





# Selection for constrained peptides that bind to a single target protein

Andrew M. King<sup>1,2</sup>, Daniel A. Anderson <sup>1,3</sup>, Emerson Glassey<sup>1,3</sup>, Thomas H. Segall-Shapiro<sup>1</sup>, Zhengan Zhang<sup>1</sup>, David L. Niquille<sup>1</sup>, Amanda C. Embree<sup>2</sup>, Katelin Pratt<sup>2</sup>, Thomas L. Williams <sup>2</sup>, D. Benjamin Gordon <sup>1,2</sup> & Christopher A. Voigt <sup>1,2</sup>✉

Peptide secondary metabolites are common in nature and have diverse pharmacologically-relevant functions, from antibiotics to cross-kingdom signaling. Here, we present a method to design large libraries of modified peptides in *Escherichia coli* and screen them in vivo to identify those that bind to a single target-of-interest. Constrained peptide scaffolds were produced using modified enzymes gleaned from microbial RiPP (ribosomally synthesized and post-translationally modified peptide) pathways and diversified to build large libraries. The binding of a RiPP to a protein target leads to the intein-catalyzed release of an RNA polymerase  $\sigma$  factor, which drives the expression of selectable markers. As a proof-of-concept, a selection was performed for binding to the SARS-CoV-2 Spike receptor binding domain. A 1625 Da constrained peptide (AMK-1057) was found that binds with similar affinity ( $990 \pm 5$  nM) as an ACE2-derived peptide. This demonstrates a generalizable method to identify constrained peptides that adhere to a single protein target, as a step towards “molecular glues” for therapeutics and diagnostics.

<sup>1</sup>Synthetic Biology Center, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>2</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>3</sup>These authors contributed equally: Daniel A. Anderson, Emerson Glassey. ✉email: [cavoigt@gmail.com](mailto:cavoigt@gmail.com)

Short peptides have vast potential as therapeutics due to their intrinsic ability to bind to protein targets and the simplicity of creating combinatorial libraries. However, linear peptides are generally not good drugs due to their conformational flexibility, inability to cross the cell membrane, and sensitivity to proteases<sup>1</sup>. Bacteria overcome these issues by chemically modifying peptides to introduce functional groups and constrain their structures to improve affinity and stability. The modified peptides are secreted and act on eukaryotic cells by binding to surface proteins or crossing their membrane to interact with intracellular targets<sup>2–4</sup>. The peptides can inhibit cellular processes by occluding the active site of an enzyme or blocking a protein–protein interaction by binding to the interface. They can also act as a “molecular glue” to bring two proteins together, the most well-known example being rapamycin, which binds to two human proteins involved in signaling in order to suppress the host immune response<sup>5</sup>. Secreted peptides evolve to bind new targets as bacteria adapt to environmental niches<sup>6–8</sup>.

Libraries of linear peptides can be easily created using combinatorial chemistry or by encoding them as genes using oligonucleotide pools and expressing them in vivo<sup>9</sup>. Sometimes, it is possible to add modifications to a linear peptide that is a natural substrate or emerges from a screen, but this can disrupt affinity<sup>10</sup>. Cells produce modified peptides through large non-ribosomal peptide synthetases (NRPSs) or by expressing them as genes that are post-translationally modified (RiPPs)<sup>11,12</sup>. Many pharmaceuticals have been derived from NRPS products, but the synthetases are notoriously difficult to genetically diversify<sup>13–15</sup>. Therefore, libraries of NRPS analogs have been built using synthetic chemistry, which can be used to diversify the amino acids around a scaffold constrained by cycles and other modifications<sup>16</sup>. For example, rapamycin is produced by an NRPS and a library of 45,000 macrocycles was chemically synthesized to expand the targets beyond mTOR<sup>17</sup>. In contrast, RiPPs are encoded as genes, thus simplifying the creation of libraries of diverse peptides<sup>18</sup>. Once a precursor peptide is expressed, modifying enzymes bind to a leader or follower sequence and modify the core, which is then proteolytically released<sup>11</sup>. Libraries of RiPPs have been generated in heterologous hosts, including *E. coli*, and can be produced using cell-free protein synthesis (CFPS)<sup>19,20</sup>.

Screening methods have been developed to find functional peptides within large libraries. If each peptide variant is purified to high concentration, they can be individually screened using cell-based or biochemical assays<sup>21</sup>. Depending on the assay complexity, up to tens of thousands of variants can be evaluated through automation<sup>22</sup>. Alternatively, the peptides can be pooled and “panned” for those that bind to a protein target. By fusing the peptide to its genetic material (e.g., with mRNA–peptide fusions or phage display), the sequences of binders can be determined. It is more difficult to combine peptide modifications with these techniques, but innovative approaches have been applied to cyclize displayed peptides<sup>23–31</sup>.

Screens have been developed to identify compounds that disrupt protein–protein interactions by binding to their interface<sup>32</sup>. One approach is to fuse each protein to one half of a split reporter (luciferase, GFP, or T7 RNAP), but the disruption of this interaction leads to the loss of signal which is less desirable for screens<sup>33–37</sup>. This is addressed by reverse two-hybrid systems, where the proteins are fused to the DNA-binding domains of repressors, such that the disruption of their interaction leads to the derepression of a promoter<sup>38</sup>. This can be used to drive a selection that links the disruption of the interaction with the survival of the cell, thus allowing orders of magnitude more variants to be evaluated in an experiment. This has been applied to finding cyclized peptides and RiPPs that are  $\mu\text{M}$  inhibitors of

the Gag–TSG1010 and p6–UEV interactions, respectively, that are critical for the human immunodeficiency virus (HIV) life cycle<sup>39,40</sup>.

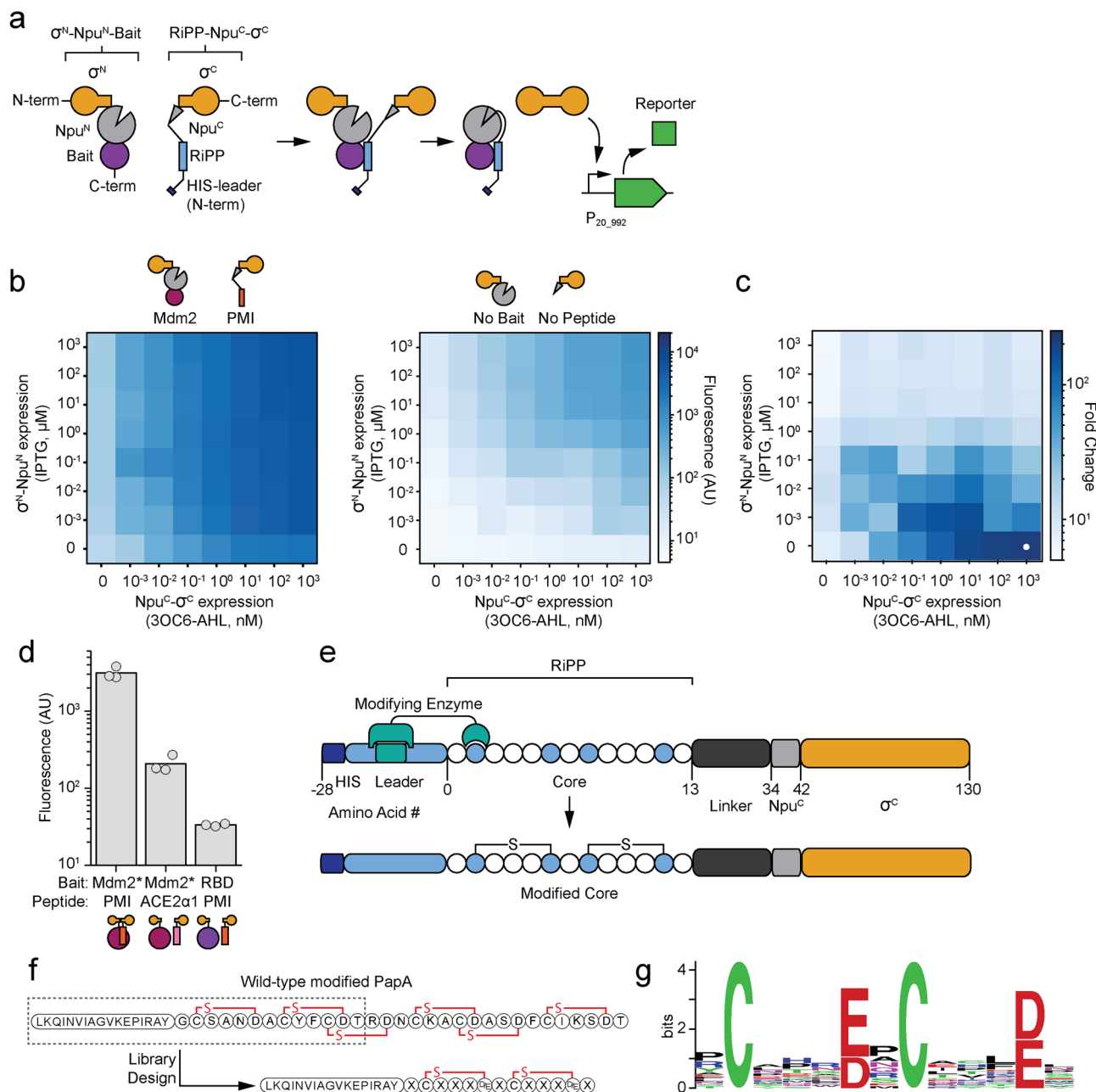
For many applications, it is desirable to identify a peptide that binds to a single target. For example, this could be to detect a protein as a diagnostic<sup>41,42</sup> or to degrade a therapeutic target by directing it to cellular proteolytic machinery<sup>43,44</sup>. Here, we present a selection strategy to identify RiPPs that bind to a single target-of-interest, without specifying its binding position. A synthetic genetic circuit was constructed in *E. coli* that responds when a modified peptide binds to a bait protein. A RiPP, including leader and core sequence, was fused to one half of a split intein and split  $\sigma$  factor and the target protein was fused to the other half of the intein and  $\sigma$  factor. When the RiPP binds the target, the intein covalently releases the complete  $\sigma$  factor, which turns on a promoter to drive a selectable marker. This approach was demonstrated by identifying a synthetic RiPP constrained by a thioether macrocycle that binds to the Spike receptor binding domain (RBD) from the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2)<sup>45</sup>. Rounds of selection were performed to identify a hit, from which a 14 amino acid modified sequence is derived that binds to a unique site that is outside of the regions bound by the native human substrate (ACE2) and neutralizing antibodies ( $K_D = 990 \pm 5 \text{ nM}$ ). This demonstrates the design of cells to simultaneously make peptidic small molecules and evaluate their ability to bind to a defined therapeutic target, thus allowing billions of compounds to be tested in parallel in single experiments.

## Results

### Genetic circuit to detect peptide binding to a target protein.

The circuit converts a binding event into a transcriptional response. It is based on two fusion proteins containing the RiPP and the bait that, upon binding, bring together two halves of a split intein that catalyze the release of a  $\sigma$  factor that recruits RNA polymerase (RNAP) to a promoter (Fig. 1a). We selected a version of the split intein from *Nostoc punctiforme* PCC73102 (Npu) that has been artificially fused and re-split at a different position to produce large N-terminal and short C-terminal fragments. This version has much less activity compared to the native intein’s exceptional stability and splicing kinetics<sup>46,47</sup>. The reduced association between fragments facilitates their use as a protein–protein interaction detector. The  $\sigma$  factor ECF20\_992 was chosen because of its strong transcriptional output and sensitivity to linker variations between the DNA binding domains<sup>48,49</sup>. The first fusion protein consists of the N-terminal  $\sigma$  factor fragment ( $\sigma^N$ ) placed upstream of the N-terminal Npu fragment (Npu<sup>N</sup>) and linker followed by the bait protein. The second fusion protein has the C-terminal Npu fragment (Npu<sup>C</sup>), followed by the C-terminal  $\sigma$  factor fragment ( $\sigma^C$ ). The N-terminus of this protein has the leader peptide that recruits modifying enzymes and the core sequence of the RiPP followed by a flexible 20 aa linker (Methods). Successful splicing of full length  $\sigma$  factor leads to the activation of the P<sub>20\_992</sub> promoter, thereby turning on the downstream genes.

To develop the circuit, we selected the well-studied interaction between the proteins p53 and Mdm2 as a test case<sup>50</sup>. Specifically, we used residues 17–124 of Mdm2 (Mdm2\*) and a high affinity ( $K_D = 0.5 \text{ nM}$ ) variant of residues 15–29 of p53 (PMI)<sup>51,52</sup> as bait and peptide fusions to the  $\sigma^N$ -Npu<sup>N</sup> and Npu<sup>C</sup>- $\sigma^C$  fragments, respectively. The two genes were placed under the control of 3OC6-AHL-inducible P<sub>LuxB</sub> and IPTG-inducible P<sub>Tac</sub> promoters in *E. coli* Marionette Clo<sup>53</sup>. For characterization, *gfp* was cloned downstream of P<sub>20\_992</sub>. The expression of the two fusion proteins was controlled using 3OC6-AHL and IPTG and the



**Fig. 1** Split intein system for in vivo detection of protein–protein interactions. **a** A schematic is shown for the detection of a binding event between a RiPP and target. A split intein from *Nostoc punctiforme* PCC73102 (Npu), is used to splice together a  $\sigma$  factor for downstream transcription. **b** Inducible interaction-mediated splicing. The median fluorescence is shown as a function of the expression of the two halves of the sensor proteins. The induction of PMI- and Mdm2\*-driven association (left) or split intein alone (right) are shown. The plots are a representative replicate selected from three performed on different days. **c** The specificity is calculated using the data in part b: (Fluorescence<sup>Mdm2\*+PMI</sup>/Fluorescence<sup>no bait+no peptide</sup>). The white dot marks the highest fold-change in expression. **d** The fluorescence measured from the circuit containing a binding pair (Mdm2\*:PMI) and non-binding pairs are compared. The 3OC6-AHL concentration for all inductions was 1  $\mu$ M. Three replicates performed on different days are shown. **e** A schematic of the RiPP-containing half of the split intein system is shown, with the modified core residues in blue. **f** The structure of the wild-type PapA modified peptide<sup>56</sup> is shown with a box around the region used to design the peptide library and the resulting scaffold sequence. **g** Library sequence weblogo for the 14 modified library variants observed (Methods).

fluorescence output of the circuit was measured using flow cytometry (Methods). When the interaction was driven by Mdm2\* and PMI, there was a large induction over a wide range of expression levels (Fig. 1b). However, when there was no driving protein–protein interaction, the split inteins and split  $\sigma$  factors only turned on the output promoters at high levels of expression. When performing a selection, it is important that the background expression of the  $\sigma^N$ -Npu<sup>N</sup> and Npu<sup>C</sup>- $\sigma^C$  fragments

not induce the circuit. The difference between the presence and absence of a peptide:bait interaction was maximized to >200-fold when Npu<sup>C</sup>- $\sigma^C$  was fully induced and  $\sigma^N$ -Npu<sup>N</sup> fragment uninduced (Fig. 1c). To remove the need to add inducer,  $\sigma^N$ -Npu<sup>N</sup>-bait was placed under the control of a weak constitutive promoter. The peptide-Npu<sup>C</sup>- $\sigma^C$  was kept under 3OC6-AHL control so that its concentration can be decreased for progressive rounds of selection to bias towards higher-affinity binders.

The inducible range of the sensor was then determined when either the bait or peptide were changed to disrupt the interaction (Fig. 1d). When a peptide based on the N-terminal residues 19–56 of ACE2 (ACE2a1)<sup>54</sup> was used, which does not bind to the Mdm2\* target, the fluorescence output of the circuit drops 15-fold. Similarly, when the target peptide was swapped to be residues 328–533 of the SARS-CoV-2 Spike protein (RBD)<sup>45</sup>, to which PMI does not bind, the fluorescence output of the circuit was almost two orders-of-magnitude lower. Using published affinity values for PMI variants<sup>52</sup>, we estimated that the threshold for detection of this system is 5  $\mu$ M, with a linear increase in fluorescence to 10 nM. Outside of this range, the circuit cannot distinguish variants (Supplementary Figure 1).

The peptide needs to be able to be modified by RiPP enzymes in the context of its fusion to C-terminal Npu<sup>C</sup>- $\sigma^C$  (Fig. 1e). RiPP modifying enzymes bind to an N-terminal leader sequence and then modify the core sequence. This process is not typically influenced by the addition of C-terminal fusions, including entire proteins<sup>24,26,55</sup>. RiPP cores are released from the leader by proteases<sup>11</sup>, which were not included in our system. To improve the accessibility of the core, we added a 20 amino acid linker (GGKGGPGGRGGVGGGGGIGG) between it and the split intein.

We performed a preliminary experiment to test whether a RiPP enzyme could modify a large fraction of core sequences in a library in the context of the fusion protein. The PapA/PapB peptide/enzyme pair from the *Paenibacillus polymyxa* freyrasin biosynthetic gene cluster was selected<sup>56</sup>. PapB introduces a thioether macrocycle between core C and D/E residues and it was shown to be tolerant to amino acid diversity at the unmodified residues<sup>56,57</sup>. Based on this system, we designed a simplified core and leader peptide containing two cycles (Fig. 1f). A library was constructed allowing full (NNK) degeneracy at 9 unconserved amino acid positions and D/E at the two macrocyclized positions, resulting in 10<sup>12</sup> theoretical peptides. We selected 19 random members from this library, expressed them with PapB, and evaluated cyclization by measuring the mass shift observed using liquid chromatography-mass spectrometry (LC-MS) (Methods). From this subset, 16 complete peptides were observed and 14 had masses consistent with either one or two macrocycles (Supplementary Table 4). The modification did not bias the core positions towards specific amino acids (Fig. 1g).

**Complete selection system.** The genetic system used for the selections, involving seven genes, is shown in Fig. 2a. It has been previously demonstrated that the SARS-CoV-2 Spike RBD can be expressed in *E. coli*<sup>58,59</sup>. This domain was used as part of  $\sigma^N$ -Npu<sup>N</sup>-RBD, which is placed under the control of the weak BBa\_J23105 constitutive promoter. The DNA encoding the peptide library (*ripp*) was used to build a fusion library (*ripp-npu<sup>C</sup>- $\sigma^C$* ), the expression of which was controlled with 3OC6-AHL. The modifying enzyme was placed under the control of the cumate-inducible promoter. When the  $\sigma$  factor is released, the P<sub>20-992</sub> promoter drives the expression of an operon containing the selectable marker fusion *gfp-cat*. This enabled positive selection through the addition of chloramphenicol (Cm) to the media. The operon also encodes *pheS* because it can be used in negative selections<sup>60</sup>, but it was not used in this study.

**Library design and selection.** The selection scheme is shown in Fig. 2b. Each cell makes a different variant of the peptide. When it binds to the target, *gfp-cat* is expressed. Over multiple rounds of selection, increased stringency was applied by increasing the concentration of Cm or decreasing peptide induction (by decreasing the concentration of 3OC6-AHL) (Fig. 2c).

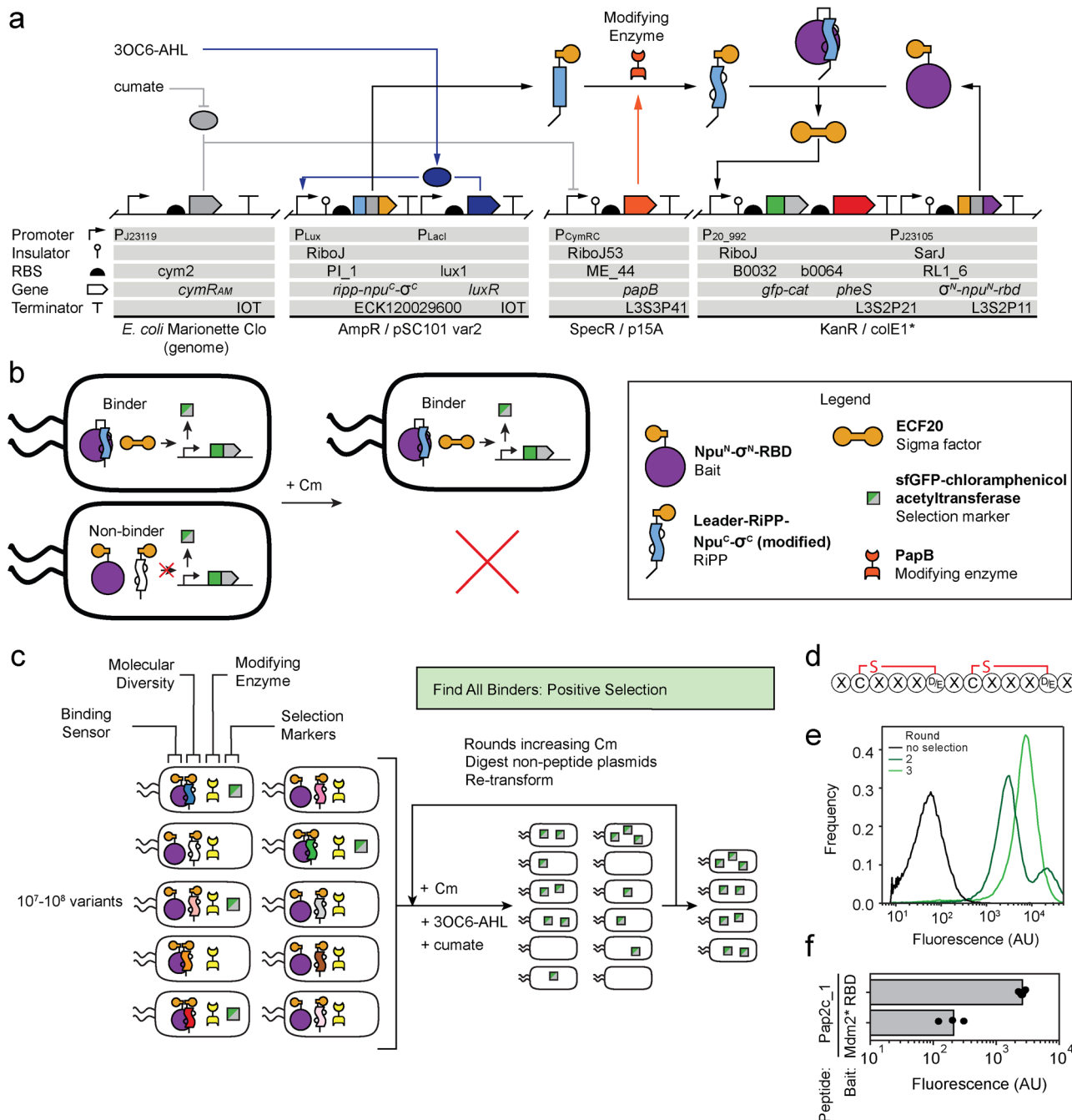
The library was based on a simplified PapB-modified 13 amino acid core structure (Fig. 2d). The positions not modified by PapB were diversified using NNK codons, leading to a theoretical library size of 10<sup>12</sup> variants. When the library was introduced to cells using electrocompetence, it was limited in practice to 10<sup>8</sup> per transformation. Rounds of positive selection were performed; after each round, the plasmids from the surviving cells were isolated and retransformed (Fig. 2c) (Methods). Cells were grown overnight in increasing concentrations of Cm: Round 1 (300  $\mu$ M), Round 2 (800  $\mu$ M), and Round 3 (1200  $\mu$ M). GFP expression was measured after each round using flow cytometry, showing a continuous increase in the fluorescence output (Fig. 2e).

When using 300  $\mu$ M Cm for the initial selection, two peaks were observed: one lower (2,000 AU) and one higher (20,000 AU). This higher peak was attributed to escape mutants from the selection plasmid breaking. To eliminate escapes from round to round, we digested the selection plasmid and retransformed enriched peptide plasmids into the expression strain. This eliminated the peak corresponding to escapes (Fig. 2e). All selection rounds were then analyzed using next-generation sequencing (NGS). The number of unique RiPP sequences decreased after each round, indicating enrichment: 139,320, 88,229, and 63,344. The abundance of each sequence was calculated and 32 were found to represent >1% of the population after Round 3. These were further reduced to 20 by only considering those that showed consistent enrichment between rounds 1→2 and 2→3.

Next-generation sequencing (NGS) was used to analyze the product of each selection, rather than physically isolating the top variants (Methods). Because of this, we had to re-obtain the top 20 hits using DNA synthesis. Each was cloned into the *ripp-npu<sup>C</sup>- $\sigma^C$*  plasmid and re-assessed for activity. This process eliminates any cheaters that may have arisen throughout the selection process due to mutations to the plasmid or genome. These constructs were transformed into strains containing the modifying enzyme and either Spike RBD or Mdm2\* as bait, with the latter selected to measure off-target binding. The fluorescence output of the circuit was measured using flow cytometry under the same growth conditions and inducer concentrations used for the selections. From these, the core VCKYGEWCEIVEI encoded within Pap2c\_1 demonstrated a strong fluorescence output and 14-fold specificity for the Spike RBD as compared to Mdm2\* (Fig. 2f).

Variants of Pap2c\_1 were constructed and the impact on binding to Spike RBD evaluated using the fluorescent reporter and flow cytometry (Methods). Any truncations from the N-terminus of the core sequence led to lower fluorescence (Fig. 3a, Supplementary Fig. 2). However, even as few as six amino acids still showed a comparable binding profile (up to 3-fold higher fluorescence output of the circuit) than the off-target control PMI (dashed lines in Fig. 3a and data in Fig. 1d). These truncations lead to an affinity equivalent to ACE2a1 binding to Spike RBD, which has been reported to bind with low  $\mu$ M affinity<sup>54</sup>.

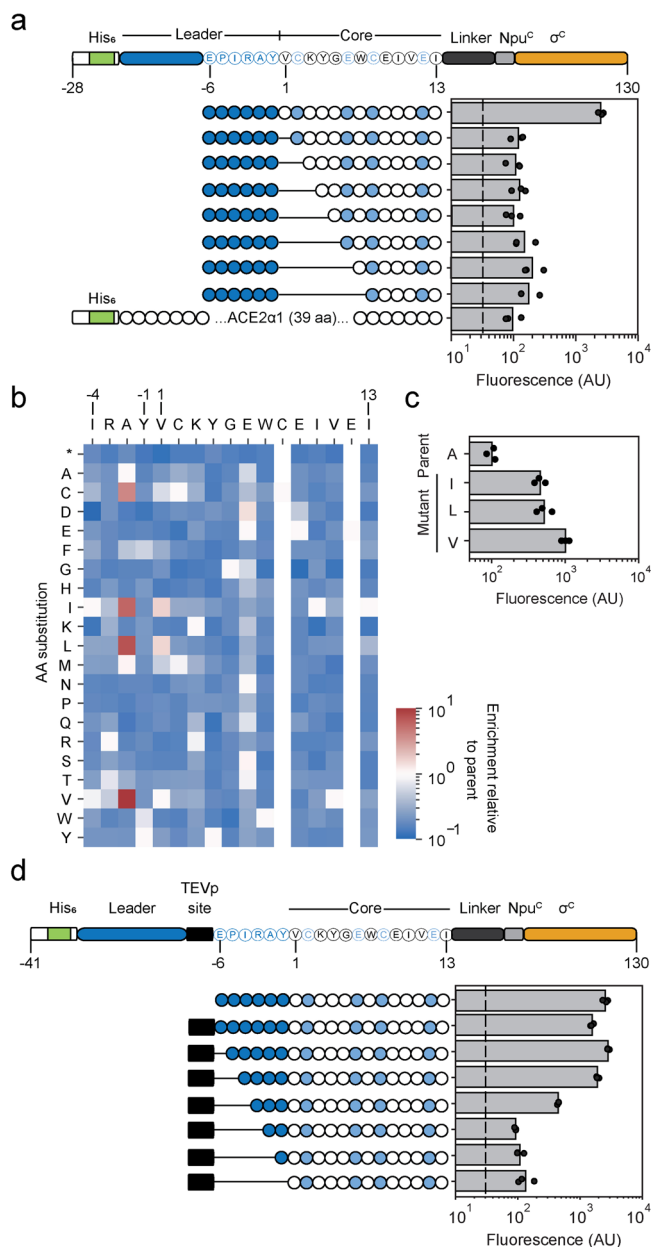
Next, single-site saturation mutagenesis (NNK) was conducted for each residue of the core. The 15 libraries, each corresponding to a different residue, were pooled and enriched for binding using the selection (250  $\mu$ M Cm) while inducing Pap2c\_1 (3.2 nM OC6-AHL). The selected and unselected libraries were analyzed by NGS and the enrichment of each amino acid substitution calculated (Methods) (Fig. 3b). The majority of single amino acid substitutions led to lower affinity for RBD, indicating that a near-optimal sequence was identified in the selection (Fig. 3b). Substitutions at residue 3 from alanine to cysteine or branched hydrophobic amino acids (I, L, V) led to greater enrichment. These also led to an increase in fluorescence when measured by cytometry (Fig. 3c).



**Fig. 2 Selection system to find RBD-binding RiPPs.** **a** The genetic circuit diagram for the RBD-binding RiPP selection system is distributed across three plasmids and two genomic regions. Two small molecules: 3OC6-AHL and cumate are used to control the expression of the RiPP peptide and modifying enzyme, respectively. **b** Schematic of selection output under two conditions: with an RBD-binding RiPP and without an RBD-binding RiPP. Binding is shown to result in the production of GFP-chloramphenicol acetyltransferase (GFP-CAT; green/gray squares). **c** Overview of the positive selection. Selection rounds were conducted in the presence of the RiPP peptide and modifying enzymes and used increasing Cm concentrations for increased stringency. **d** The core scaffold for the Pap2c library (IbAMK-103). Predicted macrocycles are colored between constrained residues and “X” residues correspond to NNK translated amino acids. **e** Cytometry distributions for positive selections on the Pap2c library beginning with no selection, round 2, and round 3 of positive selections (0, 800, and 1200 μM Cm). Fluorescence of the GFP fused to CAT is reported. **f** RBD-specific hit (Pap2c<sub>1</sub>) isolated from genetic selection. Pap2c<sub>1</sub> was used as peptide against the non-specific bait (Mdm2\*) and specific bait (RBD). 3OC6-AHL (1 μM) was used for inductions and three replicates performed on different days are shown.

Next, we tested how the insertion of a TEV protease site (ENLYFQG) between the core and leader impacts affinity (Fig. 3d, Supplementary Figure 2). This leads to 10-fold lower fluorescence that can be recovered by moving the cleavage site four amino acids (IRAY) upstream of the core. The lower affinity may be due to the disruption of an α-helix (Supplementary Figure 3).

**AMK-1057 binding to Spike RBD.** To aid with downstream purification of peptides that emerge from the selection, the leader and core were fused to the N-terminal RiPP stabilization tag, RST<sub>N</sub>. TEV protease cleavage sites (ENLYFQG) are placed between RST<sub>N</sub> and the leader and between the leader and the core (Fig. 4a). The resulting fusion protein is referred to as



**Fig. 3** Sequence-activity profiling of Pap2c<sub>1</sub> yields better binders in vivo.

**a** Schematic of Pap2c<sub>1</sub> peptide fusion used for in vivo measurements. Leader and core amino acids considered for profiling are displayed and predicted modified core residues highlighted in light blue. A fusion protein of the 39 amino acid ACE2α1 peptide to the C-terminal NpuC-σC was included as an off-target control. Dashed lines indicate the fluorescence output of the circuit driven by the RBD:PMI interaction measured in Fig. 1d. 3OC6-AHL (1 μM) was used for inductions and three replicates performed on different days are shown. **b** Heat map of variant enrichment from selection and NGS. Amino acid substitutions enriched relative to the parent sequence are colored red. **c** Fluorescence output of consensus variant strains constructed based on data from (b). Individually constructed variants at position -3 (A, I, L, V) were evaluated for fluorescence output of the circuit. 3OC6-AHL (0 μM) was used for inductions and three replicates performed on different days are shown. **d** Schematic of the Pap2c<sub>1</sub> peptide fusion used for in vivo measurements and TEV protease site insertion analysis. Modified core residues are shown in light blue. Dashed lines indicate the fluorescence output of the circuit driven by the RBD:PMI interaction measured in Fig. 1d. 3OC6-AHL (1 μM) was used for inductions and three replicates performed on different days are shown.

RST<sub>N</sub>-Pap2c<sub>1</sub>. Note that when TEV protease is used to release the core, it leaves an additional glycine on its N-terminus, which did not emerge from the selection.

We expressed RST<sub>N</sub>-Pap2c<sub>1</sub> and purified the modified core peptide in order to determine the location of its binding to cell-derived Spike RBD. The cleaved, modified peptide is referred to as AMK-1057. RST<sub>N</sub>-Pap2c<sub>1</sub> was co-expressed at liter scale with the modifying enzyme, Ni-NTA purified, and cleaved using TEV protease (Methods). The cleaved peptide was isolated using solid phase extraction (SPE) and semiprep HPLC purification. Three peptides were isolated: leader (200 μg/L), unmodified core (640 μg/L) and AMK-1057 (360 μg/L) (Fig. 4b).

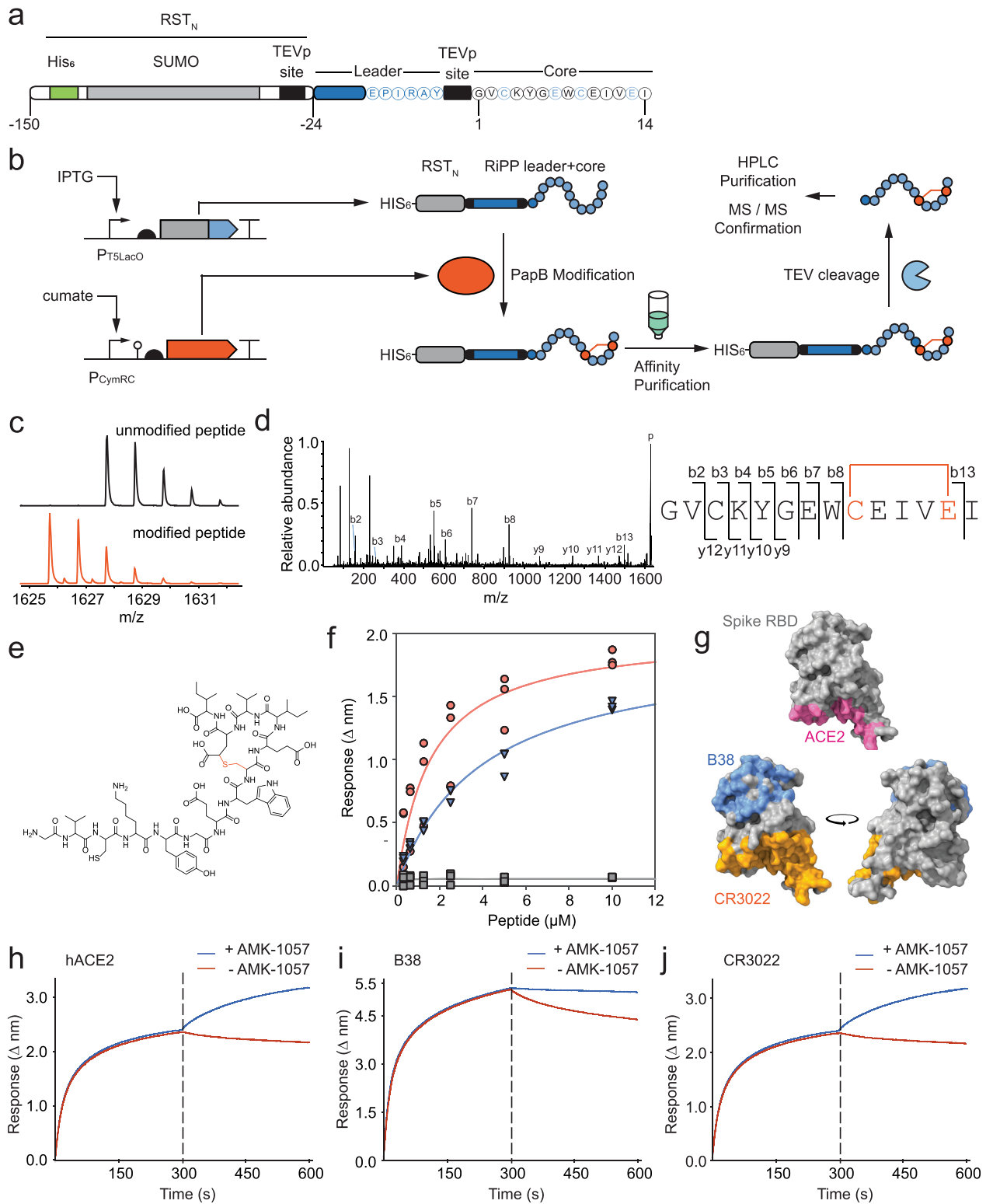
High-resolution LC-MS analysis of both modified (expected *m/z*: 1625.7338; observed *m/z*: 1625.7332) and unmodified (expected *m/z*: 1627.7494; observed *m/z*: 1627.7484) peptide showed a mass shift corresponding to formation of a single cycle, despite the library being based on a two-cycle scaffold (Fig. 4c). The macrocycle found in AMK-1057 is predicted to be formed through the covalent linkage of a side chain cysteine sulfur atom to the Cβ on the downstream glutamate residue, which is stable to standard collision-induced dissociation conditions<sup>56</sup>. This property was used to annotate the macrocycle placement using high-resolution tandem MS (HR-MS/MS) and hypothetical structure enumeration and evaluation (Fig. 4d)<sup>61</sup>. Fragmentation analysis indicated that the macrocycle forms at the C-terminal end of AMK-1057, between C9 and E13 (Fig. 4e).

We then performed in vitro binding experiments using Expi293F human cell-derived and purified Spike RBD. Bio-layer interferometry (BLI) was used to measure the affinity of AMK-1057 to Spike RBD as  $990 \pm 5$  nM (Fig. 4f; Supplementary Fig. 2) (Methods). AMK-1057 is in the higher range of natural RiPPs binding to their target (e.g., lassomycin at 0.41 μM, microcin J25 at 2 μM) and some peptidic drugs (e.g., vancomycin at ~1 μM)<sup>62–64</sup>. Neither the purified unmodified core peptide nor the leader sequence bound to the target. We purified the ACE2α1 peptide and found that it bound to the Spike RBD with a  $K_D$  of  $647 \pm 9$  nM, which is similar to the reported value of 1.3 μM<sup>54</sup> (Fig. 4f, Supplementary Fig. 2).

To define a more specific region of the Spike RBD that AMK-1057 targets, BLI was used to measure binding in the presence of competitors with known epitope binding profiles<sup>65</sup>. Recombinant, human-expressed ACE2 (hACE2;  $K_D = 44.2$  nM)<sup>66</sup> and candidate therapeutic antibodies B38 ( $K_D = 70.1$  nM)<sup>67</sup> and CR3022 ( $K_D = 6$  nM)<sup>68</sup> bind to different regions of Spike RBD, visualized in Fig. 4g. BLI traces show association (300 seconds, pre-dashed line) of either hACE2, B38, or CR3022 with target RBD (Fig. 4h, i, j). After reaching steady-state, the biosensor containing bound complex was added to either control wells or AMK-1057 and an increase in Δnm observed as a measurement of target binding. In all three experiments, AMK-1057 associated with the RBD, indicating that hACE2, B38, and CR3022 do not occlude binding. Therefore, the binding epitope for AMK-1057 must be distinct from the regions highlighted in Fig. 4g (shown in gray).

## Discussion

This work introduces a technique to capture modified peptides that bind to a single target protein without specifying the binding location, which can be valuable for “undruggable” targets lacking an obvious binding pocket. This is in contrast to two-hybrid screens, where it is the disruption of a protein–protein interaction or a functional screen that requires blocking an enzymatic activity. Our approach could be useful in finding binders that disrupt a protein–protein interaction when one of the participating proteins cannot be expressed in functional form in a recombinant host or when binding to one of the two targets is



preferred. However, there is no guarantee that the peptide that emerges from the selection will bind to a therapeutic site at the interface. Indeed, while our original intent was to discover peptides that block Spike binding to ACE2, the peptide appears to bind in a region distinct from ACE2, and even outside of the regions bound to by therapeutic antibodies. Still, there are therapeutic modalities that do not require binding to a specific position, such as PROTACs (PROteolysis Targeting Chimeras)

that make use of a weak (as high as 10  $\mu$ M) peptide binding-moiety to bring E3 ubiquitin ligase to the protein target<sup>41–43,69–71</sup>. The modified peptides could also be applicable to diagnostics, for example, when conjugated to gold nanoparticles they have been explored as diagnostics in point-of-care devices used to detect viral epitopes in HIV<sup>72</sup>.

The genetic circuit is based on a split intein and  $\sigma$  factor to which the peptide library is fused. This could be used to screen/select

**Fig. 4** AMK-1057 is a cyclic peptide that binds human-derived Spike RBD *in vitro*. **a** Schematic of the construct used to express AMK-1057. Leader and core amino acids are shown and predicted modified core residues highlighted in light blue. TEV protease sites are annotated as TEVp sites. **b** The expression, modification, cleavage, and purification steps are shown. TEV protease cleavage removes the SUMO tag/leader peptide and HPLC purification produces the final product; it leaves a single G, shown as a darker amino acid in the core. **c** High-resolution MS of unmodified AMK-1057 (black trace) and singly modified AMK-1057 (orange trace). **d** High-resolution MS/MS of modified AMK-1057 and fragment mapping to the amino acid sequence. Numbered peaks correspond to fragment ions observed and represented as lettered amino acids next to the MS/MS spectrum. **e** Structural annotation of AMK-1057. Thioether modification is colored to highlight the macrocycle. See Methods for details. **f** Binding of purified peptides (modified AMK-1057, blue; unmodified AMK-1057, gray; ACE2 $\alpha$ 1, orange) to human cell line-derived Spike RBD<sub>296-531</sub>. Three replicates performed on different days are shown. **g** Space-filling models of the Spike RBD (PDB ID: 6M17) colored to indicate annotated binding regions of ACE2 (pink)<sup>66</sup>, as well as antibodies CR3022 (orange)<sup>67</sup> and B38 (blue)<sup>68</sup>. Images were generated using ChimeraX<sup>82</sup>. The model is rotated 180°. **h–j** Competition BLI experiments measuring binding of purified, modified AMK-1057 to human cell-derived Spike RBD pre-incubated with **h** soluble hACE2 ( $\Delta\Delta\text{nm} = 0.99 \pm 0.02$ ), **i** antibody B38 ( $\Delta\Delta\text{nm} = 0.81 \pm 0.03$ ), and **j** antibody CR3022 ( $\Delta\Delta\text{nm} = 0.71 \pm 0.03$ ).  $\Delta\text{nm}$  was measured at 600 s and normalized to no AMK-1057 control to account for competitor off-rates ( $\Delta\Delta\text{nm}$ ). Dashed lines separate association reactions from dissociation reactions (Methods). For all experiments, three replicates were performed on different days and a representative trace is shown.

libraries of linear peptides or protein domains for binding to the target. Here, we decided to use modified peptides because they have advantages as pharmaceuticals, including more compact structures that are resistant to proteolysis and permeate the mammalian cell membrane<sup>1</sup>. Modified peptides have been approved for the treatment of cancer, inflammation, and infection and increasing numbers are entering all phases of clinical development for diverse indications<sup>73–76</sup>. For example, the FDA-approved HIV antiviral Enfuvirtide is a 36 amino acid (aa) linear peptide that binds to a transmembrane glycoprotein; however, it suffers from rapid proteolysis, thus requiring twice daily injections<sup>77</sup>. Crosslinking HIV-1 mimetic peptides makes them proteolytically-stable, acid-resistant, and orally bioavailable<sup>78</sup>. Further, RiPPs are a large class of natural products and the rules to diversify natural RiPPs or combine modifications to create synthetic ones are growing<sup>26,40,79</sup>. Unlike NRPS pathways, RiPPs are simple to fuse to the intein/ $\sigma$  factor domains and to create a large and diverse library of modified peptides. Combining molecular diversity creation with a selection circuit in the same cell enables massive libraries to be evaluated to populate pharmaceutical discovery pipelines with binders to a target-of-interest with minimal biochemical information.

## Methods

**Strains, media, and chemicals.** *E. coli* NEB 10-beta (C3019I, New England BioLabs, MA) was used for all routine cloning. *E. coli* Marionette-Clo<sup>53</sup> was used for all selection experiments. *E. coli* Marionette-X, a Marionette-compatible derivative of NEB Express (C2523I, New England BioLabs, MA) was used for large-scale peptide expression experiments. TB media (T0311, Teknova, CA) supplemented with 0.4% glycerol (BDH1172-4LP, VWR, OH, USA) was used for peptide expression and modification. 2xYT liquid media (B244020, BD, NJ) and 2xYT + 2% agar (B214010, BD, NJ) plates were used for routine cloning and strain maintenance. SOB liquid media (S0210, Teknova, CA) was used for making competent cells. SOC liquid media (B9020S, New England BioLabs, MA) was used for outgrowth. Unless noted otherwise, cells were induced with the following chemicals: cuminic acid (268402, Millipore Sigma, MO) added as 1000 $\times$  stock (200 mM) in EtOH or DMSO; 3-oxohexanoyl-homoserine lactone (3OC6-AHL) (K3007, Millipore Sigma, MO) added as a 1000 $\times$  stock (1 mM) in DMSO; anhydrotetracycline (aTc) (37919, Millipore Sigma, MO) added as a 1000 $\times$  stock (100  $\mu$ M) in DMSO; isopropyl  $\beta$ -D-1-thiogalactopyranoside (IPTG) (I2481C, Gold Biotechnology, MO) added as 1000 $\times$  stock (1 M) in water. Cells were selected with the following antibiotics: carbenicillin (Carb, C-103-5, Gold Biotechnology, MO) added as 1000 $\times$  stock (100 mg/mL in H<sub>2</sub>O); kanamycin (Kan, K-120-10, Gold Biotechnology, MO) as 1000 $\times$  stock (50 mg/mL in H<sub>2</sub>O); spectinomycin (Spec, S-140-5, Gold Biotechnology, MO); and chloramphenicol (Cm, C-105-5, Gold Biotechnology, MO). Liquid chromatography was performed with Optima Acetonitrile (A996-4, Thermo Fisher Scientific, MA) and water (Milli-Q Advantage A10, Millipore Sigma, MO) supplemented with LC-MS Grade Formic Acid (85178, Thermo Fisher Scientific). The following solvents/chemicals were also used: Ethanol (V1001, Decon Labs, PA), Methanol (3016-16, Avantor, PA), Ammonium bicarbonate (A6141 Millipore Sigma, MO), dimethyl sulfoxide (DMSO) (32434, Alfa Aesar, MA), Imidazole (IX0005, Millipore Sigma, MO), sodium chloride (X190, VWR, OH), sodium phosphate monobasic monohydrate (20233, USB Corporation, OH), sodium phosphate dibasic anhydrous (204855000, NJ), guanidine hydrochloride (50950, Millipore Sigma, MO), tris (75825, Affymetrix, OH), TCEP (51805-45-9, Gold Biotechnology, MO), and EDTA (0.5 M stock, 15694,

USB Corporation, OH). DNA oligos and gBlocks were ordered from Integrated DNA Technologies (IDT) (CA).

**Plasmids and genes.** Plasmids pTHSS-1282, pDAA558, and pAMK-267 were constructed from the parental pTHSSe-44 backbone, which has a pSC101 origin variant (var 2) and ampicillin resistance<sup>80</sup>. Plasmids pTHSS-1282, pDAA-558, and pAMK-267 contain a flexible linker sequence (GSSRGGKGGPGGRRGGVGGGGGIGG) between the peptide/GFP and Npu<sup>C</sup> regions. Plasmids pAMK-925, pTHSS-2132, pAMK-866, and pAMK-870 were constructed from the parental pTHSS-1458 backbone, which has a colE1\* origin variant and a kanamycin resistance marker<sup>80</sup>. The plasmid pEG06\_044 contains the PapB enzyme, a p15A origin of replication and spectinomycin resistance. The parental backbone pTHSS-2012, which has a p15a origin and spectinomycin resistance was used for additional cloning experiments<sup>80</sup>. The plasmids pAMK-926 and pTHSS-2137 that contain the P<sub>Lux</sub> promoter expressing Npu<sup>C</sup>- $\sigma^C$  and PMI-Npu<sup>C</sup>- $\sigma^C$ , respectively, were constructed from pTHSS-2012. The plasmids pAMK-925 and pTHSS-2132 that contain the P<sub>Tac</sub> promoter expressing  $\sigma^N$ -Npu<sup>N</sup> and residues 17–124 of Mdm2 (Mdm2\*<sup>-</sup>- $\sigma^N$ -Npu<sup>N</sup>), respectively, were constructed from pTHSS-1458. The plasmid pAMK-870 that contains the constitutive P<sub>J23105</sub> promoter expressing Mdm2\*<sup>-</sup>- $\sigma^N$ -Npu<sup>N</sup> and the P<sub>20\_992</sub> promoter expressing CAT-GFP was constructed from pTHSS-1458. The plasmid pAMK-866 that contains the constitutive P<sub>J23105</sub> promoter expressing 328–533 of the SARS-CoV-2 Spike protein (RBD)- $\sigma^N$ -Npu<sup>N</sup> and the P<sub>20\_992</sub> promoter expressing CAT-GFP was constructed from pTHSS-1458. The peptide cloning plasmid pAMK-267, constructed from pTHSSe-44, contains the P<sub>Lux</sub> promoter upstream of an RBS-His tag-SapI-GFP-SapI-Npu<sup>C</sup>- $\sigma^C$  where the *gfp* gene can be replaced by a peptide gene through Type IIs assembly methods using the enzyme SapI (NEB). The RBS from pAMK-267 was chosen from a library of RBS variants upstream of a His tag-PMI-Npu<sup>C</sup>- $\sigma^C$  that was tuned for co-expression with constructs containing the P<sub>J23105</sub> promoter expressing Mdm2\*<sup>-</sup>- $\sigma^N$ -Npu<sup>N</sup>. The N-terminal His tag in pAMK-267 was left in place to provide a constant 11 aa for consistent levels of expression between different peptide sequences. The plasmid pAMK-670 that contains the P<sub>Lux</sub> promoter expressing PMI-Npu<sup>C</sup>- $\sigma^C$  was constructed from pAMK-267. The plasmid pAMK-857 that contains the P<sub>Lux</sub> promoter expressing N-terminal residues 19–56 of ACE2 (ACE2 $\alpha$ 1)-Npu<sup>C</sup>- $\sigma^C$  was constructed from pAMK-267. The plasmid pAMK-917 that contains the P<sub>Lux</sub> promoter expressing the hit sequence fused to Npu<sup>C</sup>- $\sigma^C$  was constructed from pAMK-267. Note that the pTHSSe-44 and pTHSS-1458 backbones have origin variants that alter their copy numbers, making them approximately equivalent to a p15a backbone. Genes encoding Npu intein, PMI, Mdm2\*, ACE2 $\alpha$ 1, and RBD were synthesized as gBlocks. The ECF20\_992 gene was sourced from a previous publication<sup>48</sup>.

**Cytometry analysis.** Fluorescence characterization was performed on a BD LSR Fortessa flow cytometer with the HTS attachment and BD FACSDiva software version 8.0.3 (BD, NJ). Samples were prepared by diluting overnight cultures 1:400 by adding 0.5  $\mu$ l of cell culture into 200  $\mu$ l of PBS containing 1 mg/mL Kan. Fluorescence measurements were made using the blue (488 nm) laser and all data was derived from the FITC-A channel (PMT voltage of 400 V). At least 30,000 events were collected for each sample, without gating, and the Cytoflow Python package (v 1.1.1) was used for downstream analysis. When presented, the median value is used.

**Evaluation of the split-intein  $\sigma$  factor circuit.** Strains of *E. coli* Marionette Clo harboring a combination of plasmids pTHSS-1282, pTHSS-2132, and pTHSS-2137 or pTHSS-1282, pAMK-925, and pAMK-926 were used for assessing intein splicing with or without PMI-Mdm2\* induced association, respectively. Strains were grown in 1 mL of LB media + antibiotics for 20 h in a deep well 96-well plate (1896–2000, USA Scientific, FL) at 30 °C, 900 rpm in an Infors HT Multitron Pro (Infors USA, MD). Cultures were then diluted 1:100 into fresh 1 mL of LB media + antibiotics and serial 1:10 dilutions of inducers (IPTG, 10<sup>-3</sup>–10<sup>3</sup>  $\mu$ M; 3OC6-AHL, 10<sup>-3</sup>–10<sup>3</sup> nM) for 20 h in a deep well 96-well plate at 30 °C, 900 rpm in the



Multitron Pro. 0.5  $\mu$ l of saturated cell culture were then diluted into 200  $\mu$ l of PBS containing 1 mg/mL Kan for cytometry analysis.

**PMI variant induction profiling.** PMI variants without any N-terminal sequence additions were appended to the linker-Npu<sup>C</sup>- $\sigma^C$  and inserted into the pDAA558 backbone using standard cloning methods. These plasmids were then co-transformed into a *Marionette Clo* strain alongside either pAMK-878 (off-target, Spike RBD) or pAMK-877 (on-target, Mdm2\*) and then grown overnight in 1 mL LB media + Kan/Carb at 30 °C, 900 rpm in a deep well 96-well plate. The next day, cultures were diluted 1:100 into 1 mL TB media + Kan/Carb containing either 0 or 1  $\mu$ M 3OC6-AHL. After overnight growth in the same growth conditions, the variants were assayed as described in Cytometry Analysis.

**Two-hybrid assay for RBD/Mdm2\* association.** To assay for protein–protein mediated splicing the following plasmid combinations were transformed into *E. coli* *Marionette Clo* and fluorescence was measured via cytometry: pAMK-866/pAMK-670 (RBD/PMI); pAMK-866/pAMK-857 (RBD/ ACE2a1); pAMK-870/pAMK-670 (Mdm2\*/PMI); pAMK-870/pAMK-857 (Mdm2\*/ACE2a1). Strains were grown in 1 mL of LB media + antibiotics for 20 h in a deep well 96-well plate at 30 °C, 900 rpm in a Multitron Pro. Cultures were then diluted 1:100 into fresh 1 mL of LB media + antibiotics + 1  $\mu$ M 3OC6-AHL (full induction of peptide plasmid) for 20 h in a deep well 96-well plate at 30 °C, 900 rpm in the Multitron Pro. 0.5  $\mu$ l of saturated cell culture were then diluted into 200  $\mu$ l of PBS containing 1 mg/mL Kan for cytometry analysis.

**Library generation.** The Pap library was designed with diversity at the ends and middle of the peptide and included either glutamate or aspartate as a cyclization partner, for a final sequence design of XCXXX[D/E]XCXXX[D/E]X. Using the degenerate nucleotide sequences NNK for any amino acid and GAW for aspartate or glutamate, we generated a library of 10<sup>12</sup> peptides encoded by 10<sup>14</sup> unique codon sequences. The library of plasmids lbAMK-103, which contains the P<sub>Lux</sub> promoter expressing the Pap library-Npu<sup>C</sup>- $\sigma^C$  was constructed from pAMK-267. The *pap* library was amplified from pEG03\_283 using degenerate oligonucleotides oAMK-915/916 (IDT). Gel purification was used to isolate the 124 bp amplicon, which was then cloned into pAMK-267 using the type IIS restriction enzyme SapI (NEB).

Linear insert and plasmid were mixed at a 1:1 molar ratio (200 fmol each) along with 10  $\mu$ l 10 $\times$  DNA ligase buffer, 2  $\mu$ l T4 DNA ligase (HC) (20 U/ $\mu$ l) (M1794, Promega, WI) and 4  $\mu$ l SapI in 100  $\mu$ l total volume. Reactions were cycled 25 times for 2 min at 37 °C and 5 min at 16 °C then incubated for 30 min at 50 °C, 30 min at 37 °C, and 10 min at 80 °C in a DNA Engine cycler (Bio-Rad, CA). An additional 2  $\mu$ l SapI was then added, and the assembly was incubated for 1 h at 37 °C. Assemblies were then purified using Zymo Spin I columns (Zymo Research, CA). Library assemblies were initially transformed into electrocompetent *E. coli* NEB 10 $\beta$  (C3020K, NEB, MA). We observed 1.5  $\times$  10<sup>7</sup> colony forming units (CFU)/mL for lbAMK-103. Total transformants were estimated by colony counting after 10<sup>7</sup>-fold dilution and plating of liquid outgrowths on selective media.

**Calculation of the modified fraction of the library.** The initial, unselected *papA* library was transformed and plated to resolve individual colonies. A set of 19 random colonies were picked and sequenced via colony PCR. Of the 19 sequenced colonies, 18 were properly assembled. These 18 library members were then assessed for post-translational modification via LC-MS. Plasmids were transformed into either *E. coli* NEB Express or *E. coli* *Marionette-X* using 30  $\mu$ l of competent cells and 1  $\mu$ l of each plasmid being transformed in a PCR strip tubes (1402-4700, USA Scientific, FL or 951020401, Eppendorf, NY). Transformations were incubated on ice (20–30 min), heat shocked (42 °C, 30 s), and incubated on ice again (5 min). Cells were then transferred to a deep well 96-well plate (1896–2000, USA Scientific, FL) with 120  $\mu$ l of SOC media. After outgrowth (Multitron Pro, 1 h, 30 °C) in an Infors HT Multitron Pro (Infors USA, MD), 900  $\mu$ l LB was added with appropriate antibiotics (at 1.1 $\times$  final concentration) and incubated (Multitron Pro, 30 °C, 900 r.p.m.) until all wells reached saturation (12–30 h). Overnight cultures were diluted 1:100 into 1 mL TB media with Carb/Kan in deep well plates. After 3 h incubation (Multitron Pro, 30 °C, 900 r.p.m.), inducer was added (1 mM IPTG, 100  $\mu$ l cumate), and cultures were incubated for 20 h (Multitron Pro, 30 °C, 900 r.p.m.). To purify the peptides, the 96-well plates were centrifuged (Legend XFR, 4,500 g, 4 °C, 20 min), pellets were resuspended in 600  $\mu$ l lysis buffer (5 M guanidinium hydrochloride, 50 mM sodium phosphate, pH 8), and freeze-thawed (frozen at –80 °C; thawed in Multitron Pro at 37 °C, 900 r.p.m.). Cell lysates were centrifuged (Legend XFR, 4,500 g, 4 °C, 60 min) and peptides affinity purified using His MultiTrap TALON plates (29-0005-96, GE Life Sciences (now Cytiva), MA), following manufacturer instructions, using 1  $\times$  600  $\mu$ l water and 2  $\times$  600  $\mu$ l lysis buffer for column equilibration, 4  $\times$  600  $\mu$ l wash buffer (50 mM Tris, 50 mM NaCl, 5 mM Imidazole pH 8), and 1  $\times$  200  $\mu$ l elution buffer (50 mM Tris, 50 mM NaCl, 150 mM Imidazole pH 8). TCEP (2 mM) was added to the peptides, they were digested with TEV protease (0.1 mg/mL) at 4 °C overnight (~16) and the samples were analyzed using the QTOF (see below). The 14 modified library sequences were then aligned and WebLogos generated (<https://weblogo.berkeley.edu/logo.cgi>) with default parameters, except without small sample correction.

**Selection of Pap library lbAMK-103.** Assembled library of plasmids lbAMK-103 was transformed into an electrocompetent *Marionette Clo* strain harboring the PapB modifying enzyme plasmid, pEG06\_044, and the selection plasmid, pAMK-866 (all non-assembly transformation steps were > 1  $\times$  10<sup>8</sup> efficiency). A 1 mL of liquid outgrowth of library transformants was diluted 1:50 in TB media + Carb/Kan/Spec + 1  $\mu$ M 3OC6-AHL and 100  $\mu$ M cumate to induce peptide + modifying enzyme, and grown at 30 °C, 250 r.p.m. for 20 h. For the first round of selection, cultures were then diluted 1:100 in TB Carb/Kan/Spec + 1  $\mu$ M 3OC6-AHL and 100  $\mu$ M cumate + 300  $\mu$ M Cm and grown at 30 °C, 250 r.p.m. for at least 20 h (until cultures were saturated). A 0.5  $\mu$ l aliquot was taken for cytometry analysis and 2 mL of culture was also taken to harvest plasmid. A 5  $\mu$ l sample of purified plasmid was stored for NGS analysis and the rest was digested with 1  $\mu$ l SapI (NEB) for 1 h at 37 °C to remove the background pEG06\_044/pAMK-866 plasmid. The selected lbAMK-103 plasmid was then re-transformed into the strain of electrocompetent *E. coli* *Marionette Clo* strain harboring the PapB modifying enzyme, pEG06\_046, and the selection plasmid, pAMK-866. The selection process was repeated once more with a Cm concentration of 800  $\mu$ M and then once more with a Cm concentration of 1200  $\mu$ M.

**NGS analysis.** Library construction for NGS was performed using the protocol for KAPA Hyper Prep Kits with PCR Library Amplification/Illumina series (KK8504, Roche). First, miniprep library plasmids were amplified with Q5 polymerase (#M0492L, New England BioLabs, MA) with the primers oAMK-946/947 (Pap library) and oAMK-997/998 (Tgn/Lyn library). A 150 bp band was isolated via gel extraction. Indexed adapters were ligated and reamplified with 10 cycles of PCR. Gel extraction was then used to isolate the resultant 260 bp PCR product. Sample concentrations were calculated using a BioAnalyzer on a High Sense DNA chip (5067-4626, Agilent). Samples were diluted to 2 nM, denatured, and further diluted to 10 pM, with a 10% phiX spike in. Samples were run on a HiSeq 2500 using HiSeq v2 reagents for Paired End Clustering and a 200 cycle SBS kit (PE-402-4002 and FC-402-4021, Illumina) using the Illumina HCS software version 2.2.68. Forward and reverse reads were both 110 cycles, with an 8 cycle single index read. Base-calling and demultiplexing were performed using the bcl2fastq software (Illumina) with default settings. After basecalling and indexing, sequences were identified and aligned using the leader sequence and then binned by sequence.

**Validation of sequences from NGS.** Hit peptides from NGS were resynthesized as gBlocks (IDT). These gBlocks were used as template for PCR to introduce SapI restriction sites compatible for re-cloning into the pAMK-267 library backbone. Newly reconstructed library members were transformed into *Marionette-Clo* cells containing modifying enzyme and selection plasmids and were then plated on media containing Carb/Kan/Spec. Individual transformants were then cultured in TB media + Carb/Kan/Spec in a deep well 96-well plate (1896–2000, USA Scientific, FL) and incubated overnight (Multitron Pro, 30 °C, 900 rpm). These cultures were then subcultured 1:100 in TB media + Carb/Kan/Spec either fully induced (1  $\mu$ M 3OC6-AHL, and 100  $\mu$ M cumate) or uninduced and incubated for 20 hr (Multitron Pro, 30 °C, 900 rpm) before taking 0.5  $\mu$ l for standard flow cytometry analysis.

**Analysis of Pap2c\_1 mutants.** Truncations of Pap2c\_1 were generated via insertion into pAMK-267 with standard cloning methods. These variants were then transformed into a *Marionette Clo* strain with the PapB modifying enzyme plasmid, pEG06\_044, and the selection plasmid, pAMK-866, and then grown overnight in 1 mL LB media + Kan/Carb at 30 °C, 900 rpm in a deep well 96-well plate. The next day, cultures were diluted 1:100 into 1 mL TB media + Kan/Carb containing either 0 or 1  $\mu$ M 3OC6-AHL. After overnight growth in the same growth conditions, the variants were assayed as described in Cytometry Analysis.

**Site-saturation mutagenesis of Pap2c\_1.** A library containing all possible single site substitutions was created from the TEV-GIRAY-core parent construct (pAMK-1056). 15 PCR reactions were run to create NNK substitutions at each position except C8 and E12. These 15 PCRs were then combined in equimolar amounts and assembled into the pAMK-267 backbone using SapI (see Library Generation). Colony counting indicated a library size of 1.3  $\times$  10<sup>6</sup>. This library was then miniprep and transformed into the *Marionette Clo* strain harboring the PapB modifying enzyme plasmid, pEG06\_044, and the selection plasmid, pAMK-866. After overnight growth in LB media 30 °C, the library with modifying enzymes and selection plasmid was then induced by diluting 20  $\mu$ l into 1 mL TB media + Carb/Kan/Spec and 3.2 nM 3OC6-AHL. Overnight growth was conducted in a deep well 96-well plate at 30 °C with 900 rpm shaking. After the culture reached saturation (20 hours), a selection was conducted by diluting 10  $\mu$ l into 1 mL TB media + Carb/Kan/Spec, 3.2 nM 3OC6-AHL, and 250  $\mu$ M Cm. The library was also diluted into media without any Cm to provide a “pre-selection” library. These cultures were grown until saturated (20 hours) before being miniprep. The pre- and post-selection library plasmids were then amplified via PCR using the osDAA154 and osDAA155 primers (Supplementary Table 3) and 200 bp bands were gel isolated. These fragments were then submitted for paired-end reads on an Illumina MiSeq, utilizing V2 chemistry. Paired-end alignments were conducted with PANDAseq and only sequences without any mismatches between forward and reverse reads

were kept for downstream processing, resulting in >50 K reads per sample. Read fractions for each core sequence variant were calculated by dividing the number of reads for each DNA sequence by total number of reads in the sample. Enrichment values were then calculated from these by dividing the read fraction of each sequence in the post-selection library by its read fraction in the pre-selection library. Degenerate DNA sequences corresponding to the same amino acid sequence were averaged to get a single enrichment value for each core variant. These amino acid sequence enrichments were then normalized to the enrichment for the parent core sequence to derive “Enrichment relative to parent”.

**Peptide purification.** Peptide hit gBlocks were cloned into the peptide expression plasmid, pEG03-119 using their flanking SapI restriction sites. The peptide and modifying enzyme plasmids were co-transformed into *E. coli* Marionette-X, streaked onto 2xYT agar with Carb/Spec and incubated at 30 °C overnight. Individual colonies were used to inoculate 20 mL of LB media in a 125 mL shake flask and incubated overnight at 30 °C and 250 rpm in an Innova44 (Eppendorf, NY). A 5 mL aliquot of overnight starter culture was diluted in 500 mL total volume TB media with Carb/Spec in Fernbach flasks and grown at 30 °C and 250 rpm until reaching OD<sub>600</sub> 0.8–1.0, at which point 1 mM IPTG and 200 μM cumate are added. Induced cultures were grown for a further 20 h at 30 °C and 250 rpm and then centrifuged (4,000 g, 4 °C, 10 min) in a Sorvall RC 6+ centrifuge (Thermo Fisher Scientific, MA). Pellets were resuspended in 30 mL lysis buffer (5 M guanidinium hydrochloride, 300 mM NaCl, 10 mM imidazole, 50 mM sodium phosphate, pH 7.5), and freeze-thawed twice (frozen in –80 °C freezer; thawed in innova44 incubator at 30 °C, 250 rpm). Cell lysates were centrifuged (Eppendorf 5424, 20,000 g, 18 °C, 45 min) in a Sorvall RC 6+ centrifuge (Thermo Fisher Scientific, MA) and the peptides affinity purified via gravity-flow using 3 mL resin-bed volume of Ni-NTA agarose resin (88223, Thermo Fisher Scientific, MA), following manufacturer instructions, using 2 resin-bed volumes water and lysis buffer for column equilibration, 4 resin-bed volumes of wash buffer (5 M guanidinium hydrochloride, 300 mM NaCl, 25 mM imidazole, 50 mM sodium phosphate, pH 7.5), 4 resin-bed volumes of elution buffer (5 M guanidinium hydrochloride, 300 mM NaCl, 250 mM imidazole, 50 mM sodium phosphate, pH 7.5). Eluate from Ni-NTA purification was then subjected to solid-phase extraction (SPE) using Strata-XL 500 mg tubes (8B-S043-HCH, Phenomenex, CA). The solid phase was first conditioned with 4 bed volumes of methanol and then water. Eluate was then loaded, washed with 8 bed volumes of 10 mM NH<sub>4</sub>CO<sub>3</sub>, and eluted with 8 bed volumes of 1:1 acetonitrile:aqueous 10 mM NH<sub>4</sub>CO<sub>3</sub>. Solvent was removed via lyophilization at –80 °C for 24–48 h. To cleave the SUMO and leader peptide from the core, the extracted peptide was resuspended in 20 mL TE buffer and 100 μl 20 mg/mL TEV protease and incubated overnight at room temperature with slow orbital shaking. The cleaved peptides were then desalted using a Strata-X PRO 500 mg SPE tubes (8B-S536-HCH, Phenomenex, CA). The solid phase was first conditioned with 4 bed volumes of methanol and then water. Eluate was then loaded, washed with 8 bed volumes of 10 mM NH<sub>4</sub>CO<sub>3</sub>, and eluted with 8 bed volumes of 1:1 acetonitrile:aqueous 10 mM NH<sub>4</sub>CO<sub>3</sub>. Solvent was removed via lyophilization at –80 °C for 24–48 h. After solvent removal, a 5 mL aliquot of the mixture resuspended in 10:90 acetonitrile:water was injected into a Agilent Technologies 1260 Infinity system HPLC (HPLC) system (Agilent Technologies, Santa Clara, CA) and separated using a 150 mm × 10 cm Aeris PEPTIDE XB-C18 column (100 Å, 5 μm) at a flow rate of 2 mL/min. Separation was carried out with a gradient program, with 0.1% formic acid as solvent A and acetonitrile with 0.1% formic acid as solvent B. The % B was held at 25% for 3 minutes, then increased to 50% over the next 17 min. The eluent was passed through a diode array detector (DAD) and absorbance at 270 nm was recorded. Detected peaks were collected using an Agilent G1364B Fraction Collector and again solvent was removed via lyophilization at –80 °C for 24–48 h. Samples were resuspended in 1 mL of 1:1 acetonitrile:aqueous 10 mM NH<sub>4</sub>CO<sub>3</sub> in pre-weighed 2 mL microcentrifuge tubes (Eppendorf) and solvent was removed via lyophilization at –80 °C for 24–48 h. Yields were measured by comparing mass of empty tubes to tubes containing lyophilized powder.

**Liquid chromatography/mass spectrometry.** All chromatography was performed using the mobile phases ACN (acetonitrile supplemented with 0.1% formic acid and 0.1% water) and water (supplemented with 0.1% formic acid). The “QTOF” is an Agilent 1260 Infinity II liquid chromatograph with binary pump configured in low-dwell volume mode and column oven set to 40 °C, coupled to an Agilent 6545 QTOF mass spectrometer equipped with an Agilent electrospray ionization (ESI) source. Nitrogen gas is building-supplied and ESI source parameters are 350 °C gas temperature, 12 L/min gas flow, 30 psig nebulizer pressure, 350 °C sheath gas temperature, 8 L/min sheath gas flow, 3000 V capillary voltage, 1000 V nozzle voltage, 135 V fragmentor voltage, 15 V skimmer voltage, 600 V Oct 1 RF Vpp; the mass spectrometer was run in MS mode with reference mass enabled and tuned in positive mode with standard mass range (3200 *m/z*) and 2 GHz extended dynamic range. QTOF analysis was performed with a Phenomenex Aeris PEPTIDE XB-C18 2.6 μm 50 mm × 2.1 mm column. The flow rate was set at 0.5 mL/min and 5 μl sample was injected. The gradient used was 20% ACN for 0.5 min, 20–55% ACN over 5.5 min, 55–90% ACN over 0.5 minutes, 90% ACN for 1.5 min, with 0.8 min re-equilibration. LC–MS data were analyzed using

MassHunter version 10.0. Accurate mass predictions of peptides were generated using the online resource, ChemCalc<sup>81</sup>.

**Bio-layer interferometry.** Assays were performed on an Octet Red (ForteBio) instrument at 30 °C with shaking at 1,000 rpm. Ni-NTA biosensors (18-5101, ForteBio, NY) were hydrated in 1× kinetics buffer (diluted from 10× buffer; 18-5032, ForteBio, NY) for 30 min before the measurement. Expi293F human cell-derived and purified SARS-CoV-2 RBD (RBD<sub>296-531</sub>) was loaded at 10–20 μg/mL in 1× Kinetics Buffer for 300 s prior to baseline equilibration for 180 s in 1× kinetics buffer. Association reactions of the peptide to RBD<sub>296-531</sub> were carried out in 1× kinetics buffer at various concentrations in a two-fold dilution series from 80 mM to 1.25 mM was carried out for 900 s. Then dissociation reactions were observed for 900 s. Data were acquired using the Octet Data Acquisition software version 9.0 and response data were generated using FortéBio data analysis software version 9.0.

**Competition analysis using Bio-layer interferometry.** Competition of peptide AMK-1057 with hACE2, B38, and CR3022 for binding to recombinant SARS-CoV-2 RBD was evaluated using a ForteBio Octet Red instrument at 30 °C with shaking at 1,000 rpm. Ni-NTA biosensors (18-5101, ForteBio, NY) were hydrated in 1× kinetics buffer (diluted from 10× buffer; 18-5032, ForteBio NY) for 30 min before the measurement. Expi293F human cell-derived and purified SARS-CoV-2 RBD was loaded at 20 μg/mL in 1× Kinetics Buffer for 450 s prior to baseline equilibration for 300 s in 1× kinetics buffer. The RBD-loaded biosensor tips were exposed (300 s) to the hACE2 soluble receptor (300 nM), B38 (300 nM), or CR3022 IgG (300 nM), and then exposed (300 s) to either AMK-1057 peptide (10000 nM) or 1× kinetics buffer. The data was interstep corrected using the FortéBio Data Analysis Software. Additional binding by the secondary molecule indicates an unoccupied epitope (non-competitor), while no binding indicates epitope blocking (competitor).

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

## Data availability

Source data are provided with this paper. Genetic part sequences are available in the Supplementary Information. Plasmids are available from Addgene. Any other data are available from the corresponding author upon reasonable request.

## Code availability

Python scripts used for cytometry processing are released as open source software under the MIT license (GitHub repository: <https://github.com/VoigtLab/constr-peptide-paper-figs>).

Received: 10 November 2020; Accepted: 27 September 2021;

Published online: 03 November 2021

## References

- Bluntzer, M. T. J., O’Connell, J., Baker, T. S., Michel, J. & Hulme, A. N. Designing stapled peptides to inhibit protein-protein interactions: an analysis of successes in a rapidly changing field. *Pept. Sci.* **113**, <https://doi.org/10.1002/pep2.24191> (2021).
- Zhang, H. et al. Structure-activity relationship and conformational studies of the natural product cyclic depsipeptides YM-254890 and FR900359. *Eur. J. Med. Chem.* **156**, 847–860 (2018).
- Okino, T., Matsuda, H., Murakami, M. & Yamaguchi, K. New microviridins, elastase inhibitors from the blue-green alga *Microcystis aeruginosa*. *Tetrahedron* **51**, 10679–10686 (1995).
- Schreiber, S. L. & Crabtree, G. R. The mechanism of action of cyclosporin A and FK506. *Immunol. Today* **13**, 136–142 (1992).
- Banaszynski, L. A., Liu, C. W. & Wandless, T. J. Characterization of the FKBP.rapamycin.FRB ternary complex. *J. Am. Chem. Soc.* **127**, 4715–4721 (2005).
- Cubillos-Ruiz, A., Berta-Thompson, J. W., Becker, J. W., van der Donk, W. A. & Chisholm, S. W. Evolutionary radiation of lanthipeptides in marine cyanobacteria. *Proc. Natl Acad. Sci. USA* **114**, E5424–E5433 (2017).
- Culp, E. J. et al. Evolution-guided discovery of antibiotics that inhibit peptidoglycan remodelling. *Nature*, <https://doi.org/10.1038/s41586-020-1990-9> (2020).
- Shigdel, U. K. et al. Genomic discovery of an evolutionarily programmed modality for small-molecule targeting of an intractable protein surface. *Proc. Natl Acad. Sci. USA* **117**, 17195–17203 (2020).
- Davis, A. M., Plowright, A. T. & Valeur, E. Directing evolution: the next revolution in drug discovery? *Nat. Rev. Drug Discov.*, <https://doi.org/10.1038/nrd.2017.146> (2017).

10. White, A. M. & Craik, D. J. Discovery and optimization of peptide macrocycles. *Expert Opin. Drug Discov.* **11**, 1151–1163 (2016).
11. Arnison, P. G. et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).
12. Fischbach, M. A. & Walsh, C. T. Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
13. Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat. Chem. Biol.* **11**, 660–670 (2015).
14. Zhang, J. J., Tang, X. & Moore, B. S. Genetic platforms for heterologous expression of microbial natural products. *Nat. Prod. Rep.* **36**, 1313–1332 (2019).
15. Nivina, A., Herrera Paredes, S., Fraser, H. B. & Khosla, C. GRINS: Genetic elements that recode assembly-line polyketide synthases and accelerate their diversification. *Proc. Natl Acad. Sci. USA* **118**, <https://doi.org/10.1073/pnas.2100751118> (2021).
16. Peel, M. & Scribner, A. Semi-synthesis of cyclosporins. *Biochim Biophys. Acta* **1850**, 2121–2144 (2015).
17. Guo, Z. et al. Rapamycin-inspired macrocycles with new target specificity. *Nat. Chem.* **11**, 254–263 (2019).
18. Montalbán-López, M. et al. New developments in RiPP discovery, enzymology and engineering. *Nat. Prod. Rep.*, <https://doi.org/10.1039/d0np00027b> (2020).
19. Hegemann, J. D., Bobeica, S. C., Walker, M. C., Bothwell, I. R. & van der Donk, W. A. Assessing the flexibility of the prochlorosin 2.8 scaffold for bioengineering applications. *ACS Synth. Biol.*, <https://doi.org/10.1021/acssynbio.9b00080> (2019).
20. Liu, R. et al. A cell-free platform based on nisin biosynthesis for discovering novel lanthipeptides and guiding their overproduction in vivo. *Adv. Sci. (Weinh.)* **7**, 2001616 (2020).
21. Gonzalez-Nicolini, V. & Fussenecker, M. In vitro assays for anticancer drug discovery—a novel approach based on engineered mammalian cell lines. *Anticancer Drugs* **16**, 223 (2005).
22. Ermert, P., Luther, A., Zbinden, P. & Obrecht, D. Frontier between cyclic peptides and macrocycles. *Methods Mol. Biol.* **2001**, 147–202 (2019).
23. Scott, J. K. & Smith, G. P. Searching for peptide ligands with an epitope library. *Science* **249**, 386–390 (1990).
24. Urban, J. H. et al. Phage display and selection of lanthipeptides on the carboxy-terminus of the gene-3 minor coat protein. *Nat. Commun.* **8**, 1500 (2017).
25. Wang, X. S. et al. A genetically encoded, phage-displayed cyclic-peptide library. *Angew. Chem. - Int. Ed.* **58**, 15904–15909 (2019).
26. Hetrick, K. J., Walker, M. C. & van der Donk, W. A. Development and application of yeast and phage display of diverse lanthipeptides. *ACS Cent. Sci.* **4**, 458–467 (2018).
27. Owens, A. E., Iannuzzelli, J. A., Gu, Y. & Fasan, R. MOrPH-PhD: an integrated phage display platform for the discovery of functional genetically encoded peptide macrocycles. <https://doi.org/10.1021/acscentsci.9b00927> (2019).
28. Tiede, C. et al. Adhiron: a stable and versatile peptide display scaffold for molecular recognition applications. *Protein Eng., Des. Selection* **27**, 145–155 (2014).
29. Kale, S. S. et al. Cyclization of peptides with two chemical bridges affords large scaffold diversities. *Nat. Chem.* **10**, 715–723 (2018).
30. Guillen Schlippe, Y. V., Hartman, M. C. T., Josephson, K. & Szostak, J. W. In vitro selection of highly modified cyclic peptides that act as tight binding inhibitors. *J. Am. Chem. Soc.* **134**, 10469–10477 (2012).
31. Sakai, K. et al. Macrocyclic peptide-based inhibition and imaging of hepatocyte growth factor. *Nat. Chem. Biol.* **15**, 598–606 (2019).
32. Wang, T. et al. Detecting protein–protein interaction based on protein fragment complementation assay. *Curr. Protein Pept. Sci.* **21**, 598–610 (2020).
33. Yao, Z. et al. Split intein-mediated protein ligation for detecting protein–protein interactions and their inhibition. *Nat. Commun.* **11**, 2440–2440 (2020).
34. Pinto, F., Thornton, E. L. & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat. Commun.* **11**, 1–15 (2020).
35. Azad, T., Tashakor, A. & Hosseinkhani, S. Split-luciferase complementary assay: applications, recent developments, and future perspectives. *Anal. Bioanal. Chem.* **406**, 5541–5560 (2014).
36. Magliery, T. J. et al. Detecting protein–protein interactions with a green fluorescent protein fragment reassembly trap: scope and mechanism. *J. Am. Chem. Soc.* **127**, 146–157 (2005).
37. Tebo, A. G. & Gautier, A. A split fluorescent reporter with rapid and reversible complementation. *Nat. Commun.* **10**, 1–8 (2019).
38. Horswill, A. R., Savinov, S. N. & Benkovic, S. J. A systematic method for identifying small-molecule modulators of protein–protein interactions. *Proc. Natl Acad. Sci. USA* **101**, 15591–15596 (2004).
39. Tavassoli, A. et al. Inhibition of HIV budding by a genetically selected cyclic peptide targeting the Gag-TSG101 interaction. *ACS Chem. Biol.* **3**, 757–764 (2008).
40. Yang, X. et al. A lanthipeptide library used to identify a protein–protein interaction inhibitor. *Nat. Chem. Biol.* **14**, 375–380 (2018).
41. Sun, X. et al. Peptide-based imaging agents for cancer detection. *Adv. Drug Deliv. Rev.* **110–111**, 38–51 (2017).
42. Kwon, D. et al. High-contrast CXCR4-targeted 18F-PET imaging using a potent and selective antagonist. *Mol. Pharm.* **18**, 187–197 (2021).
43. Lai, A. C. & Crews, C. M. Induced protein degradation: an emerging drug discovery paradigm. *Nat. Rev. Drug Discov.*, <https://doi.org/10.1038/nrd.2016.211> (2016).
44. Sakamoto, K. M. et al. Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc. Natl Acad. Sci. USA* **98**, 8554–8559 (2001).
45. Walls, A. C. et al. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* <https://doi.org/10.1016/j.cell.2020.02.058> (2020).
46. Aranko, A. S., Wlodawer, A. & Iwai, H. Nature's recipe for splitting inteins. *Protein Eng. Des. Sel.* **27**, 263–271 (2014).
47. Zettler, J., Schutz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).
48. Rhodiou, V. A. et al. Design of orthogonal genetic switches based on a crosstalk map of *os*, anti-*os*, and promoters. *Mol. Syst. Biol.* **9**, 702 (2013).
49. Pinto, F., Thornton, E. L. & Wang, B. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat. Commun.* **11**, 1529 (2020).
50. Moll, U. M. & Petrenko, O. The MDM2-p53 interaction. *Mol. Cancer Res.* **1**, 1001–1008 (2003).
51. Kussie, P. H. et al. Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948–953 (1996).
52. Li, C. et al. Systematic mutational analysis of peptide inhibition of the p53-MDM2/MDMX interactions. *J. Mol. Biol.* **398**, 200–213 (2010).
53. Meyer, A. J., Segall-Shapiro, T. H., Glassey, E., Zhang, J. & Voigt, C. A. *Escherichia coli* “Marionette” strains with 12 highly optimized small-molecule sensors. *Nat. Chem. Biol.*, <https://doi.org/10.1038/s41589-018-0168-3> (2018).
54. Zhang, G. et al. Investigation of ACE2 N-terminal fragments binding to SARS-CoV-2 Spike RBD. <https://doi.org/10.1101/2020.03.19.999318> (2020).
55. Zong, C., Maksimov, M. O. & Link, A. J. Construction of lasso peptide fusion proteins. *ACS Chem. Biol.* **11**, 61–68 (2016).
56. Hudson, G. A. et al. Bioinformatic mapping of radical S-adenosylmethionine-dependent ribosomally synthesized and post-translationally modified peptides identifies new Ca, C $\beta$ , and C $\gamma$ -linked thioether-containing peptides. *J. Am. Chem. Soc.*, <https://doi.org/10.1021/jacs.9b01519> (2019).
57. Precord, T. W., Mahanta, N. & Mitchell, D. A. Reconstitution and substrate specificity of the thioether-forming radical S-adenosylmethionine enzyme in freyrasin biosynthesis. *ACS Chem. Biol.* **14**, 1981–1989 (2019).
58. Noy-Porat, T. et al. A panel of human neutralizing mAbs targeting SARS-CoV-2 spike at multiple epitopes. *Nat. Commun.* **11**, 4303 (2020).
59. Chen, J. et al. Receptor-binding domain of SARS-Cov spike protein: soluble expression in *E. coli*, purification and functional characterization. *World J. Gastroenterol.* **11**, 6159–6164 (2005).
60. Maranhao, A. C. & Ellington, A. D. Evolving orthogonal suppressor tRNAs to incorporate modified amino acids. *ACS Synth. Biol.* **6**, 108–119 (2017).
61. Zhang, Q. et al. Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS. *Proc. Natl Acad. Sci. USA* **111**, 12031–12036 (2014).
62. Gavrish, E. et al. Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chem. Biol.* **21**, 509–518 (2014).
63. Delgado, M. A., Rintoul, M. R., Farias, R. N. & Salomon, R. A. *Escherichia coli* RNA polymerase is the target of the cyclopeptide antibiotic microcin J25. *J. Bacteriol.* **183**, 4543–4550 (2001).
64. Rao, J., Lahiri, J., Weis, R. M. & Whitesides, G. M. Design, synthesis, and characterization of a high-affinity trivalent system derived from vancomycin and L-Lys-d-Ala-d-Ala. *J. Am. Chem. Soc.* **122**, 2698–2710 (2000).
65. Wec, A. Z. et al. Broad neutralization of SARS-related viruses by human monoclonal antibodies. *Science*, <https://doi.org/10.1126/science.abc7424> (2020).
66. Lan, J. et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature*, <https://doi.org/10.1038/s41586-020-2180-5> (2020).
67. Yuan, M. et al. A highly conserved cryptic epitope in the receptor-binding domains of SARS-CoV-2 and SARS-CoV. *Science*, <https://doi.org/10.1126/science.abb7269> (2020).
68. Wu, Y. et al. A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274–1278 (2020).

69. Zhang, Z. et al. GTP-State-Selective Cyclic Peptide Ligands of K-Ras(G12D) Block Its Interaction with Raf. *ACS Cent. Sci.* **6**, 1753–1761 (2020).
70. Bond, M. J. & Crews, C. M. Proteolysis targeting chimeras (PROTACs) come of age: entering the third decade of targeted protein degradation. *RSC Chem. Biol.*, <https://doi.org/10.1039/D1CB00011J> (2021).
71. Bondeson, D. P. et al. Lessons in PROTAC Design from Selective Degradation with a Promiscuous Warhead. *Cell Chem. Biol.* **25**, 78–87 (2018). e75.
72. Edward Williams, M. & Tincho, M. Molecular validation of putative antimicrobial peptides for improved human immunodeficiency virus diagnostics via HIV protein p24. *J. AIDS Clin. Res.* **07**, <https://doi.org/10.4172/2155-6113.1000571> (2016).
73. Luther, A., Moehle, K., Chevalier, E., Dale, G. & Obrecht, D. Protein epitope mimetic macrocycles as biopharmaceuticals. *Curr. Opin. Chem. Biol.* **38**, 45–51 (2017).
74. Giordanetto, F. & Kihlberg, J. Macrocyclic drugs and clinical candidates: what can medicinal chemists learn from their properties? *J. Med. Chem.* **57**, 278–295 (2014).
75. Morrison, C. Constrained peptides' time to shine? *Nat. Rev. Drug Discov.* **17**, 531–533 (2018).
76. Newman, D. J. & Cragg, G. M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **83**, 770–803 (2020).
77. Kilby, J. M. et al. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* **4**, 1302–1307 (1998).
78. Bird, G. H. et al. Hydrocarbon double-stapling remedies the proteolytic instability of a lengthy peptide therapeutic. *Proc. Natl Acad. Sci. USA* **107**, 14093–14098 (2010).
79. Burkhart, B. J., Kakkar, N., Hudson, G. A., van der Donk, W. A. & Mitchell, D. A. Chimeric leader peptides for the generation of non-natural hybrid RiPP products. *ACS Cent. Sci.* **3**, 629–638 (2017).
80. Segall-Shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat. Biotechnol.*, <https://doi.org/10.1038/nbt.4111> (2018).
81. Patiny, L. & Borel, A. ChemCalc: a building block for tomorrow's chemical infrastructure. *J. Chem. Inf. Model* **53**, 1223–1228 (2013).
82. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

## Acknowledgements

This research was funded by the US Defense Advanced Research Projects Agency (DARPA) program award HR0011-15-C-0084, a research award from Novartis Institute for BioMedical Research (Cambridge, USA), and the Banting Fellowships Program (A.M.K.). This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant no. #1745302 (D.A.A.). Expi293F human cell-derived and purified SARS-CoV-2 RBD was obtained as a gift from the King

lab (University of Washington). hACE2, B38, and CR3022 were obtained as a gift from the Schmidt lab (Harvard University).

## Author contributions

A.M.K., E.G., D.A.A., T.H.S.S., and C.A.V. conceived the study and designed the experiments. T.H.S.S. constructed initial split intein selection system. Z.Z. performed BLI experiments. D.L.N. performed pap2c library assessment experiments. A.M.K., Z.Z., and T.L.W. HPLC purified peptides. K.P. conducted NGS experiments. A.M.K. and A.C.E. cloned constructs, A.M.K., E.G., and D.A.A. performed all other experiments. A.M.K., E.G., D.A.A., and D.B.G. analyzed the data. A.M.K., E.G., D.A.A., and C.A.V. wrote the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26350-4>.

**Correspondence** and requests for materials should be addressed to Christopher A. Voigt.

**Peer review information** *Nature Communications* thanks Igor Stagljar and Wilfred van der Donk for their contribution to the peer review of this work. Peer reviewer reports are available.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021