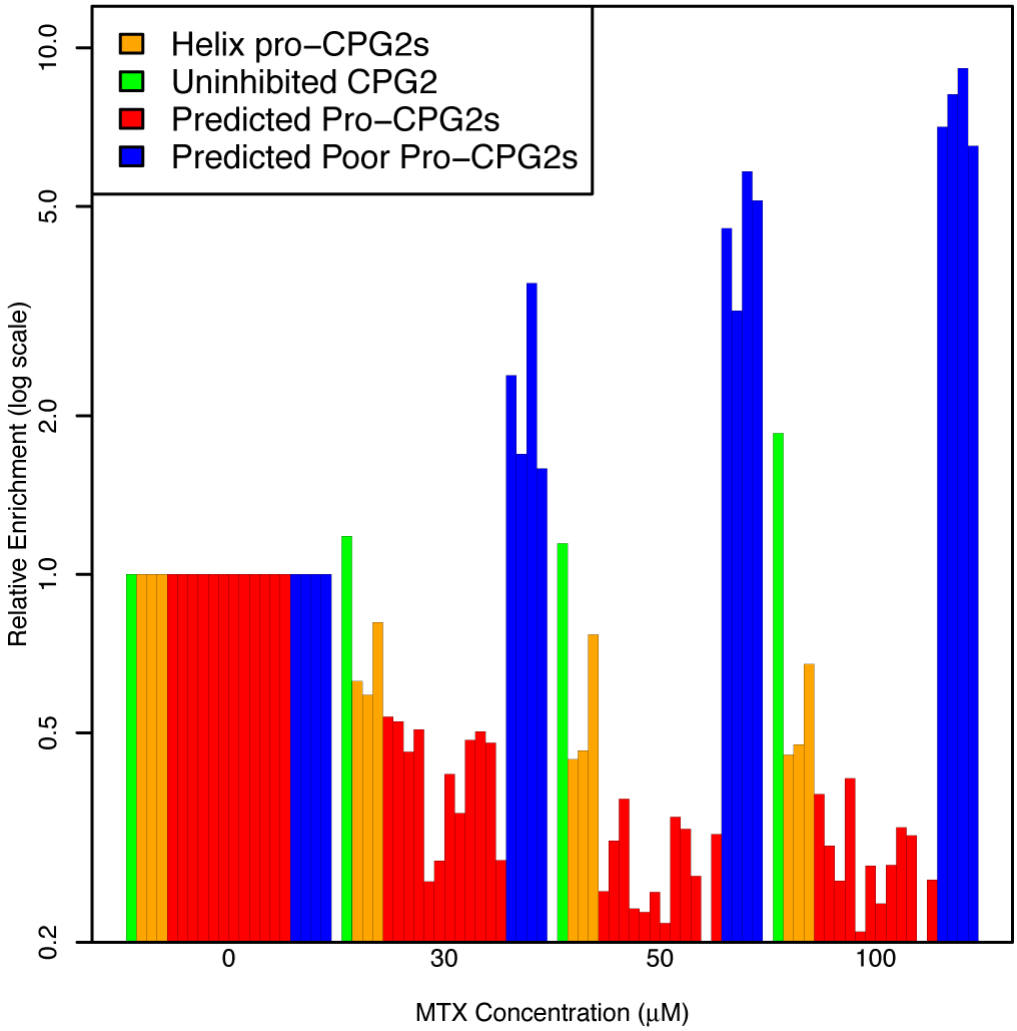

Figure 3: Next-generation sequencing results from the massively parallel selection assay. The logarithm of the relative enrichment ratio is shown at increasing methotrexate concentrations. The helical designs that are known not to inhibit (green) or to mildly inhibit (orange) CPG2CP-N89 are shown alongside the twelve designs predicted to be Pro-CPG2s (red) and predicted not to be Pro-CPG2s (blue).

biased towards a particular fold. Each of these sixteen designs were purified as individual constructs alongside the most inhibited helical pro-CPG2 design and the $CPG2_{CP\text{-}N89}$-K177A construct. All twelve of the predicted inhibiting pro-CPG2 designs were more strongly inhibited than the helical pro-CPG2 design, while the four predicted non-inhibiting pro-CPG2 designs exhibited activities resembling $CPG2_{CP\text{-}N89}$-K177A (Figure S4). These results confirm that our screening system accurately discriminates between inhibiting and non-inhibiting designs, using the helical pro-CPG2 controls as benchmarks. Of the twelve design constructs, we identified

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

three designs for further characterization, which are referred to hereafter as Pro-CPG2-1, Pro-CPG2-2, and Pro-CPG2-3.

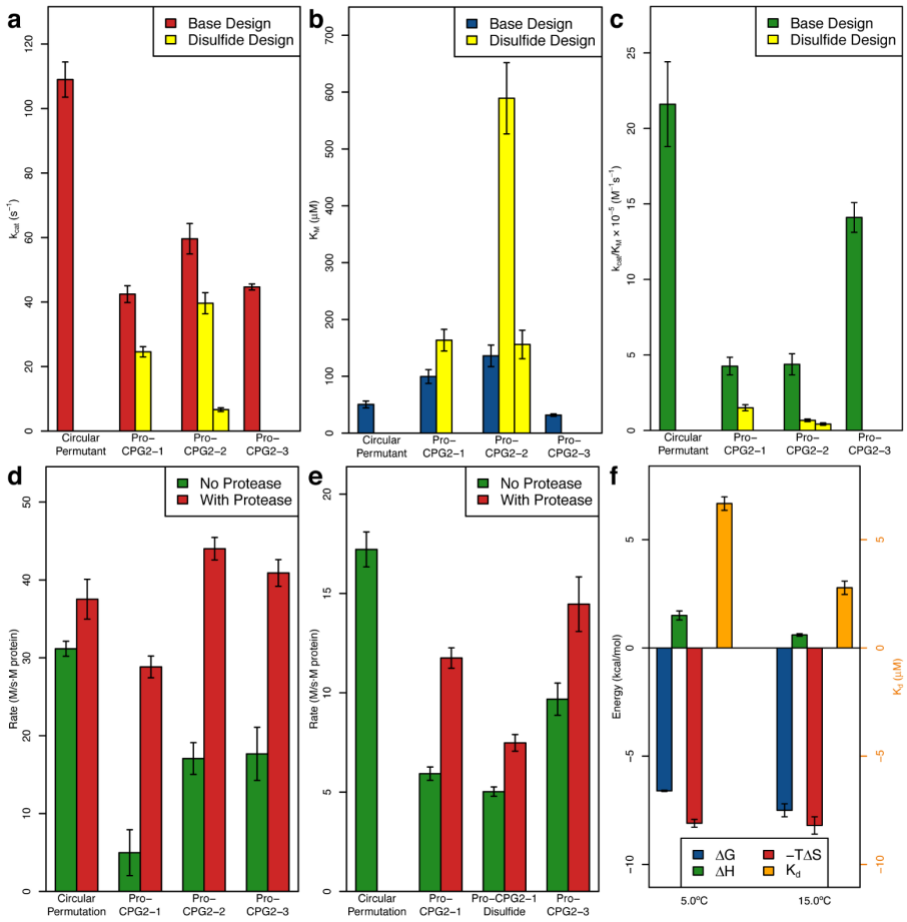

Figure 4: Activity and thermodynamic data for the Pro-CPG2 designs and their disulfide variants. (**a-c**) The Michaelis-Menten kinetics (a, K_M; b, k_{cat}; c, k_{cat}/K_M) for the three base Pro-CPG2 designs are shown. In yellow, the disulfide variants are grouped with the base design they are derived from. See also Table S3. (**d-e**) The recovery of activity following incubation with (**d**) MMP-2 and (**e**) TEV protease is shown. (**f**) Temperature-dependent thermodynamic binding signatures (left axis) and affinity (right axis) characterizing the association of $CPG2_{CP\text{-}N89}$-K177A and Pro-Domain-1 deduced from ITC measurements. See also Table S4.

The Michaelis-Menten kinetic parameters of Pro-CPG2-1, -2, and -3 were determined and compared to the fully active $CPG2_{CP\text{-}N89}$-K177A (Figure 4abc and Table S3). Significantly, there is an approximate 50% reduction in k_{cat} for all three designs, indicating a dramatic decrease in the enzyme's ability to complete its catalytic cycle. In addition, Pro-CPG2-1 and -2 elicited a two to threefold increase in K_M, suggesting that the enzyme's effective substrate binding affinity

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

is reduced. Collectively, these three designs exhibit up to an 80% decrease in catalytic efficiency, k_{cat}/K_M.

Recovery of activity using specific proteases. In order to function as switches, zymogens must be re-activated via proteolytic processing by a specific protease. To probe the activatability of the pro-CPG2 designs, we generated interdomain linker variants containing the canonical cleavage sequences of two dissimilar proteases: the 3C-type TEV cysteine protease[29] and zinc-dependent matrix metalloproteinase-2 (MMP-2).[30] The activity of all Pro-CPG2 designs was nearly fully recovered following incubation with the appropriate protease (Figure 4d,e). As demonstrated by mass spectroscopy, both residual uncleaved pro-enzyme and retention of the bound but cleaved pro-domain may be responsible for the incomplete recovery of protease activity in some cases (Figure S5). As two unrelated proteases are able to cleave and re-activate the corresponding CPG2 pro-enzyme, it is likely that a broad range of other site-selective proteases could be used to re-activate pro-CPG2 by including the appropriate cleavage site in the inter-domain linker.

Characterization of the designed pro-domains. As an initial, computational characterization of the designed pro-enzymes, we performed a series of simulations to assess their quality (Supplementary Results S2 and Figure S6). *Ab initio* forward folding was able to successfully "refold" the designs both as isolated proteins and in the context of CPG2, suggesting that the appended domains are well-folded in the designed conformations.

To further investigate the pro-domains as independent proteins, the pro-domains of Pro-CPG2-1, -2, and -3 were expressed individually and their secondary structure features analyzed using

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

circular dichroism (CD) spectroscopy. All three pro-domains exhibit CD spectra characteristic of well-folded proteins that are stable at physiologically-relevant temperatures. Inspection of thermal melting profiles reveals relatively broad non-cooperative unfolding transitions that do not plateau even when heated to 100 ºC, thereby precluding accurate determination of a discrete melting temperature (Figure S7). The high stability of the pro-domain constructs further indicates that the pro-domains interact with and inhibit CPG2 as well-folded domains in the pro-enzyme context.

Next, we assessed the binding affinity of the pro-domain constructs for the CPG2 catalytic domain "in *trans*". We conducted isothermal titration calorimetry (ITC) experiments in which Pro-Domain-1 is titrated successively into $CPG2_{CP\text{-}N89}$-K177A. Acquisition of a series of binding profiles at varying Pro-Domain (150-500 µM**)** and enzyme (15-50 µM) concentrations yielded a consistent set of thermodynamic parameters, thereby ensuring that aggregation or precipitation (at higher concentrations) or dissociation (at lower concentrations) do not appreciably impact binding energetics. Binding isotherms acquired at 5.0ºC reveal that the interaction heats are relatively modest and endothermic. Moreover, analysis of these association profiles indicates that formation of the $CPG2_{CP\text{-}N89}$-Pro-Domain-1 complex is characterized by binding affinities on the order of 10^5 M^{-1} (i.e., K_d ~ 3-6 µM) over the temperature range studied. Inspection of the thermodynamic binding parameters (Figure 4f and Table S4) acquired at 5.0ºC reveals that Pro-Domain-enzyme association is entropy-driven, as manifested by the relative contributions of ΔH (1.5 kcal·mol^{-1}) and TΔS (8.1 kcal·mol^{-1}) to the Gibbs Free Energy (ΔG = -6.6 kcal·mol^{-1}). Significantly, the interaction is accompanied by a negative heat capacity (ΔC_p ~ -0.098 kcal·mol^{-1}·deg^{-1}) as reflected in the temperature-dependence of binding enthalpies

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

measured at 5 and 15ºC. Extrapolation of the thermodynamic binding profiles to 25ºC reveals that the enthalpic term approaches zero at this temperature and is presumably slightly exothermic under selection, assay, and physiological conditions. The observed temperature-dependence is a characteristic hallmark of hydrophobic interactions[31–33] that presumably reflects the desolvation and burial of hydrophobic surfaces upon pro-domain binding.

Examination of the binding interface in our model reveals several non-polar contact regions: 766 $Å^2$ of the 1117 $Å^2$ buried solvent accessible surface area calculated using Rosetta is hydrophobic. The discrete surface burial, particularly in terms of hydrophobic residues, is generally consistent with a modest heat capacity change estimated for the interaction and may provide further support for our design model. Given that the binding interface includes a zinc ion-coordinating glutamate from the pro-domain, it is entirely plausible that metal ion-bound water molecules may be dislodged upon association. Therefore, the zinc-glutamate coordination might contribute additional unfavorable enthalpic contributions as previously reported for protein-metal interactions,[34] resulting in an overall net entropy-driven association. An evaluation of the interdependence of the CPG2 pro-domain binding sites is included in the Supplementary Results (S3).

An open-closed equilibrium associated with CPG2 activity revealed by small-angle X-ray scattering. Based on the crystal structure of CPG2 reported here (Figure S2), as well as a previously published structure of CPG2[23] and closely-related homologues such as mouse carnosinase 2 (mCN2),[35] Sapep,[36] PepV,[37] peptidase T,[38] aminoacylase-1,[39] and *N*-succinyl-L,L-diaminopimelic acid desuccinylase (DapE),[40] we hypothesized that CPG2 exists in two distinct

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

conformations: an "open" conformation responsible for substrate binding, as observed in the CPG2 crystal structures, and a "closed" conformation optimized for catalytic turnover, as seen in many of the homologous structures (Figure 2d). In order to test our hypothesis regarding conformational changes in the catalytic cycle of CPG2 and the Pro-CPG2 designs, we utilized small-angle X-ray scattering (SAXS) experiments (Figure 5). Anticipating that the addition of excess substrate (which is converted entirely to product by the enzyme prior to data collection) would push the conformational equilibrium to the "closed" conformation, we acquired SAXS data in the presence and absence of methotrexate to monitor changes in the accessibility of these two states. A model of the "closed" state was generated by individually superimposing the CPG2 catalytic and dimerization domains onto the structure of mCN2, and using generalized kinematic loop closure[41] to re-build the interdomain linkers.

Analysis of the SAXS data (Figure 5a, Table S5) for $CPG2_{CP\text{-}N89}$-K177A in the absence of substrate resulted in a particle with a radius of gyration (R_g) of 38.9 Å. While neither the "open" crystal structure nor the "closed" model produced a reasonable fit of the data alone ($\chi^2 > 25$), we observed a good fit ($\chi^2 = 2.2$) for a two-state model using FoXS.[42] The two states ("open" and "closed") each contribute roughly 50% of the total scattering (Figure 5b). In contrast, when methotrexate is added to the sample, one obtains a smaller R_g of 36.0 Å which is consistent with the more compact "closed" model (Figure 5a). The "closed" model produced a reasonable fit to the data using CRYSOL[43] ($\chi^2 = 4.9$), though a moderately better fit is obtained with the two-state FoXS model ($\chi^2 = 2.3$), with 75% and 25 % contributions from the "closed" conformation and "open" conformation, respectively (Figure 5b). Thus, these data support our hypothesis regarding the involvement of domain motions in the catalytic cycle of CPG2 and are consistent

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with the mechanisms of homologous enzymes,[35–40] namely that addition of the substrate triggers a shift in the conformational equilibrium towards the "closed" conformation of CPG2.

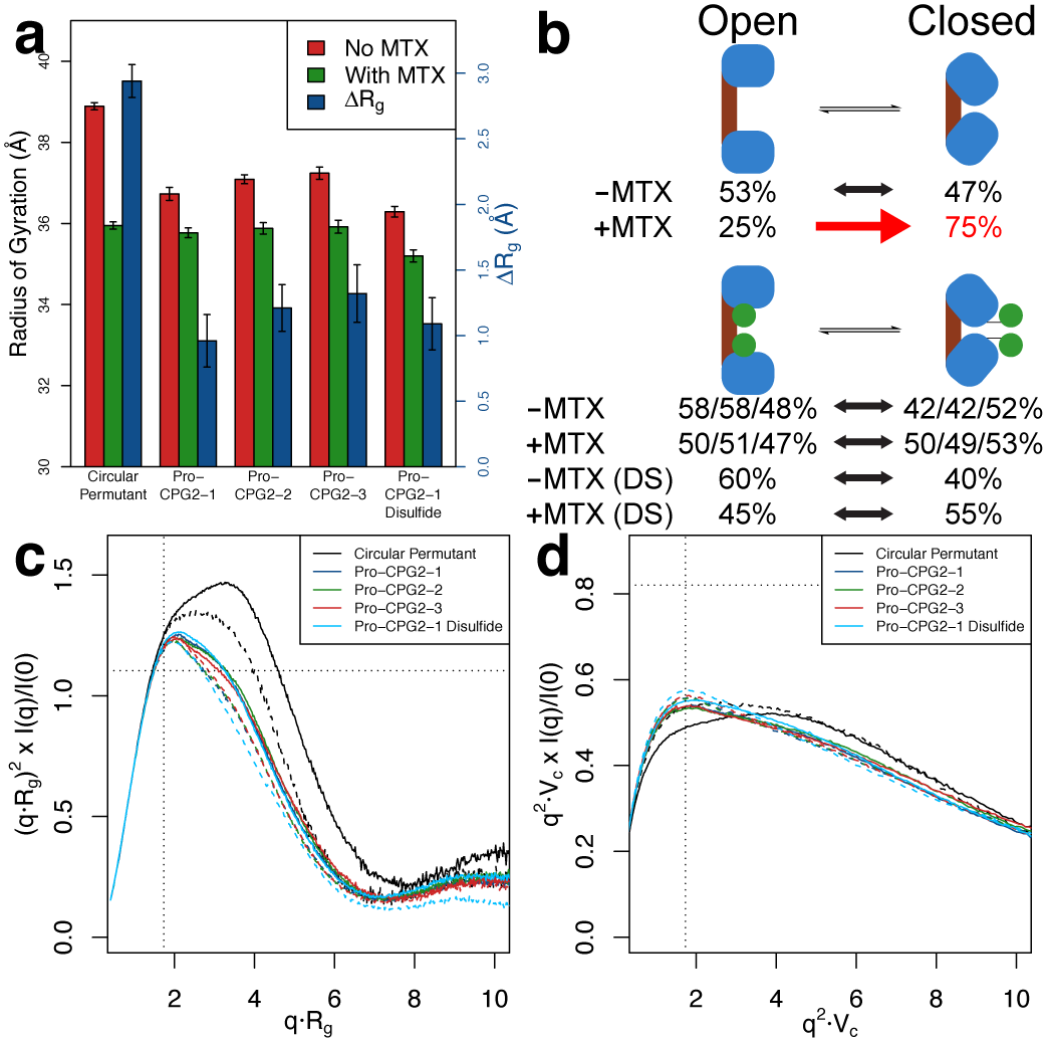

Figure 5: Small angle X-ray scattering (SAXS) analysis for CPG2CP N89 K177A and the Pro-CPG2 designs. (a) The radius of gyration of the enzymes in the absence (red) and presence (green) of methotrexate (left axis) and difference in radius of gyration (blue, right axis) are shown. **(b)** Results from fitting the two-state model (open and closed) to the SAXS data using FoXS are shown, with "-MTX" and "+MTX" indicating the absence or presence of methotrexate. The disulfide variant of Pro-CPG2-1 is indicated by "DS." The quality of the fits is shown in Figure S8. **(c-d)** The **(c)** radius of gyration-based and **(d)** volume of correlation-based dimensionless Kratky plots are shown. Samples with solid lines have no methotrexate present, while samples with dashed lines have methotrexate present. The horizontal and vertical dotted lines indicate cross at the peak position of a compact, spherical particle. In the Rg-based plot, if the peak is shifted up and to the right, this indicates that the particle is more elongated, and if the curve becomes a hyperbolic plateau, this is indicative of a disordered particle. In the Vc-based plot, a downwards shift indicates an increase in surface-to-volume ratio relative to that of a sphere.

A similar analysis was conducted on Pro-CPG2-1, -2, and -3 (Figure 5 and Table S5). In all three cases, the R_gs of samples containing substrate are reduced when compared to the samples

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

in the absence of substrate; however, the magnitudes of these changes are smaller than those for $CPG2_{CP\text{-}N89}$-K177A, primarily owing to a smaller observed R_g for samples without substrate (Figure 5a). When two-state models of the Pro-CPG2s are used to fit the scattering data, the conformational equilibrium is approximately 50%/50% open/closed, as observed for $CPG2_{CP\text{-}N89}$-K177A, ranging between 48-59% in the "open" conformation. In contrast, the conformational equilibrium does not shift with the addition of substrate, remaining at roughly 50%/50% levels for the pro-enzymes (Figure 5b). While the lack of a shift towards the "closed" conformation is consistent regardless of the Pro-enzyme model used, the specific percentages are dependent on the quality of models used. These data are consistent with a scenario in which the "open" state is preferentially stabilized by the designed inhibitory domain. The population of the "open" state in solution also provides an experimental observable to quantify improvements in inhibition obtained by the next generation of designed pro-domains.

To corroborate our finding and investigate the impact of bias introduced by the quality of both the $CPG2_{CP\text{-}N89}$-K177A and pro-enzyme models, we utilized model-free analyses of the SAXS data, including dimensionless Kratky[44] and Porod exponent analyses (Supplementary Results S4 and Figure 5). These analyses confirm that in the absence of substrate, $CPG2_{CP\text{-}N89}$-K177A behaves as a well-folded but flexible particle. While addition of the substrate renders this particle less flexible, the pro-enzymes' flexibility is even further reduced, with the concentration of mass occurring closer to the particle's center of mass. These results are therefore consistent with our initial findings that the pro-enzymes exhibit reduced conformational flexibility as compared to the circular permutant, with a greater population in the inactive "open" state.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Effects of disulfide stapling on Pro-CPG2. To validate our structural models of Pro-CPG2 and assess whether further reduction of Pro-CPG2 activity is possible by stabilization of the "open" state using a covalent linkage between CPG2 and the pro-domain, we generated disulfide-stapled variants of Pro-CPG2-1 and -2 with the goal of restricting the pro-domain to its designed "open"-state conformation. For each pro-enzyme, we used a Rosetta-based approach to identify locations where a disulfide bond spanning the CPG2-pro-domain interface is feasible. We found that one designed disulfide-containing variant of Pro-CPG2-1 and two variants of Pro-CPG2-2 exhibited substantially reduced activity (Figure 4abc) relative to the CPG2 cysteine variant. SAXS experiments on the Pro-CPG2-1-Disulfide variant reveal a further moderate decrease in flexibility as deduced from the dimensionless Kratky plots (Supplementary Results S4 and Figure 5cd). We conclude that a disulfide bridge is effectively created between the pro-domain and scaffold, thereby validating our structural models. Moreover, "pinning down" the pro-domain using a secondary covalent linkage assists in strengthening the effectiveness of the pro-domain, presumably by decreasing its conformational freedom.

Pro-CPG2-1-Disulfide has reduced ability to detoxify methotrexate in cancer cell culture. To assess the function of the Pro-CPG2-1-Disulfide variant in a more complex biological context, we treated an aggressive prostate cancer cell line, PC3, with Pro-CPG2-1-Disulfide and $CPG2_{CP\text{-}N89}$-K177A. The PC3 cell line was derived from a bone metastasis of a grade IV prostate cancer patient and is commonly used to assess treatment responses, including chemotherapies, making it ideal to evaluate the capability of designed pro-enzymes to degrade the chemotherapeutic methotrexate.[45] PC3 cells were incubated briefly with Pro-CPG2-1-Disulfide or $CPG2_{CP\text{-}N89}$-K177A and then subsequently treated with 50 nM of methotrexate over

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

six days. Cells treated with the Pro-CPG2-1-Disulfide variant displayed significantly lower cell viability at enzyme concentrations of 15-62 pM when compared to CPG2$_{CP-N89}$-K177A (Figure 6). These data highlight that in the context of a cancer cell line, the Pro-CPG2-1-Disulfide variant exhibits a lower ability to detoxify methotrexate as compared to CPG2$_{CP-N89}$-K177A.

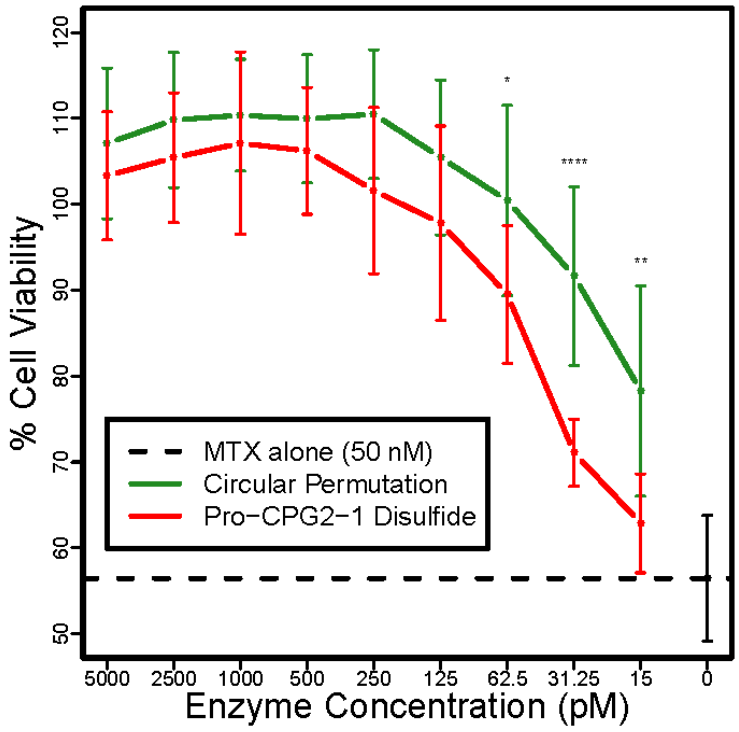

Figure 6: Pro-CPG2-1-Disulfide has reduced ability to cleave and detoxify methotrexate (MTX) compared to CPG2CP-N89-K177A. PC3 cells were treated with MTX at 50 nM, and then treated with either CPG2CP-N89-K177A or Pro-CPG2-1-Disulfide. Enzyme concentrations between 250-5,000 pM showed 100% viability. Enzyme concentrations between 15-62pM showed a significant difference (unpaired t-test, $p<0.05$) of viability in PC3 cells between the two enzyme treatment groups, with cells treated with Pro-CPG2-1-Disulfide exhibiting a lower survival rate than cells treated with active CPG2. The dashed line indicates the cell viability in the presence of MTX alone (no enzyme treatment). Error bars indicate the standard deviation. *= $p< 0.05$, **= $p< 0.01$, ****= $p < 0.0001$

Discussion

We have developed a massively parallel enzyme design and screening approach to engineer a pro-enzyme form of CPG2, an FDA-approved drug and one of the best-studied DEPT enzymes. This approach has allowed us to achieve a reversible decrease in activity of up to 80% without further design optimization, which can be reversed by two structurally and functionally distinct

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

proteases, including one that is typically overexpressed by cancer cells. Introduction of a disulfide bond between the pro-domain and the enzyme results in 98% inhibition compared to the base $CPG2_{CP\text{-}N89}$-K177A variant.

The design challenge for developing a pro-CPG2 enzyme is substantial as the pro-domain binding site is not known to naturally form protein-protein interactions. Moreover, considering its proximity to the active site, any substitution made on the catalytic domain may affect enzyme activity, which must be deconvoluted from changes in activity induced by the pro-domain. Therefore, we employed a one-sided design strategy against a relatively polar surface (with two Zn^{2+} ions) on the catalytic domain, focusing on creating hydrogen bond networks for stabilizing an appropriately placed metal co-ordinating glutamate sidechain. We reasoned that a massively parallel design and screening approach with a diversity of pro-domain structures would maximize opportunities to identify strongly inhibiting pro-domains given the known challenges in accurately designing polar interfaces and metal chelation sites.[46–52] It is noteworthy that the pro-domain structures and sequences that we identified are highly varied, suggesting that the design approach developed here is able to identify multiple, distinct solutions to the problem, rather than converging on a single solution, as is often the case with “winner-takes-all” directed evolution strategies.

At least two putative mechanisms may underlie the inhibition of activity observed in Pro-CPG2 designs. In one case, if the pro-domain blocks the active site, it may impede binding of the substrate. Alternatively, CPG2 is expected to undergo a “closing” conformational change during catalysis and the pro-domain may act as a “door stop,” preventing the enzyme from closing after

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

substrate binding, thereby impeding catalysis. Our structural and kinetic characterization provides insight into the mechanisms of inhibition. If the pro-domain is simply acting to block the active site, rather than directly interfering with the catalytic machinery of the enzyme, one might reasonably expect that this would primarily manifest as a decrease in affinity between the enzyme and substrate, as evidenced by an increase in K_M. As the catalytic residues and conformational changes are unimpeded in this model, the k_{cat} should be minimally affected by this mechanism of inhibition. Conversely, if the substrate is free to bind the enzyme, but the enzyme's ability to pass through its catalytic cycle is impeded, the K_M should be largely unaffected, while the k_{cat} should be highly reduced. In two of the three most inhibited designs, we observe a decrease in both K_M and k_{cat}, suggesting that both mechanisms are at play. The third design is characterized solely by a decrease in k_{cat}, suggesting that the primary mechanism of inhibition occurs *via* disruption of the catalytic turnover.

The SAXS data provide further insight into the molecular mechanism of inhibition. We present the first direct evidence that CPG2 undergoes open-closed conformational changes, although similar observations were made for homologous enzymes.[35–40] Our data highlight the conformational selection underlying CPG2 activity: $CPG2_{CP\text{-}N89}$-K177A, upon addition of substrate, shifts towards the more compact, "closed" conformation. In the Pro-CPG2 designs, this conformational shift does not seem to occur, suggesting that this change is impeded by presence of the pro-domain. This trend is further validated by analysis of the SAXS flexibility data; while various SAXS-derived metrics suggest that $CPG2_{CP\text{-}N89}$-K177A is more compact and less flexible in the presence of methotrexate than in its absence, the Pro-CPG2 samples display

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

minimal corresponding variation. Furthermore, these data are generally consistent with a more compact Pro-CPG2 structure.

Analysis of the calorimetric profiles characterizing in *trans* binding of the pro-domain of Pro-CPG2-1 revealed dissociation constants in the low micromolar range. In a therapeutic context, this magnitude of affinity likely represents an ideal balance between sufficiently tight binding to effectively inhibit the enzyme when covalently linked to the pro-domain, yet weak enough to be released from the binding site following proteolytic processing. In order to evaluate whether pro-enzymes with similar inhibitory characteristics have potential utility for *in vivo* or clinical applications, we examined whether the pro-enzyme form of CPG2 has diminished ability to protect PC-3 cancer cells from methotrexate toxicity. Our results demonstrate that the decreased activity of the Pro-CPG2-1-Disulfide variant results in a statistically significant reduction in the enzyme's ability to detoxify the drug over a multi-day experiment in tumor cells. This activity differential suggests that further investigation of these pro-enzymes in more cell line and animal models as well as in clinically valuable contexts is warranted. For example, the use of a tumor activated pro-enzyme in antibody DEPT (ADEPT) chemotherapeutic strategies would further mitigate the serious side-effects associated with conventional chemotherapy beyond what is already achieved using ADEPT. ADEPT enzymes such as CPG2 are directed to the tumor using a targeting molecule, such as a tumor-directed antibody, where they can activate the drug, thereby localizing the cytotoxic effects to the tumor site. Clearance times for the circulating enzyme vary widely from patient to patient,[53,54] requiring the development of a variety of complex methods for assaying serum enzyme levels.[53] As the Pro-CPG2 enzymes that we have developed exhibit reduced activity compared to their activated form in tumor cell culture,

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

selection of an appropriate protease that is upregulated in the tumor microenvironment could facilitate synergistic improvement in the pharmacological safety profile of ADEPT therapy. The ability to easily switch the activating protease offers a distinct advantage, as the approach can be readily adapted to tumors with varied protease expression profiles.

Conclusion

The computational design of zymogen forms for any chosen enzyme is a challenging problem, particularly for enzymes acting on small molecule substrates. This study presents a massively parallel computational design method and accompanying screening approach to develop diverse pro-enzymes of CPG2, a clinically useful enzyme acting on small molecule substrates, whose inhibition is readily reversed through proteolytic cleavage of the enzyme-pro-domain linker. Our computational method incorporates specific adjustments to tailor the design process for a metal-chelating active site, and our screening approach exploits advances in synthetic DNA technology and NGS to build and screen a library of thousands of diverse pro-domains without requiring specialized assays or equipment. Biophysical characterization of our best pro-enzyme design reveals that two mechanisms of inhibition are at play: binding of the substrate to the enzyme is impeded, and domain movements required for catalytic activity are also disrupted. Finally, we demonstrate that the activity differential between the inhibited and activated forms is sufficient to significantly impact tumor cell survival under methotrexate treatment. This design and screening strategy has potential implications for the development of a broad range of enzyme inhibitors and pro-enzymes, potentially yielding novel therapeutics and diagnostic tools.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Helical pro-domain design. To assess the feasibility of auto-inhibiting CPG2 using a pro-domain, an idealized helical backbone structure was used. Either two or three repeats of the helical "EAAAK" sequence were fused, and placed in roughly the correct orientation to form either an N- or C-terminal connection with the $CPG2_{CP\text{-}N89}$ circular permutant.[24] Docking was performed using script 1 (Table S6). As a means to "filter" the resulting placements of the helix, Rosetta Match[55,56] was used to build a zinc-coordinating residue (Cys, Asp, Glu, or His) between the helix and the zinc coordination site ordinarily occupied by substrate.[35] For designs in which co-ordination was deemed possible, steric clashes between the pro-domain and catalytic domain were relieved by further sequence and structure optimization (script 2, Table S6). Favorably-scoring designs were identified, and eight "consensus designs" were selected to be tested experimentally.

Generation of the $CPG2_{CP\text{-}N89}$-K177A and Helical $CPG2_{CP\text{-}N89}$-K177A-Protease-Pro-domain constructs. Starting from the previously described $CPG2_{CP\text{-}N89}$ constructs (N- and C-terminally His-tagged),[24] the K177A mutation was introduced using standard Kunkel mutagenesis[57] or QuikChange mutagenesis (Agilent Technologies Inc.) protocols (Table S7, primer 1). To the C-terminal His-tag construct, a C-terminal TEV protease cleavage sequence was inserted upstream of the His-tag using primer extension PCR (Table S7, primers 2-3) followed by QuikChange mutagenesis. This construct was linearized at the C-terminus (Table S7, primers 4-5), and helix design sequences generated by primer extension PCR were inserted using Gibson assembly[58] (Table S7, primers 6-13). The N-terminal, thrombin cleavable His-tag variants were generated by amplifying the entire coding region (Table S7, primers 14-15), excluding the His-tag, and inserting it using Gibson assembly[58] into pET15b digested using NdeI and XhoI (New England

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Biolabs Inc.). The TEV cleavage site was substituted for an MMP cleavage site using primer extension PCR (Table S7, primers 16-17) followed by QuikChange mutagenesis.

Small domain library preparation. Following the success with the simple helical designs, a library of larger pro-domain candidates was prepared from small protein domains. The PDB was screened for high-quality crystal structures of 20-70 residue protein chains (resolution < 2.5 Å). Only monomers were considered, and a maximum of 70% sequence identity between two proteins in the library was allowed, yielding 258 small domains.

In addition, a large set of 40 residue *de novo* designed, well-folded "mini-proteins"[25] were included. Designs that had a stability score[25] of greater than 1.0 were included in our library, yielding an additional 2899 small domains with four different folds.

The structures of all 3157 domains were energy-optimized using the REF2015 scorefunction in Rosetta.

Initial placement using PatchDock. To globally dock the library of small domains in to CPG2, PatchDock[59,60] was used. A Rosetta-optimized model of the circularly permuted $CPG2_{CP\text{-}N89}$ dimer was set as the receptor. The K177A mutation was introduced in the starting model, as this mutation had only marginal effects on catalytic activity while improving the degree of inhibition in the helical pro-domain designs. The binding site included residues 114-115, 175-178, 200-201, 324, 360-362, and 384-385 (wild-type CPG2 numbering). A distance constraint of 25 Å between the C-terminus of $CPG2_{CP\text{-}N89}$-K177A and the N-terminus of the pro-domain was used. The top 50 models for each pro-domain were used in subsequent stages.

Hotspot placement and pro-domain design. Rosetta's "PlaceStub" mover[61] was employed to place small domains interacting with the catalytic zinc ion. An interaction between one of the active site zincs and a glutamate introduced in the pro-domain was defined as the "hotspot." A

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

library of 466 inverse rotamers of glutamate interacting with the active site zinc was built using the "TryRotamers" mover. Hotspot-directed pro-domain placement was performed using script 3 (Table S6). This protocol produced 4444 designs, which were further filtered to 1668 designs after analyzing the scores and filter values.

Increasing pro-domain sequence diversity. To generate a larger set of designs, further rounds of fixed backbone design were applied to the set of 4444 designs produced from the Hotspot protocol. Either the standard fixed-backbone design protocol or a design protocol (HBNet) aimed at obtaining hydrogen bond networks at the interface[62,63] was used as described in scripts 4 and 5 (Table S6). These protocols produced 8757 (fixed backbone) and 4993 (HBNet) additional designs, for a total of 15,418 designs (Table S8).

Disulfide variant design. Disulfide-bridge variants were designed starting from the base pro-domain models using the Disulfidize mover[41] (script 8, Table S6) and evaluated for quality manually.

Pro-CPG2 library design. We intended to screen 7500 sequences, including 31 controls, requiring us to reduce our 15,418 designs down to 7469 based on computational scoring metrics. First, designs were excluded based on filter cutoffs as described in Table S3. Designs in which the pro-domain clashed with its symmetric partner and duplicate sequences were also removed, leaving 7511 designs (1256 designs from the original set, 2856 from the fixed backbone design set, and 3399 from the HBNet design set). The poses with the lowest Rosetta total_scores from each design set were dropped, until a total of 7469 designs remained. These sequences, plus 31 controls known to inhibit or not inhibit CPG2 from the original helical design strategy, were codon-optimized using Biopython.[64] PCR adapters were added (5' adapter: 5'-CTTTATTTTCAAGGGGGGTCT; 3' adapter: TAAAAGCTTAATTAGCTGAG-3'), and these

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sequences were ordered as an Oligonucleotide Library (Agilent Technologies Inc.) (see https://github.com/BYachnin/pro_cpg2_design_scripts/blob/main/generate_sequences.py).

Pro-CPG2 library construction. Starting with $CPG2_{CP\text{-}N89}$ with an N-terminal pelB leader sequence in the pQE-80L expression vector, the K177A variant was prepared as described above to generate pelB-$CPG2_{CP\text{-}N89}$-K177A. This construct was linearized at the C-terminus after incorporating a C-terminal TEV protease cleavage sequence (Table S7, primers 5 and 18). The parental DNA was digested using DpnI.

The Oligonucleotide Library was amplified using the standard Herculase II Fusion PCR protocol (Agilent Technologies Inc.), using 15 PCR cycles (Table S7, primers 19-20). The PCR product was purified. In a 20 μL total volume, 100 ng each of the linearized vector and library of inserts were combined with NEBuilder HiFi DNA Assembly Master Mix and incubated at 50ºC for 1 hour. NEB 10-beta Electrocompetent *E. coli* (50 μL) were transformed with 2 μL of the resulting reaction product. After allowing the cells to recover for 1 hour at 37ºC in 1 mL of SOC, the transformants were expanded into 10 mL of LB media with 100 μg/mL ampicillin and grown overnight at 37ºC. The plasmid DNA was purified and will be referred to as the "Full Plasmid Library."

Massively parallel library screening. The Full Plasmid Library DNA was transformed into electrocompetent CAG12184 cells,[26] obtained from the Yale Coli Genetic Stock Center (CGSC #7437). The transformants were expanded into 10 mL of LB media with 100 μg/mL of ampicillin and 15 μg/mL tetracyclin and grown overnight at 37ºC. The cells were resuspended in 50 mL of fresh media and grown again overnight at 37ºC.

Multiple 50 mL cultures of M9 minimal media containing 0.4% glucose, 0.2% casamino acids (made from pure amino acids), 100 μg/mL ampicillin, 15 μg/mL tetracyclin, and 10 mM

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

isopropyl β-D-1-thiogalactopyranoside were inoculated with 1 mL of the library culture. Triplicate library cultures containing 0 μM, 30 μM, 50 μM, or 100 μM methotrexate were prepared alongside control cultures containing 30 μM methotrexate inoculated with CAG12184 cells harboring pQE-80L plasmids with pelB-CPG2$_{CP\text{-}N89}$-K177A, pelB-mCN2, or the helical pro-domain controls. These cultures were grown at 30ºC for 14 hours. Plasmid library DNA from each culture was purified and will be referred to as the "Selected Plasmid Libraries." The 12 Selected Plasmid Libraries and the Full Plasmid Library were amplified using 10-12 cycles of the standard Q5 polymerase PCR protocol (Table S7, primers 19-20). These PCR products, as well as the Oligonucleotide Library, were prepared for Illumina NGS. Samples were amplified with an additional 8 PCR cycles to incorporate DNA barcodes using the Illumina Nextera XT kit. A Bioanalyzer 2100 was used to confirm library size and quantity. Samples were mixed in equimolar ratio and sequenced on the MiSeq instrument using a 300 bp cycle that yields ~10 million reads. After demultiplexing and removing adapter sequences, forward and reverse reads with a minimum 15 bp overlap with 10% error were joined using Flash.[65] Sequences were aligned using Magic-BLAST[66] (FASTQ sequences as queries and the Oligonucleotide Library sequences as our subject database). Alignments flagged as being unmapped or a secondary alignment were discarded, as were alignments with insertions, deletions, skipped regions, hard clipping, padding, or mismatches. Sequences with soft clipping outside of the beginning and end of each sequence or longer than 70 bases each were discarded. The counts of the remaining matches to each of the designs in the Oligonucleotide Library were converted to percentages of the total number of matched sequences present in each NGS sample. Across replicates, the average, standard deviation, and percent error of the percentages for each design was calculated (see

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

https://github.com/BYachnin/pro_cpg2_design_scripts/blob/main/deep_seq_magicblast.R). Computational metrics and NGS-derived data were deposited to ProtaBank.

To exclude unreliable sequencing data, we filtered designs according to four criteria: the average percentage of the total number of sequences for that design needed to be at least 10^{-3}% in both the (1) Full Plasmid Library and (2) Selected Plasmid Library with a methotrexate concentration of 0. (3) The percent error in these libraries needed to be less than 50% for the Selected Plasmid Library with a methotrexate concentration of 0, and (4) less than 40% for the Full Plasmid Library. This reduced the number of designs under consideration from 7469 to 1396 (Table S8). Strongly inhibiting pro-CPG2 designs should decrease as a fraction of the number of sequences present in a culture as the methotrexate concentration increases. To identify such designs, an "enrichment ratio" was calculated by dividing the percentages for a design in the Selected Plasmid Libraries (with varying methotrexate concentrations) by the percentage of that same design in the Selected Plasmid Library with a methotrexate concentration of 0. We identified any design that had an enrichment ratio lower than that of all the helical pro-CPG2 controls as a "Predicted Pro-CPG2" design candidate. The 27 designs in which the enrichment ratio at 50 μM and 100 μM methotrexate were both less than the enrichment ratio at 30 μM methotrexate were identified as "High Probability Pro-CPG2" design candidates. These designed were inspected individually, and twelve were selected for expression as individual designs (Figure 3 and Table S8). In addition, four designs that were expected to be poorly inhibited were also expressed (Figure 3).

Preparation of individual pro-CPG2 design constructs. The vector backbone and CPG2 scaffold was amplified from the helical design plasmids by PCR (Table S7, primers 21-22). The

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sixteen designs were ordered as gBlocks (Integrated DNA Technologies Inc.) and inserted into this vector using Gibson assembly.[58]

The TEV cleavage site was substituted for an MMP site using primer extension PCR (Table S7, primers 29-32) followed by QuikChange mutagenesis (Agilent Technologies Inc.).

Disulfide variants were generated using QuikChange Multi-Site Mutagenesis (Agilent Technologies Inc.) (Table S7, primers 23-28), starting from the Pro-CPG2-1 and Pro-CPG2-2 TEV variants.

Preparation of pro-domain constructs. The pro-domains were amplified using PCR (Table S7, primers 33-36) and re-inserted into pET15b, linearized by NdeI and BamHI.

Protein expression. Plasmid DNA was transformed into BL21(DE3) cells. LB media containing 100 μg/mL ampicillin was inoculated with a single colony and grown overnight at 37ºC. ZYP-5052 media[67] (5 mL for screening, 500 mL for other applications) was inoculated with the starter culture, using 1/50th of the culture volume, and grown for 3 hours at 37ºC followed by 18ºC overnight. The cells were harvested by centrifugation at 4ºC and re-suspended in 50 mM Tris, 300 mM NaCl, 30 mM imidazole, pH 7.4 supplemented with lysozyme, PMSF, and DNase I. For screening, 1.5X BugBuster (EMD Millipore) was also added.

Protein purification for pro-CPG2 design activity screening. The BugBuster cell suspension was incubated at room temperature for 20 minutes, and then 40 minutes at 4ºC. The lysate was centrifuged at 20,000×g, and the supernatant was applied to 150 μL of Ni-NTA agarose resin (Qiagen), shaking for 1 hour. The resin was washed with 50 mM Tris, 300 mM NaCl, 30 mM imidazole, pH 7.4, and bound protein was eluted using 50 mM Tris, 300 mM NaCl, 300 mM imidazole, pH 7.4. The protein purity was analyzed using SDS-PAGE, and the protein was dialyzed against 50 mM Tris, 100 mM NaCl, pH 7.4.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Protein purification for Michaelis-Menten kinetics and structural biology. The cells were lysed by sonication (Fisherbrand Model 705 Sonic Dismembrator), and the soluble fraction was isolated by centrifugation at 20,000×g for 30-60 minutes. After filtering, the supernatant was applied to a 5 mL NiNTA Superflow column (Qiagen) equilibrated with 50 mM Tris, 300 mM NaCl, 30 mM imidazole, pH 7.4 using an Äkta Pure (GE Life Sciences). The protein was eluted with a linear gradient of 30 mM – 300 mM imidazole, and fractions containing the CPG2 designs were pooled.

The samples were desalted into 50 mM Tris, 300 mM NaCl, pH 7.4 using a HiPrep 26/10 Desalting column (GE Life Sciences) and then supplemented with 2.5 mM $CaCl_2$ and ~1 unit of thrombin/mg protein. The sample was incubated at room temperature for 3 hours. Uncleaved protein and thrombin were removed by re-applying the protein to a 5 mL NiNTA Superflow column (Qiagen) with a 1 mL HiTrap Benzamidine Fast Flow (high sub) column (GE Life Sciences) connected in series. The flowthrough was concentrated to ~2 mL and applied to a HiLoad 16/600 Superdex 200 pg column (GE Life Sciences) equilibrated with 50 mM Tris, 100 mM NaCl, pH 7.4. The purity of the sample was assessed using SDS-PAGE, and fractions containing pure protein were pooled and concentrated to 10-20 mg/mL.

Protein purification of the Isolated Pro-Domains. The cells were lysed by sonication as described above. The supernatant was applied to 5 mL of Ni-NTA agarose resin (Qiagen), shaking for 1 hour. The resin was washed with 50 mM Tris, 300 mM NaCl, 30 mM imidazole, pH 7.4, and bound protein was eluted using 50 mM Tris, 300 mM NaCl, 300 mM imidazole, pH 7.4. The protein purity was analyzed using SDS-PAGE, and the protein was dialyzed against 20 mM sodium phosphate, 100 mM NaCl, pH 7.4.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

CPG2 activity assays. CPG2 was first diluted to 100 μg/mL in 50 mM Tris, 100 mM NaCl, pH 7.4, and then to 1-10 μg/mL in the same buffer, after which 10 μL was mixed with 90 μL of 50 mM Tris, 50 mM NaCl, 0.1 mM $ZnSO_4$, pH 7.4 containing 100 μM of methotrexate pre-heated to 37ºC (final protein concentration of 100-1000 ng/mL). The absorbance at 320 nm was measured over time in a 96-well plate (Greiner Bio-One UV-Star Half Area 96-well Microplate) using a Tecan Infinite M200 Pro.

Pro-CPG2 proteolytic re-activation. For the TEV protease variants, the protein was diluted to 1.0-1.5 mg/mL in 50 mM Tris, 100 mM NaCl, pH 7.4. TEV protease, stored at 1 mg/mL in 50 mM sodium phosphate, 200 mM NaCl, 10% glycerol, 2 mM EDTA, 10 mM DTT, pH 7.5, was added to a final concentration of 0.1 mg/mL. The samples were incubated at 4ºC for 48 hours. In some experiments, these samples were dialysed against 50 mM Tris, 100 mM NaCl, pH 7.4 using a 20,000 molecular weight cutoff membrane sufficient to retain the CPG2 protein while allowing passage of the "free" pro-domains following cleavage. Following proteolysis, the samples were tested for activity as described above. In addition, samples were analyzed using LC-MS with a Velos Mass Spectrometer (Thermo Scientific). In brief, the sample was loaded in a mixture of 85% buffer A (0.2% formic acid) and 15% buffer B (0.1% formic acid in acetonitrile). After 3 minutes, a linear gradient from 15%-100% buffer B was applied over 7 minutes, after which the column was held at 100% B for another 3 minutes. Data were analyzed using Thermo Scientific Protein Deconvolution software.

For the MMP Pro-CPG2 variants, the protein was diluted to 2 μM (approximately 100 μg/mL) in 50 mM Tris, 100 mM NaCl, 10 mM $CaCl_2$, 5 μM $ZnSO_4$, pH 7.4, with a final concentration of 40 nM MMP-2 (Sigma Aldrich, 10 μg in 10 mM sodium phosphate pH 7.5, 0.1 mM $CaCl_2$

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

reconstituted in 0.1% bovine serum albumin). The samples were incubated at 4ºC for 24 hours, and the samples were tested for activity as described above.

Michaelis-Menten kinetics. CPG2 was first diluted to 100 µg/mL in 50 mM Tris, 100 mM NaCl, pH 7.4, and then to 50 ng/mL, or 500 ng/mL for the disulfide variants, in the same buffer. 10 µL of this sample was mixed with 90 µL of 50 mM Tris, 50 mM NaCl, 0.1 mM $ZnSO_4$, pH 7.4 containing between 1-300 µM methotrexate, pre-heated to 37ºC. The final protein concentration was approximately 1 nM (or 10 nM for the disulfide variants). The absorbance at 320 nm was measured over time as described above. Absorbance values were converted to methotrexate concentrations, and initial rates were determined using linear regression. Michaelis-Menten parameters were determined using non-linear regression.

Circular dichroism of individual pro-domains. The pro-domains were filtered and diluted to 0.1 mg/mL using 20 mM sodium phosphate, 100 mM NaCl, pH 7.4. CD data were recorded in an Aviv model 420SF CD spectrometer at 4°C in 1.0 mm curvettes. CD spectra were acquired over the wavelength range of 190-260 nm in 2 nm increments with an averaging time interval of 2 s. Next, CD spectra were recorded for the protein samples starting at 10ºC and increasing in 5ºC increments to 100ºC. Buffer and duplicate measurements were recorded periodically. The CD spectra were smoothed using the Savitzky-Golay filter,[68] and the buffer spectrum at the appropriate temperature was subtracted from the sample spectra. The average baseline molar ellipticity between 250-260 nm was subtracted from the CD curve.

Isothermal titration calorimetry (ITC). Thermodynamic binding parameters for the association of $CPG2_{CP\text{-}N89}$-K177A with Pro-Domain-1 were determined calorimetrically employing a MicroCal VP-ITC (Malvern Panalytical, Northampton, MA). Protein ($CPG2_{CP\text{-}N89}$-K177A) and Pro-Domain-1 stock solutions were prepared and dialyzed exhaustively against 50

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mM sodium phosphate, 100 mM NaCl, pH 7.4. Each ITC experiment consisted of thirty consecutive 10.0 µL injections during which the reaction heat is monitored and integrated over a 5.0 minute period under constant stirring conditions. The experimental protocol has been designed to ensure that there is a tenfold excess of Pro-Domain-1 in the titration syringe (i.e., 150, 250, or 500 µM) relative to the initial enzyme concentration (i.e., 15, 25, or 50 µM) in the sample cell. The resultant binding isotherms are generated by recording the integrated reaction heats normalized for peptide concentration versus the Pro-Domain:Enzyme ratio. Inspection of these profiles reveals that formation of the enzyme-Pro-Domain-1 complex is characterized by mid-low micromolar binding affinities and relatively low enthalpies. Accordingly, we have employed the NITPIC/SEDPHAT suite of software programs[69,70] to facilitate unbiased baseline assignment and peak integration. A nonlinear least squares fit of the resultant profile to a single site binding model yields thermodynamic parameters for the $CPG2_{CP\text{-}N89}$-K177A:Pro-Domain-1 complex including the affinity (K_a), Gibbs free energy (ΔG), enthalpy (ΔH), and entropy (ΔS).

Small-angle X-ray scattering. For SAXS, protein samples were prepared as described above. The protein was diluted to 10 mg/mL using 50 mM Tris, 100 mM NaCl, pH 7.4, and 2 mM methotrexate (if applicable). These samples were microdialyzed against the same buffer to establish a "matching" buffer for SAXS buffer subtraction. Following dialysis, the protein samples were diluted with their matched buffer by a factor of two and four, yielding protein concentrations of ~2.5 and ~5 mg/mL. SAXS data were collected for these samples, the ~10 mg/mL sample, and two matched buffer samples at the SIBYLS beamline (beamline 12.3.1) at the Advanced Light Source.[71] An exposure of 10 seconds split over 32 frames was collected. The buffer signal was subtracted from the protein curves, and the concentration series was merged in PRIMUS[72] such that noisy data at higher q-ranges and interparticle repulsion at low q-

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ranges were eliminated from the merged scattering curves. Particle distance distribution plots were generated using GNOM,[73] and ScÅtter (Rodic & Rambo, www.bioisis.net) was used to determine other SAXS parameters, including the particle density,[74] Porod exponent,[74] and volume of correlation.[75]

X-ray Crystallography. CPG2$_{CP-N89}$-K177A-MMP-Helix (best helical pro-domain sequence: AEAAWKEAEAKDWAAKA) was prepared as described above to a concentration of 19 mg/mL in 50 mM Tris 100 mM NaCl pH 7.4 and 0.2 $ZnSO_4$. Crystals were grown using hanging drop vapor diffusion. Reservoirs containing 750 µL of 20 mM Tris pH 8.0, 10% glycerol, and 10% PEG 3350 were prepared in 24 well hanging drop vapor diffusion plates. Equal volumes of protein solution and reservoir solution were mixed on a cover slip and suspended over the reservoir. The plate was incubated at 20ºC, and crystals were obtained in 1-2 weeks.

Crystals were harvested and flash-frozen in a nitrogen cryostream. Data were collected on a Rigaku MicroMax-007 generator equipped with an RAxis IV++ detector. Data were processed using xia2[76] with the XDS and SCALE[77] pipeline, indexing with peaks from all images and with the ice rings filter enabled. The structure was solved by molecular replacement using Phaser[78] and positional and B-factor refinement was performed using Refmac[79] with non-crystallographic constraints. Manual inspection and model building was performed with Coot.[80]

CPG2 Cell Viability Assay. PC-3 cells were seeded at 1,400 cells/100 µL in 96 well plates in RPMI media (Sigma-Aldrich) with 10% FBS, 1% Pen/Strep, and 1% GlutaGro (Corning). After an overnight incubation at 37°C and 5% CO_2, media in the wells was replaced with fresh RPMI media with 5% charcoal-stripped (CSS)-FBS, 1% Pen/Strep, and 1% GlutaGro (Corning). Cells were then grown for 3 days. We then added one of the following three drug groups: 50 nM

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

methotrexate alone, 50 nM methotrexate + CPG2$_{CP-N89}$-K177A (100-5000 pM), or 50 nM methotrexate + Pro-CPG2-1-Disulfide (100-5000 pM). Treatment lasted for 6 days with replenishment of the media and enzymes after 3 days. Cell viability was measured using WST-1 at 1:10 dilution with CSS-FBS media at absorbance of 450 nm (Tecan 1100 Plate Reader).

Acknowledgements

This work was supported by grants from NIH (R01GM132565 to SDK) ******. BJY was supported by a Canadian Institutes of Health Research Post-Doctoral Fellowship and a Mistletoe Research Fellowship. LRA was a recipient of a RosettaCommons Research Experience for Undergraduates internship program. VMT was funded by the National Institute of General Medical Sciences (NIGMS) Grant T32 GM135141. We would like to thank Joseph D. Bauman and Eddy Arnold (Rutgers University Center for Advanced Biotechnology and Medicine) for providing support and resources for the crystallography experiments, and Vikas Nanda (Rutgers University Center for Advanced Biotechnology and Medicine) for providing access to his Aviv model 420SF CD spectrometer. We would like to acknowledge Dr. Dibyendu Kumar, Dr. Min Tu, and the Waksman Genomics Core Facility for performing the NGS analysis. SAXS data were collected at the Advanced Light Source (ALS), SIBYLS beamline on behalf of US DOE-BER, through the Integrated Diffraction Analysis Technologies (IDAT) program. Additional support comes from the NIGMS project ALS-ENABLE (P30 GM124169) and a High-End Instrumentation Grant S10OD018483. Mass spectroscopy data were collected by Amenah Soherwardy and Haiyan Zheng (Rutgers University Biological Mass Spectrometry Facility). The authors acknowledge the Office of Advanced Research Computing (OARC) at Rutgers, The State University of New Jersey for providing access to the Amarel cluster and associated research computing resources that have contributed to the results reported here.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author contributions

BJY and SDK designed the project. BJY, CASAM, DPR, JMD, and SDK designed the experiments. BJY performed all experiments and analysis of data unless otherwise indicated. LRA performed the computational design of the first-generation helical pro-enzymes. REW and VMT performed the PC-3 cell survival assays and analyzed the data. CASAM and DPR performed the ITC experiments and analyzed the data. All authors contributed to the writing and/or editing of the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Khan, A. R. & James, M. N. G. Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes. *Protein Sci.* **7**, 815–836 (1998).
2. Johnson, R. J., Lin, S. R. & Raines, R. T. A ribonuclease zymogen activated by the NS3 protease of the hepatitis C virus. *FEBS J.* **273**, 5457–5465 (2006).
3. Turcotte, R. F. & Raines, R. T. Design and Characterization of an HIV-Specific Ribonuclease Zymogen. *AIDS Res. Hum. Retroviruses* **24**, 1357–1363 (2008).
4. Mitrea, D. M., Parsons, L. S. & Loh, S. N. Engineering an artificial zymogen by alternate frame protein folding. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2824–9 (2010).
5. Plainkum, P., Fuchs, S. M., Wiyakrutta, S. & Raines, R. T. Creation of a zymogen. *Nat. Struct. Biol.* **10**, 115–119 (2003).
6. Puskas, J. *et al.* Development of an attenuated interleukin-2 fusion protein that can be

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

activated by tumour-expressed proteases. *Immunology* **133**, 206–220 (2011).

7. Gerspach, J. *et al.* Restoration of membrane TNF-like activity by cell surface targeting and matrix metalloproteinase-mediated processing of a TNF prodrug. *Cell Death Differ.* **13**, 273–284 (2006).

8. Yogo, T., Umezawa, K., Kamiya, M., Hino, R. & Urano, Y. Development of an Activatable Fluorescent Probe for Prostate Cancer Imaging. *Bioconjug. Chem.* (2017) doi:10.1021/acs.bioconjchem.7b00233.

9. DeAngelis, L. M., Tong, W. P., Lin, S., Fleisher, M. & Bertino, J. R. Carboxypeptidase G2 rescue after high-dose methotrexate. *J. Clin. Oncol.* **14**, 2145–2149 (1996).

10. Green, J. M. Glucarpidase to combat toxic levels of methotrexate in patients. *Ther. Clin. Risk Manag.* **8**, 403–13 (2012).

11. Wyatt, K. D. *et al.* Delayed Methotrexate Clearance Despite Carboxypeptidase-G2 (Glucarpidase) Administration in 2 Patients With Toxic Methotrexate Levels. *J. Pediatr. Hematol. Oncol.* **40**, 152–155 (2018).

12. Christensen, A. M. *et al.* Resumption of high-dose methotrexate after acute kidney injury and glucarpidase use in pediatric oncology patients. *Cancer* **118**, 4321–4330 (2012).

13. Francis, R. J. *et al.* A phase I trial of antibody directed enzyme prodrug therapy (ADEPT) in patients with advanced colorectal carcinoma or other CEA producing tumours. *Br. J. Cancer* **87**, 600–607 (2002).

14. Napier, M. P., Sharma, S. K. & Springer, C. J. Antibody-directed Enzyme Prodrug Therapy: Efficacy and Mechanism of Action in Colorectal Carcinoma. *Clin. Cancer Res.* **6**, 765–772 (2000).

15. Springer, C. J. *et al.* Optimization of Alkylating Agent Prodrugs Derived from Phenol and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Aniline Mustards: A New Clinical Candidate Prodrug (ZD2767) for Antibody-Directed Enzyme Prodrug Therapy (ADEPT). *J. Med. Chem* **38**, 5051–5065 (1995).

16. Capucha, V., Mendes, E., Francisco, A. P. & Perry, M. J. Development of triazene prodrugs for ADEPT strategy: New insights into drug delivery system based on carboxypeptidase G2 activation. *Bioorg. Med. Chem. Lett.* **22**, 6903–6908 (2012).

17. Davies, L. C. *et al.* Novel Fluorinated Prodrugs for Activation by Carboxypeptidase G2 Showing Good in vivo Antitumor Activity in Gene-Directed Enzyme Prodrug Therapy. *J. Med. Chem.* **48**, 5321–5328 (2005).

18. Sharma, S., Bagshawe, K., Burke, P., Boden, R. & Rogers, G. Inactivation and clearance of an anti-CEA carboxypeptidase G2 conjugate in blood after localisation in a xenograft model. *Br. J. Cancer* **61**, 659–662 (1990).

19. Jamin, Y. *et al.* Noninvasive detection of carboxypeptidase G2 activity in vivo. *NMR Biomed.* **24**, 343–350 (2011).

20. Bagshawe, K. D. *et al.* A cytotoxic agent can be generated selectively at cancer sites. *Br. J. Cancer* **58**, 700–3 (1988).

21. Martin, J. *et al.* Antibody-directed enzyme prodrug therapy: pharmacokinetics and plasma levels of prodrug and drug in a phase I clinical trial. *Cancer Chemother. Pharmacol.* **40**, 189–201 (1997).

22. Blacklock, K. M., Yachnin, B. J., Woolley, G. A. & Khare, S. D. Computational Design of a Photocontrolled Cytosine Deaminase. *J. Am. Chem. Soc.* **140**, 14–17 (2018).

23. Rowsell, S. *et al.* Crystal structure of carboxypeptidase G2, a bacterial enzyme with applications in cancer therapy. *Structure* **5**, 337–347 (1997).

24. Yachnin, B. J. & Khare, S. D. Engineering carboxypeptidase G2 circular permutations for

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the design of an autoinhibited enzyme. *Protein Eng. Des. Sel.* **30**, 321–331 (2017).

25. Rocklin, G. J. *et al.* Global analysis of protein folding using massively parallel design, synthesis, and testing. *Science* **357**, 168–175 (2017).

26. Kopytek, S. J., Dyer, J. C., Knapp, G. S. & Hu, J. C. Resistance to methotrexate due to AcrAB-dependent export from Escherichia coli. *Antimicrob. Agents Chemother.* **44**, 3210–2 (2000).

27. Volpato, J. P., Fossati, E. & Pelletier, J. N. Increasing Methotrexate Resistance by Combination of Active-site Mutations in Human Dihydrofolate Reductase. *J. Mol. Biol.* **373**, 599–611 (2007).

28. Volpato, J. P. *et al.* Multiple conformers in active site of human dihydrofolate reductase F31R/Q35E double mutant suggest structural basis for methotrexate resistance. *J. Biol. Chem.* **284**, 20079–89 (2009).

29. Phan, J. *et al.* Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277**, 50564–72 (2002).

30. Morgunova, E. *et al.* Structure of human pro-matrix metalloproteinase-2: Activation mechanism revealed. *Science* **284**, 1667–1670 (1999).

31. Minetti, C. A. S. A. *et al.* Energetics of lesion recognition by a DNA repair protein: Thermodynamic characterization of formamidopyrimidine-glycosylase (Fpg) interactions with damaged DNA duplexes. *J. Mol. Biol.* **328**, 1047–1060 (2003).

32. Spolar, R. S., Ha, J. H. & Record, M. T. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 8382–8385 (1989).

33. Ha, J. H., Spolar, R. S. & Record, M. T. Role of the hydrophobic effect in stability of site-

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

specific protein-DNA complexes. *J. Mol. Biol.* **209**, 801–816 (1989).

34. Nunes, A. M., Minetti, C. A. S. A., Remeta, D. P. & Baum, J. Magnesium Activates Microsecond Dynamics to Regulate Integrin-Collagen Recognition. *Structure* **26**, 1080-1090.e5 (2018).

35. Unno, H. *et al.* Structural Basis for Substrate Recognition and Hydrolysis by Mouse Carnosinase CN2. *J. Biol. Chem.* **283**, 27289–27299 (2008).

36. Girish, T. S. & Gopal, B. Crystal structure of Staphylococcus aureus Metallopeptidase (Sapep) Reveals Large Domain Motions between the Manganese-bound and Apo-states. *J. Biol. Chem.* **285**, 29406–29415 (2010).

37. Jozic, D. *et al.* Crystal Structure of the Dinuclear Zinc Aminopeptidase PepV from Lactobacillus delbrueckii Unravels Its Preference for Dipeptides. *Structure* **10**, 1097–1106 (2002).

38. Håkansson, K. & Miller, C. G. Structure of peptidase T from <I>Salmonella typhimurium</I>. *Eur. J. Biochem.* **269**, 443–450 (2002).

39. Lindner, H. A. *et al.* Essential Roles of Zinc Ligation and Enzyme Dimerization for Catalysis in the Aminoacylase-1/M20 Family. *J. Biol. Chem.* **278**, 44496–44504 (2003).

40. Nocek, B. *et al.* The Dimerization Domain in DapE Enzymes Is required for Catalysis. *PLoS One* **9**, e93593 (2014).

41. Bhardwaj, G. *et al.* Accurate de novo design of hyperstable constrained peptides. *Nature* **538**, 329–335 (2016).

42. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A. & Sali, A. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res.* **44**, W424–W429 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

43. Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28**, 768–773 (1995).

44. Reyes, F. E., Schwartz, C. R., Tainer, J. A. & Rambo, R. P. Methods for Using New Conceptual Tools and Parameters to Assess RNA Structure by Small-Angle X-Ray Scattering. *Methods Enzymol.* **549**, 235–263 (2014).

45. Tai, S. *et al.* PC3 is a cell line characteristic of prostatic small cell carcinoma. *Prostate* **71**, 1668–1679 (2011).

46. Obexer, R., Pott, M., Zeymer, C., Griffiths, A. D. & Hilvert, D. Efficient laboratory evolution of computationally designed enzymes with low starting activities using fluorescence-activated droplet sorting. *Protein Eng. Des. Sel.* **29**, 355–366 (2016).

47. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S. & Tawfik, D. S. High-Throughput Screening of Enzyme Libraries: Thiolactonases Evolved by Fluorescence-Activated Sorting of Single Cells in Emulsion Compartments. *Chem. Biol.* **12**, 1281–1289 (2005).

48. Yang, G. & Withers, S. G. Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution. *ChemBioChem* **10**, 2704–2715 (2009).

49. Chen, B. *et al.* High-throughput analysis and protein engineering using microcapillary arrays. *Nat. Chem. Biol.* **12**, 76–81 (2016).

50. Longwell, C. K., Labanieh, L. & Cochran, J. R. High-throughput screening technologies for enzyme engineering. *Curr. Opin. Biotechnol.* **48**, 196–202 (2017).

51. Mair, P., Gielen, F. & Hollfelder, F. Exploring sequence space in search of functional enzymes using microfluidic droplets. *Curr. Opin. Chem. Biol.* **37**, 137–144 (2017).

52. Bunzel, H. A., Garrabou, X., Pott, M. & Hilvert, D. Speeding up enzyme discovery and

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

engineering with ultrahigh-throughput methods. *Curr. Opin. Struct. Biol.* **48**, 149–156 (2018).

53. Springer, C. J., Poon, G. K., Sharma, S. K. & Bagshawe, K. D. Analysis of antibody-enzyme conjugate clearance by investigation of prodrug and active drug in an ADEPT clinical study. *Cell Biophys.* **24**–**25**, 193–207 (1994).

54. Bagshawe, K. D. Targeting: The ADEPT Story So Far. in *Current Drug Targets* 152–157 (2009). doi:10.2174/138945009787354520.

55. Zanghellini, A. *et al.* New algorithms and an in silico benchmark for computational enzyme design. *Protein Sci.* **15**, 2785–2794 (2006).

56. Richter, F., Leaver-Fay, A., Khare, S. D., Bjelic, S. & Baker, D. De Novo Enzyme Design Using Rosetta3. *PLoS One* **6**, e19230 (2011).

57. Kunkel, T. A. Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl. Acad. Sci. U. S. A.* **82**, 488–92 (1985).

58. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009).

59. Duhovny, D., Nussinov, R. & Wolfson, H. J. Efficient Unbound Docking of Rigid Molecules. in *Proceedings of the 2nd Workshop on Algorithms in Bioinformatics (WABI) Rome, Italy, Lecture Notes in Computer Science 2452* (eds. Guigó, R. & Gusfield, D.) 185–200 (Springer-Verlag Berlin Heidelberg, 2002).

60. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363–W367 (2005).

61. Fleishman, S. J. *et al.* Hotspot-Centric De Novo Design of Protein Binders. *J. Mol. Biol.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

413, 1047–1062 (2011).

62. Boyken, S. E. *et al.* De novo design of protein homo-oligomers with modular hydrogen-bond network–mediated specificity. *Science* **352**, 680–687 (2016).

63. Maguire, J. B., Boyken, S. E., Baker, D. & Kuhlman, B. Rapid Sampling of Hydrogen Bond Networks for Computational Protein Design. *J. Chem. Theory Comput.* **14**, 2751–2760 (2018).

64. Cock, P. J. A. *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).

65. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

66. Boratyn, G. M., Thierry-Mieg, J., Thierry-Mieg, D., Busby, B. & Madden, T. L. Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**, 405 (2019).

67. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).

68. Savitzky, A. & Golay, M. J. E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **36**, 1627–1639 (1964).

69. Brautigam, C. A., Zhao, H., Vargas, C., Keller, S. & Schuck, P. Integration and global analysis of isothermal titration calorimetry data for studying macromolecular interactions. *Nat. Protoc.* **11**, 882–894 (2016).

70. Keller, S. *et al.* High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* **84**, 5066–5073 (2012).

71. Dyer, K. N. *et al.* High-Throughput SAXS for the Characterization of Biomolecules in

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Solution: A Practical Approach. in *Structural Genomics. Methods in Molecular Biology, vol 1091* (ed. Chen, Y. W.) 245–258 (Humana Press, 2014). doi:10.1007/978-1-62703-691-7_18.

72. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS : a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

73. Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* **25**, 495–503 (1992).

74. Rambo, R. P. & Tainer, J. A. Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers* **95**, 559–71 (2011).

75. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496**, 477–481 (2013).

76. Winter, G., Lobley, C. M. C. & Prince, S. M. Decision making in xia2. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **69**, 1260–1273 (2013).

77. Kabsch, W. XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D66**, 125–132 (2010).

78. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

79. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **D53**, 240–255 (1997).

80. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

81. Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *Journal of Molecular Biology* vol. 234 946–950 (1993).

Figure Legends

Figure 1. Utility of a protease-activatable form of CPG2. (a) Mechanism of activation of antibody-linked Pro-CPG2. The antibody (black) directs the inactive enzyme (red) to a protease-expressing cell (cell is orange, protease is yellow, antibody's targets are orange diamonds). The protease (yellow) cleaves the linker that connects the pro-domain (blue) to the enzyme, activating it (green). The enzyme can then process small molecules (teal and purple) which can have an effect (red arrow) on the protease-expressing cell. (b) Activated CPG2 can cleave and detoxify the drug, methotrexate, rescuing the protease expressing cells (black to orange). (c) Activated CPG2 can cleave and active the nitrogen mustard prodrug CMDA, killing the protease expressing cells (orange to black).

Figure 2. Massively-parallel design of Pro-CPG2. (abc) The three Pro-CPG2 designs selected for detailed study: (a) Pro-CPG2-1, (b) Pro-CPG2-2, (c) Pro-CPG2-3. The pro-domain (green), catalytic domain (blue), and dimerization domain (wheat) are highlighted. Inset: Important design characteristics are shown, including (a) close contacts between the pro-domain and the enzyme, sufficient to allow design of a disulfide staples, (b) proximity between the C-terminus of the circular permutant and the N-terminus of the pro-domain, and (c) an interaction between a glutamate residue on the pro-domain and one of the active site zinc ions (grey spheres). (d) The conformational equilibrium required for CPG2 activity, highlighting the role of the pro-domain (green). (e) The design strategy is depicted as a flow chart.

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3. Next-generation sequencing results from the massively parallel selection assay. The logarithm of the relative enrichment ratio is given at increasing methotrexate concentrations. The helical designs that are known not to inhibit (green) or to mildly inhibit (orange) $CPG2_{CP\text{-}N89}$ are shown alongside the twelve designs predicted to be Pro-CPG2s (red) and predicted not to be Pro-CPG2s (blue).

Figure 4. Activity and thermodynamic data for the Pro-CPG2 designs and their disulfide variants. (a-c) The Michaelis-Menten kinetics (a, K_M; b, k_{cat}; c, k_{cat}/K_M) for the three base Pro-CPG2 designs are shown. In yellow, the disulfide variants are grouped with the base design they are derived from. See also Table S3. (d-e) The recovery of activity following incubation with (d) MMP-2 and (e) TEV protease is shown. (f) Temperature-dependent thermodynamic binding signatures (left axis) and affinity (right axis) characterizing the association of $CPG2_{CP\text{-}N89}$-K177A and Pro-Domain-1 deduced from ITC measurements. See also Table S4.

Figure 5. Small angle X-ray scattering (SAXS) analysis for $CPG2_{CP\text{-}N89}$-K177A and the Pro-CPG2 designs. (a) The radius of gyration of the enzymes in the absence (red) and presence (green) of methotrexate (left axis) and difference in radius of gyration (blue, right axis) are shown. (b) Results from fitting the two-state model (open and closed) to the SAXS data using FoXS are shown, with "-MTX" and "+MTX" indicating the absence or presence of methotrexate. The disulfide variant of Pro-CPG2-1 is indicated by "DS." The quality of the fits is shown in Figure S8. (c-d) The (c) radius of gyration-based and (d) volume of correlation-based dimensionless Kratky plots are shown. Samples with solid lines have no methotrexate

bioRxiv preprint doi: https://doi.org/10.1101/2021.03.25.437042; this version posted March 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

present, while samples with dashed lines have methotrexate present. The horizontal and vertical dotted lines indicate cross at the peak position of a compact, spherical particle. In the R_g-based plot, if the peak is shifted up and to the right, this indicates that the particle is more elongated, and if the curve becomes a hyperbolic plateau, this is indicative of a disordered particle. In the V_c-based plot, a downwards shift indicates an increase in surface-to-volume ratio relative to that of a sphere.

Figure 6. Pro-CPG2-1-Disulfide has reduced ability to cleave and detoxify methotrexate (MTX) compared to $CPG2_{CP\text{-}N89}$-K177A. PC3 cells were treated with MTX at 50 nM, and then treated with either $CPG2_{CP\text{-}N89}$-K177A or Pro-CPG2-1-Disulfide. Enzyme concentrations between 250-5,000 pM showed 100% viability. Enzyme concentrations between 15-62pM showed a significant difference (unpaired t-test, $p<0.05$) of viability in PC3 cells between the two enzyme treatment groups, with cells treated with Pro-CPG2-1-Disulfide exhibiting a lower survival rate than cells treated with active CPG2. The dashed line indicates the cell viability in the presence of MTX alone (no enzyme treatment). Error bars indicate the standard deviation. *= $p< 0.05$, **= $p< 0.01$, ****= $p < 0.0001$