Discovering Covalent Cyclic Peptide Inhibitors of Peptidyl Arginine Deiminase 4 (PADI4) Using mRNA-Display with a Genetically Encoded Electrophilic Warhead

Isabel R. Mathiesen,[1,2] Ewen D. D. Calder,[1,2] Simone Kunzelmann,[3] Louise J. Walport[1,2*]

[1] Protein-Protein Interaction Laboratory, The Francis Crick Institute, London NW1 1AT, United Kingdom

[2] Department of Chemistry, Molecular Sciences Research Hub, Imperial College London, London W12 0BZ, United Kingdom

[3]Structural Biology Scientific Technology Platform, The Francis Crick Institute, London NW1 1AT, United Kingdom

*To whom correspondence should be addressed: l.walport@imperial.ac.uk

Abstract

Covalent drugs can achieve high potency with long dosing intervals. However, concerns remain about side-effects associated with off-target reactivity. Combining macrocyclic peptides with covalent warheads provides a solution to minimise off-target reactivity: the peptide enables highly specific target binding, positioning a weakly reactive warhead proximal to a suitable residue in the target. Here we demonstrate direct discovery of covalent cyclic peptides using encoded libraries containing a weakly electrophilic cysteine-reactive fluoroamidine warhead. We combine direct incorporation of the warhead into peptide libraries using the flexible *in vitro* translation system with a peptide selection approach that identifies only covalent target binders. Using this approach, we identify potent covalent inhibitors of the peptidyl arginine deiminase, PADI4 or PAD4, that react exclusively at the active site cysteine. We envisage this approach will enable covalent peptide inhibitor discovery for a range of related enzymes and expansion to alternative warheads in the future.

Introduction

Covalent inhibitors convey beneficial properties including increased potency, simpler pharmacokinetics, due to non-equilibrium kinetics, and potential for extended dosing intervals.[1,2] Additionally, covalent inhibitors are useful for competition with high

concentrations of endogenous ligands due to their nonequilibrium binding mechanism. However, there remain concerns about off-target effects, due to reaction of the covalent warhead with other proteins.[3] Targeted covalent inhibitors (TCIs) address this by combining weakly electrophilic warheads with high affinity scaffolds which optimally position the reactive group to react at a target residue.[4,5]

Peptides make an ideal modality for the high affinity scaffold due to their tight binding affinities, high target selectivity and relative ease and low cost of synthesis. Having a comparably small size, they can be orally bioavailable, whilst still having a sufficiently large surface area to target relatively featureless protein interfaces with high specificity.[1,6–9] Macrocyclisation of peptides confers additional benefits including high proteolytic stability and increased potency.[10,11]

Classically, covalent peptides are developed through addition of a warhead into a previously identified reversible binder or substrate analogue, requiring structural information or laboriously generated structure activity relationship information.[12–15] Identifying a suitable site for warhead addition that enables efficient reaction without disrupting potent target binding is challenging. Additionally, in many cases neither a substrate analogue nor structural information is available. Addressing both these challenges, direct identification of covalent peptides from high-throughput screening offers a route to speed up covalent drug discovery.

Genetically-encoded peptide screening platforms, such as phage display, mRNA display and the related random non-standard peptides integrated discovery (RaPID) system, provide powerful approaches to identify peptide hits from enormous libraries of cyclic peptides (up to 10^{13} sequences).[16–19] These platforms have been used successfully to identify potent reversible chemical tools and drug candidates to a wide range of targets.[7,20–23] Recently these screening approaches have been modified to promote bias towards the discovery of irreversible covalent inhibitors; reactive moieties have been introduced into the peptide libraries alongside denaturing guanidine washes during the peptide selection step.[24,25] For example, a modified phage display protocol has been used to identify *de novo* covalent peptide inhibitors through post-translational modification of peptide libraries with warheads

into a fixed position — either within the cyclisation linker or at reduced disulphide bonds.[26–29]

Other methods to introduce unnatural chemistry into peptides have also been developed, including through use of modified aminoacyl tRNA synthetases, chemical aminoacylation of tRNA or use of aminoacylating ribozymes, known as flexizymes.[17,30–34] The RaPID system offers a route to identify chemically diverse peptide binders through encoding non-canonical amino acids in displayed peptides using the flexizyme-mediated flexible *in vitro* translation (FIT) system.[35,23] This enables both facile peptide cyclisation and the potential for direct warhead incorporation.[36,37] Unlike in the phage display approaches this allows incorporation of the covalent warhead at variable positions in the peptide macrocycle. Recently, this strategy has been used to incorporate phenylselenocysteine into RaPID libraries, which was then post-translationally modified to yield a dehydroalanine warhead.[24] Photoreactive covalent peptides have also been identified through incorporation of a benzophenone moiety.[25] However, the ability to directly encode an unmasked electrophilic warhead within displayed libraries, rather than relying on post-translational modification, has not yet been exploited.

Peptidyl arginine deiminase 4 (PADI4 or PAD4) is one of five enzymes in the PADI family. PADIs 1-4 catalyse the post-translational modification of peptidyl arginine residues to citrulline in a wide range of protein substrates.[38] PADI4 is involved in cell signalling processes including apoptosis, differentiation, and regulation of transcription.[39–42] Dysregulation of PADI4 is implicated in various diseases including rheumatoid arthritis, lupus and several cancers.[43–45] PADI4 has a key active site Cys residue required for catalysis and there are known covalent small molecule binders.[46,47] Fluoroamidine (**1**) is a small molecule inhibitor of PADI4; as an arginine mimetic (Figure 1A) it binds in the active site of PADI enzymes and covalently reacts with the active site cysteine (C645).[47] However, it has limited selectivity for PADI4 over PADI1.[48] From a screen of synthetic peptides containing **1**, the tripeptide TDFA was identified with an IC_{50} of 2.3 μM and >15-fold selectivity for PADI4 over PADI1.[49] Based on this precedent we envisaged that developing a high-throughput methodology to identify larger covalent peptides including the fluoroamidine warhead might provide a route to even more potent and selective inhibitors.

Here we report the direct incorporation of the cysteine reactive electrophile, fluoroamidine, into cyclic peptide RaPID libraries produced by *in vitro* translation. We apply our covalent RaPID library in a screen against PADI4 to select exclusively for peptides which are covalently bound to the target protein. Our approach yielded peptides which covalently bind to PADI4 at Cys645 and inhibit PADI4 citrullination activity, three of which have k_{inact}/K_I in the order of 10^6 M^{-1} min^{-1}. Our current method uses a cysteine reactive warhead, but it is a generally applicable strategy where any weakly reactive covalent warhead can conceivably be incorporated.

Results and Discussion

Building on previous advances in the development of covalent peptides using encoded libraries,[24,27–29] we aimed to develop a covalent peptide discovery platform using RaPID in which an unmasked electrophile would be directly incorporated into the *in vitro* translated peptides. This first required successful ribosomal incorporation of an electrophile warhead into mRNA-displayed cyclic peptide libraries. We synthesised an unnatural amino acid version of **1**, *N*-δ-fluoroacetimidoyl ornithine (FAO, **2**, Figure 1A). To allow flexizyme recognition for aminoacylation onto tRNA we activated the carboxylic acid as the 4-chlorobenzyl thioester (CBT) (FAO-CBT, **3**, Figure 1A).[33] This thioester was used because the more commonly used dinitrobenzyl ester of FAO was synthetically intractable. After confirming successful flexizyme-mediated loading onto a short tRNA mimic (Figure 1B), we tested for ribosomal compatibility. An elongator tRNA with the methionine anticodon (CAU) was aminoacylated with FAO. This aminoacylated tRNA was used to perform *in vitro* translation using the PURExpress® translation system with methionine omitted. The initiator methionine was reprogrammed to chloroacetyl-D-tyrosine. We translated a peptide template containing a single methionine in the elongator region. MALDI-TOF spectroscopy showed that ribosomal incorporation of FAO was successful (Figure S1A). As our initial translation efficiency was low, we optimised this by screening four different elongator tRNAs containing variable tRNA T-stems (Figure S1B).[50] As the T-stem number increases (1–4) this increases the affinity of the tRNA for elongation factor thermo unstable (EF-Tu), which we observed to correlate with enhanced ribosomal FAO incorporation efficiency. Optimal translation was observed with T-

stem 4 which we went on to use in all subsequent experiments. No warhead hydrolysis or reaction with components in the translation system or buffer was observed, confirming that, at least with this electrophile, covalency can be encoded for in mRNA display with genetic code reprogramming, without the need for masking and post-translational modification.

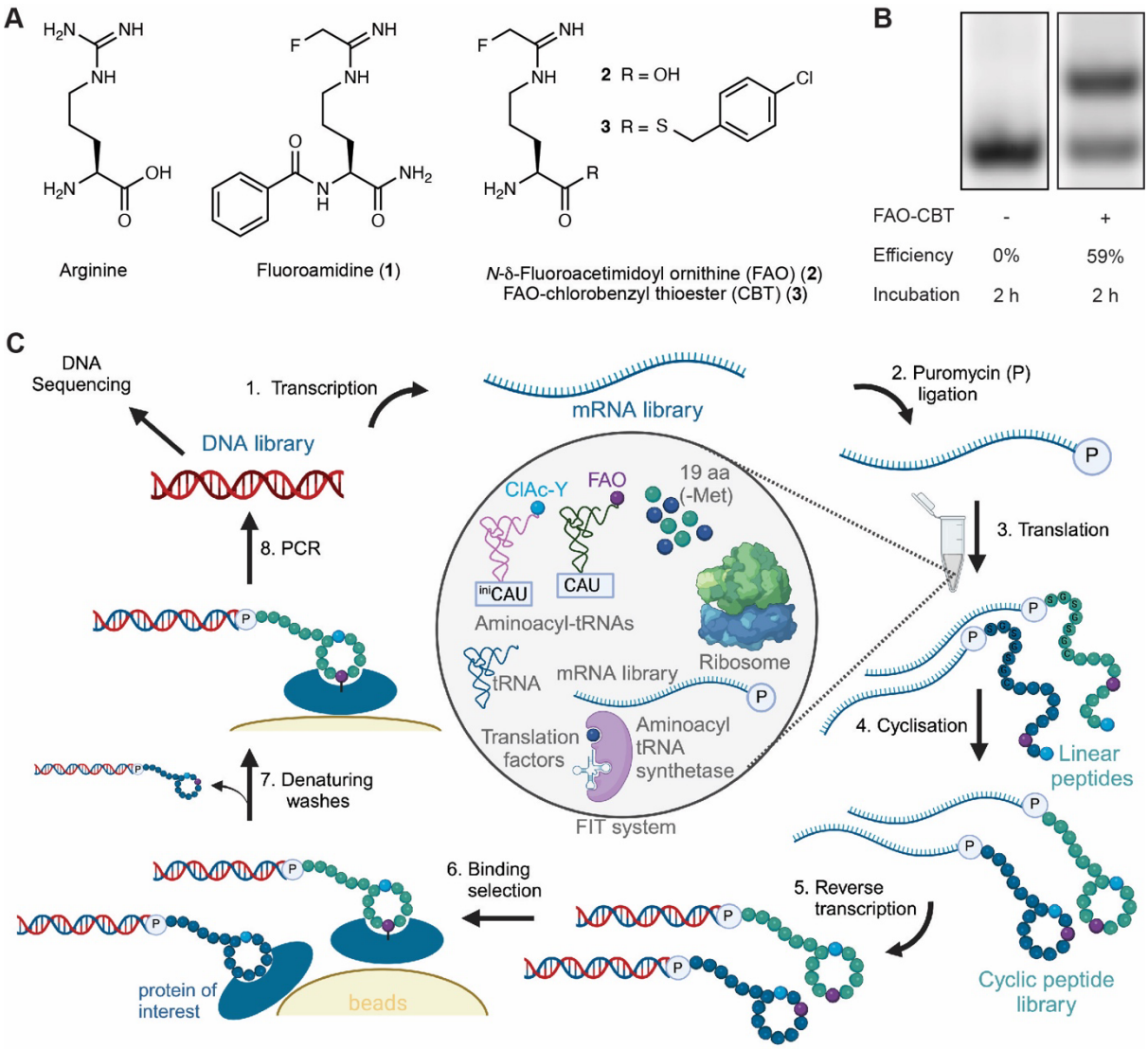

Figure 1: Covalent RaPID setup.
A The structure of arginine and related arginine-mimetic PADI4 inhibitor **1** and synthesised unnatural amino acids **2** and **3**.
B Microhelix assay using a truncated tRNA mimic to monitor loading of FAO-CBT (**3**) using eFx. The upper band indicates the presence of aminoacylated microhelix tRNA and the lower band is non-aminoacylated microhelix tRNA. After 2 h incubation at 4 °C, 59% aminoacylation is seen. The full gel is provided in Figure S23.
C The covalent RaPID cycle setup. Transcription, puromycin ligation, translation, reverse transcription, and affinity panning against immobilised PADI4 are performed as in a typical RaPID selection. However, denaturing washes are added as an additional step to remove non-covalent peptide binders to PADI4. The translation incorporates FAO (**2**) and chloroacetylated-D-tyrosine to promote covalent binding and cyclisation, respectively.

Denaturing guanidium chloride washes have been used by us and others during the affinity panning step of peptide selections to remove non-covalent binders and only retain covalently binding peptides (Figure 1C).[24,25,27] We confirmed that at concentrations up to 8 M, this did

not disrupt the interaction between biotinylated PADI4 and streptavidin beads (Figure S2). Based on previous successful selections we chose a concentration of 5 M for the washes.[25]

Before performing a full selection we wanted to test our new approach using a model protein-peptide target pair. PADI4_3 is a cyclic peptide inhibitor of PADI4 that was recently discovered using a RaPID screen.[51] A cryo-electron microscopy structure of this peptide bound to PADI4 revealed that His4 of PADI4_3 bound in the active site of PADI4, in the position normally occupied by the arginine side chain of substrate peptides (PDB ID: 8R8U). We hypothesised that substitution of this residue with **2** would enable covalent inhibition of PADI4, as the warhead should be positioned to react with the PADI4 active site cysteine, Cys645. To confirm this, PADI4_3_H4(**2**) was synthesised. The linear sequence was synthesised by solid-phase peptide synthesis (SPPS) with an ornithine residue in position 4 of the peptide. The resultant peptide was cyclised, the ornithine selectively deprotected to allow addition of the fluoroacetimadoyl group, before full peptide deprotection, resin cleavage and purification. With the purified peptide in hand, 10 equivalents PADI4_3_H4(**2**) were incubated with PADI4, and intact mass spectrometry (MS) performed. This confirmed that PADI4_3_H4(**2**) covalently bound to PADI4 at a single site (Figure S3). To evaluate whether the peptides were reacting at the active site cysteine, Cys645, we produced an inactive PADI4 variant in which the Cys645 was substituted with alanine, PADI4 C645A, and performed the same experiment (Figure S4A-S4C, S5). No covalent binding between the peptide and PADI4 C645A was observed by intact-MS, confirming that PADI4_3_H4(**2**) was reacting exclusively with the active site Cys645.

Having confirmed that PADI4_3_H4(**2**) was a covalent binder of PADI4, we synthesised three model mRNA templates for use in a test selection. The first template encoded for the wildtype PADI4_3 sequence. The second encoded a sequence where His4 in the PADI4_3 sequence was substituted for a Met codon that could be reprogrammed to FAO, PADI4_3_H4M. The third template encoded a control sequence where an arginine residue in the sequence, Arg2, was replaced by a Met codon, PADI4_3_R2M, which we anticipated would not be correctly positioned to covalently react with PADI4 (Figure 2A, Figure S6). We performed a single cycle of RaPID screening (clone assay), with each of the individual mRNA templates, to assess peptide binding. In each case, translated RaPID peptide was incubated with PADI4 at room temperature for 1 hour to allow time for covalent reaction, before affinity panning was

performed both with and without denaturing guanidinium chloride washes. qPCR was used to quantify DNA recovery for each peptide. In the absence of guanidinium chloride washes, all three peptides bound to PADI4, whilst only PADI4_3_H4M was retained after guanidinium chloride washes (Figure 2B). This confirmed that the translated warhead was competent to react with cysteine residues in the target protein, that the guanidinium chloride washes were effective at retaining only covalently bound peptides and that we were not observing high levels of non-specific peptide reaction.

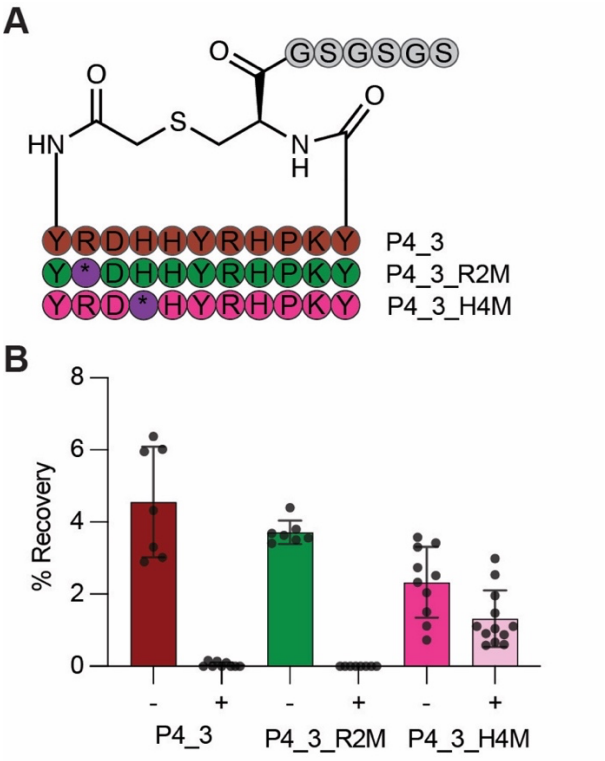

Figure 2: Proof of principle with PADI4 inhibitor PADI4_3.
A Sequences of the different PADI4_3 analogues synthesised as mRNA templates and translated, where internal 'M' codons of R2M and H4M are reprogrammed to FAO warhead (*).
B Clone assay results against PADI4 without (-) or with (+) guanidinium washes. Data shows mean percentage recovery from at least 6 replicates and error bars represent ±1 standard deviation.

Having confirmed that our encoded electrophile RaPID setup could identify covalent cyclic peptides, we set out to perform a *de novo* peptide screen for covalent binders of PADI4. For this we used an mRNA-displayed library encoding peptides with between six and ten randomised positions, flanked by an initiator codon and a CGSGSGS *C*-terminal linker. Peptide cyclisation was enabled by flexizyme-mediated reprogramming of the initiator codon to *N*-chloroacetyl-D-tyrosine which would spontaneously cyclise with the cysteine in the *C*-terminal linker. All internal Met codons were reprogrammed to **2**. Following translation and reverse transcription, this library was used in a covalent RaPID selection. In each round, the library

was preincubated with PADI4 at room temperature prior to denaturing affinity panning against PADI4. We saw low recovery when the library was incubated with biotinylated streptavidin beads (negative selection) and increasing positive library recovery for the first 4 rounds with PADI4-bound beads (Figure 3A). Following 5 rounds of selection, next-generation sequencing was performed on the DNA libraries recovered after each selection round (Supplementary Data 1). Sequencing results suggested that our RaPID setup rapidly enriched for covalent peptides, because from round 2 onwards 99% of sequences had an internal Met codon, indicating warhead presence (Figure 3B). Interestingly, however, the sequencing data from later rounds did not resemble typical successful RaPID selections. Although our total library recovery increased through the rounds, rather than finding a smaller number of highly enriched peptide sequences, contributing substantially to the total recovered libraries, we saw many different individual sequences each with relatively low abundance (Figure S7). This suggested that reaction with **2** was permissible within a wide range of peptide sequence contexts. Despite this, multiple sequence alignments indicated the enrichment of certain families and a clear increase in the abundance of certain sequences within these families round by round, indicative of target binding (Figure 3C). We selected 6 of these sequences for synthesis by SPPS and further characterisation.

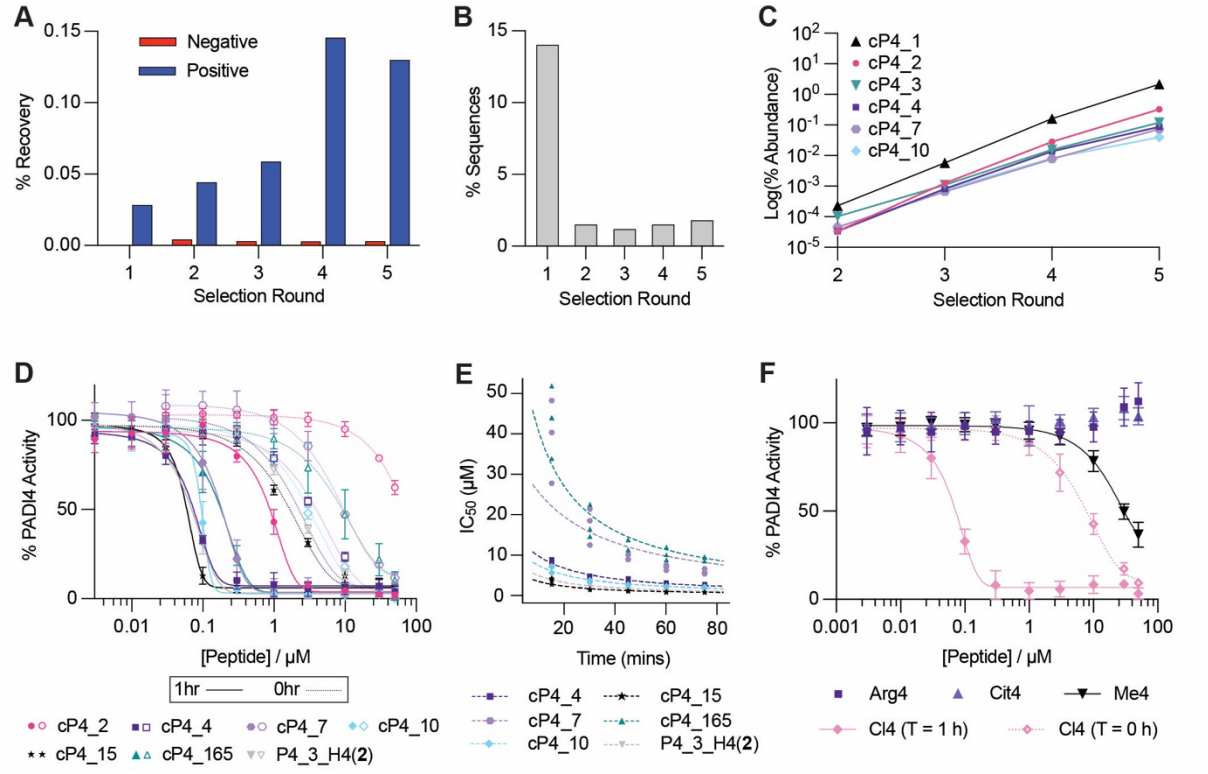

Figure 3: Covalent RaPID selection against PADI4 and peptide characterisation.
A DNA recovery from qPCR after each round of selection from biotinylated beads (negative) and PADI4 beads (positive), compared to the input DNA from each round.
B Enrichment in warhead-containing peptides. Percentage of sequences from each round of selection which do not contain an internal methionine residue, which encodes the FAO warhead.
C Enrichment of 6 key sequences over the five rounds of selection.
D Inhibition COLDER assays with PADI4 and 6 peptides identified in the selection or PADI4_3_H4(**2**). COLDER assays were performed at different peptide concentration (50 – 0.003 µM) in the presence of 10 mM $CaCl_2$. Data is normalised to activity of PADI4 in the presence of 0.1% DMSO. Data shows mean ± SEM of two independent replicates. Each replicate was done in triplicate.
E COLDER assays to determine K_I and k_{inact}. Apparent IC_{50} values were determined at 15-minute intervals from three independent replicates and the Krippendorff equation was fitted.
F Inhibition COLDER assays with PADI4 and cP4_4 variant peptides where FAO was substituted with variable groups. COLDER assays were performed at different peptide concentration (50 – 0.003 µM) in the presence of 10 mM $CaCl_2$. Data is normalised to activity of PADI4 in the presence of 0.1% DMSO. Data shows mean ± SEM of at least two independent replicates.

Initially, the synthetic peptides were incubated with PADI4 and PADI4_C645A and samples analysed by intact-MS. This confirmed all peptides were binding at Cys645, in the active site of PADI4, without any additional sites of reaction (Figure S8). Next, to determine their potency, IC_{50} values were determined using an established PADI4 activity assay, the Colour Developing Reagent (COLDER) assay, using *N*-α-benzoyl-L-arginine ethyl ester (BAEE) as the substrate.[52] All peptides showed inhibitory activity against PADI4, both with and without one-hour of preincubation between PADI4 and peptide prior to initiating the assay through addition of BAEE (Figure 3D). With preincubation, cP4_15 was the most potent peptide with an IC_{50} of 52 nM, a slight improvement over PADI4_3_H4(**2**) (IC_{50} = 61 nM). The least potent peptide, cP4_2, had an IC_{50} of 870 nM. The same trend in IC_{50} values was observed without

preincubation of the peptides with PADI4, however, the IC_{50} values were much higher (Table 1). This time-dependent improvement in IC_{50} is indicative of covalent inhibition. Therefore, to characterise the covalent behaviour further, kinetic parameters were determined using a modified COLDER assay design. Varied concentrations of each peptide were incubated with PADI4 and BAEE and the reactions quenched at 15-minute time intervals. Apparent IC_{50} values were calculated at each time point. This allowed an IC_{50} vs time correlation to be determined and fitted to the Krippendorff Equation which allows determination of K_I and k_{inact} values (Figure 3E, Table 2).[53] The k_{inact}/K_I values determined were up to 10-fold higher than those previously reported for PADI4 covalent inhibitors.[49]

Table 1: Sequences of peptides synthesised after the first selection, and the rationally designed PADI4_3_H4(**2**). Where (**2**) is the FAO warhead and y is chloroacetyl-D-tyrosine which is cyclised with the cysteine residue in each peptide. Their corresponding IC_{50} values from COLDER assay with or without 1 h preincubation of peptide and PADI4 are also shown. Data shows mean ± SEM of two independent replicates.

PEPTIDE NAME	**SEQUENCE**	**IC_{50} VALUES (μM)**	
		0 h preincubation	1 h preincubation
cP4_2	yIWGL(**2**)D(**2**)SCG	>50	0.87 ± 0.06
cP4_4	ySKYD(**2**)RSPRDCG	4.4 ± 0.07	0.070 ± 0.004
cP4_7	yVYS(**2**)KEWKYCG	8.0 ± 1.2	0.16 ± 0.02
cP4_10	yWY(**2**)NWDFNKRCG	3.1 ± 0.01	0.093 ± 0.002
cP4_15	yLD(**2**)HYSSKLYCG	1.6 ± 0.03	0.052 ± 0.002
cP4_165	yVY(**2**)DCEWINRAG	11.8 ± 4.9	0.17 ± 0.01
PADI4_3_H4(**2**)	yRD(**2**)HYRHPKYCG	2.0 ± 0.01	0.065 ± 0.007

Peptide binding to PADI4 was also characterised by surface plasmon resonance (SPR) (Figure S9). By fitting the data using a two-state reaction model, in which the rate constant for the reverse second step (k_{-2}) was set to zero, K_i and k_{inact} values could be determined (Table 2). In most cases, the k_{inact} values are similar whilst the K_I values from the COLDERs are generally an order of magnitude higher. Despite this, the rank order of peptides by k_{inact}/K_i is similar. Differences are only observed between the three most potent peptides, which is where we anticipate the most error in our fitting for both methods.

Table 2: Kinetic parameters of peptides. Both COLDER and SPR values show mean ± SEM from three independent replicates.

	SPR			COLDERS		
	K_i (μM)	k_{inact} (min^{-1})	k_{inact}/K_i (M^{-1} min^{-1})	K_I (μM)	k_{inact} (min^{-1})	k_{inact}/K_I (M^{-1} min^{-1})
cP4_2	1.3 ± 0.5	0.12 ± 0.1	92000	-	-	-
cP4_4	0.12 ± 0.01	0.11 ± 0.1	909000	1.9 ± 0.1	0.13 ± 0.01	74000
cP4_7	1.6 ± 0.1	0.55 ± 0.02	343000	4.3 ± 1.1	0.077 ± 0.052	18000
cP4_10	0.12 ± 0.1	0.23 ± 0.01	2007000	1.5 ± 0.3	0.13 ± 0.03	87000
cP4_15	0.16 ± 0.03	0.25 ± 0.04	1594000	0.78 ± 0.06	0.17 ± 0.02	213000
cP4_165	0.62 ± 0.06	0.066 ± 0.007	107000	8.9 ± 1.1	0.16 ± 0.01	17000
PADI4_3H4(**2**)	0.015 ± 0.002	0.061 ± 0.011	4083000	1.1 ± 0.2	0.16 ± 0.03	149000

To further understand the contribution of **2** to PADI4 binding and inhibition, variants of cP4_4 where **2** was substituted for arginine (Arg4) or citrulline (Cit4) were synthesised and tested. Surprisingly, neither peptide showed any inhibition of PADI4 activity as measured using the COLDER assays (Table 3, Figure 3F). Arg4 did, however, bind reversibly to PADI4 with an affinity of 2.0 μM, as measured by SPR, whilst Cit4 showed negligible binding at the concentrations tested (Figure S10). This is consistent with Arg4 acting as a substrate of PADI4; on binding to PADI4 in the COLDER assays it is converted to Cit4 which no longer binds and hence inhibition is not observed. To assess whether warhead **2** was essential for inhibition, we additionally decided to synthesise the H-amidine analogue, Me4. Me4 had a comparable affinity to Arg4, however unlike Arg4 it also acted as a weak inhibitor of PADI4, consistent with our hypothesis that lack of PADI4 inhibition by Arg4 is due to it being turned over as a substrate (Figure 3F). The reduction in affinity and PADI4 inhibition of Me4 relative to the F-amidine parent cP4_4 suggests that the fluorine atom forms important interactions within the active site of PADI4 that are crucial for binding, as well as acting as the leaving group.

Finally, we made the Cl-amidine analogue (Cl4) to see what effect this more reactive electrophile would have on the potency of the peptide. The IC_{50} values were similar to those of cP4_4, although without preincubation, the IC_{50} was slightly higher (Table 3). Interestingly,

when we determined K_i and k_{inact} values using SPR, Cl4 had a weaker K_i but higher k_{inact} (Figure S10, Table 3). This is consistent with the larger chlorine atom sterically hindering binding, but increasing the rate of the covalent reaction step.[46] This warhead was also confirmed to be more reactive by intact MS, which showed that Cl4 could covalently react twice with PADI4, once at the active site C645 and a second time at an unknown location (Figure S11).

Table 3: Summary of binding affinities and *in vitro* activity of cP4_4 and its variants where the fluoroacetimidoyl ornithine (FAO) warhead was replaced by arginine (Arg4), citrulline (Cit4), acetimidoyl ornithine (Me4) or chloroacetimidoyl ornithine (Cl4). Surface plasmon resonance (SPR) data shows mean ± SEM of three independent replicates. COLDER data shows mean ± SEM of at least two independent replicates.

	SPR				**COLDERS**	
	K_i (μM)	**k_{inact} (min^{-1})**	**k_{inact}/K_i (M^{-1} min^{-1})**	**K_D (μM)**	**IC_{50} (μM) T = 0 h**	**IC_{50} (μM) T = 1 h**
Arg4	-	-	-	2.0 ± 0.2	>100	
Cit4	-	-	-	>10	>100	
Me4	-	-	-	1.7 ± 0.4	>20	
Cl4	4.0 ± 0.2	0.28 ± 0.02	70000	-	8.2 ± 0.8	0.066 ± 0.009
cP4_4	0.12 ± 0.01	0.11 ± 0.003	909000	-	4.4 ± 0.07	0.070 ± 0.004

Although several potent covalent inhibitors had been found, we decided to test whether we could further optimise the selection conditions with the hope of promoting greater discrimination between more and less potent inhibitors. To this end, we repeated the selection starting from round 2, reducing incubation of peptides with PADI4 to only 15 minutes at 0 °C. As we expected, we saw a reduction in positive library recovery which matched the increased stringency, but recovery still increased round-by-round (Figure 4A). After sequencing the recovered libraries, we again observed that the sequences from round 2 onward had a low percentage of sequences which did not encode for a warhead (Figure 4B, Supplementary Data 2). The sequencing results showed the most enriched peptides were those found originally (cP4_4, cP4_7 and cP4_10), but with greater enrichment (Figure 4C). We saw an abolishment from the selection of the poor hit cP4_2 and the peptide hit without the warhead cP4_3. Two further peptides that were uniquely identified in this second screen, cP4_13 and cP4_18, were synthesised. Both were shown to covalently bind to C645 of PADI4 by intact MS (Figure S12). cP4_13 closely resembled cP4_2 from the first selection, but with

only one warhead **2** present. Neither peptide had strong inhibitory activity against PADI4 (Figure S13) and cP4_13 was considerably less active than cP4_2. Characterisation by SPR also showed that these peptides were among some of the poorest binders synthesised (Figure S14). These results showed that the alternative selection conditions did not eradicate the least potent hits, however their enrichment levels were lower relative to the more active cP4_4, cP4_7 and cP4_10, which might have helped with initial peptide selection for synthesis by SPPS.

Given all our identified peptides bound to the active site cysteine of PADI4, in parallel we performed a selection on biotinylated PADI4 C645A (Figure S15), to see if we could promote identification of peptide binders at alternative cysteine residues in PADI4 when the favoured C645 was not available. We chose the same selection conditions as the initial selection to increase our probability of identifying even poor binders. However, there was very minimal enrichment of positive recovery with the PADI4 C645A-bound beads, which was always far exceeded by the negative recovery with biotinylated streptavidin beads (Figure 4D). Nonetheless, we sequenced the recovered libraries. This confirmed that there was no round-by-round enrichment in warhead-containing sequences (Figure 4E, Supplementary Data 3). Of the top 30 most enriched sequences, few contained warhead **2**. Those that did, were also found in previous selections and did not significantly increase in abundance over the course of the selection, suggesting that they are contaminations that bind C645 (Figure 4F). The other sequences most frequently enriched had lost the CGSGSGS linker or contained a very large proportion of Cys residues. There was also a low number of sequences for the later rounds (Figure S16). These factors all indicated that even through presentation on a tight binding cyclic peptide, warhead **2** could not be forced to react at alternative Cys residues in PADI4, hence **2** requires a specifically activated, nucleophilic Cys, like the active site C645.

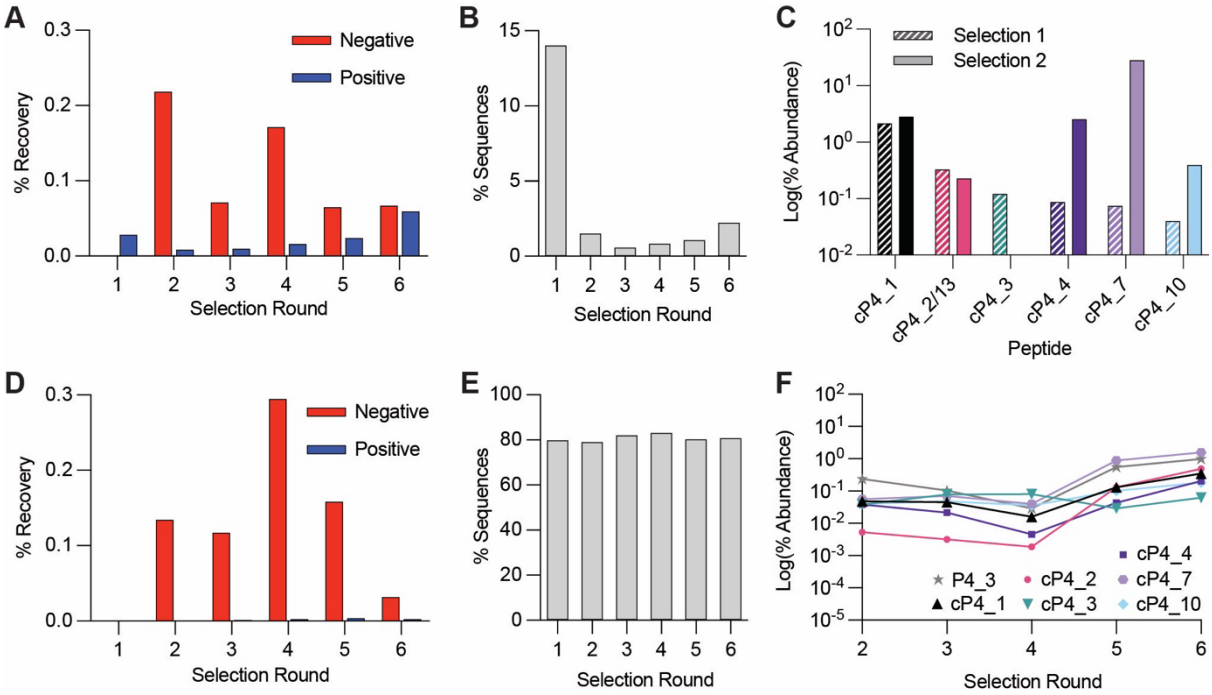

Figure 4: Covalent RaPID selection against PADI4 with increased stringency and against PADI4 C645A.
A and **D** DNA recovery from qPCR after each round of selection against PADI4 and PADI4 C645A, respectively, from biotinylated beads (negative) and PADI4 beads (positive), compared to the input DNA from each round.
B and **E** Enrichment in warhead-containing peptides for PADI4 and PADI4 C645A selections, respectively. Percentage of sequences from each round of selection which do not contain an internal methionine residue, which encodes the FAO warhead.
C Enrichment of key sequences at the 5th round of selection from selection one (hashed bars) and selection two (filled bars), where cP4_13 from selection two has high sequence homology with cP4_2 from selection one.
F Low levels of enrichment of key sequences over the six rounds of selection against PADI4 C645A.

In summary, we show here the development of a RaPID workflow that can select for covalent reaction between peptide and target by directly incorporating an electrophilic warhead into each member of the peptide library, whilst negatively selecting for non-covalent interactions, even if they are low nanomolar binders, like PADI4_3. We have used flexizymes to genetically encode covalency into mRNA display, specifically **2**, a cysteine reactive, fluoroamidine-based warhead. Using this approach we have developed some of the most potent covalent inhibitors of PADI4 identified to date. In the future, these covalent libraries could be used to identify covalent cyclic peptide inhibitors of a range of related therapeutically relevant enzymes, including bacterial arginine deiminases, the cardiovascular target, DDAH, and the inflammatory target, STING, which has previously been shown to react with Cl-amidine.[54–56] More generally, we also envisage this RaPID workflow could be applied with any other covalent warhead which can be loaded using flexizymes to target a wider range of cysteine and non-cysteine residues in therapeutic targets.

Methods

Peptide Synthesis

Peptide synthesis was performed by solid phase peptide synthesis using a Gyros Protein Technologies PreludeX automated synthesizer (Gyros Protein Technologies AB, Sweden). The purity and masses of all peptides was determined using analytical HPLC (Figure S17). Aside from citrulline and D-tyrosine, which were commercially available, unnatural amino acids were made from ornithine residues which were incorporated within the peptide sequence, orthogonally deprotected and reacted with ethyl 2-fluoroethanimidate hydrochloride (Figure S18A-B). In the case of Me4, ornithine was reacted with 2,2,2-trichloroethyl acetimidate hydrochloride (Figure S19) and for Cl4 synthesis, ethyl 2-chloroethanimidate hydrochloride was used (Figure S20). Detailed methods are provided in Supplementary Method S7.

Covalent RaPID

In vitro selections were performed against bio-His-PADI4 and bio-His-PADI4_C645A following previously described protocols. Briefly, initial DNA libraries (including 6-10 degenerate NNK codons in a ratio 0.0018 $NNK_{n=6}$:0.032 $NNK_{n=7}$:1 $NNK_{n=8}$:32 $NNK_{n=9}$:80 $NNK_{n=10}$) (see Table S1 for DNA sequence) were transcribed to mRNA using T7 RNA polymerase (37 °C, 16 h) and ligated to Pu_linker (Table S1) using T4 RNA ligase (30 min, 25 °C). First round translations were performed on a 75 µL scale, with subsequent rounds performed on a 5 µL scale. Translations were carried out (1 h, 37 °C then 12 min, 25 °C) using a custom methionine(-) Flexible *In vitro* Translation system containing additional ClAc-D-Tyr-$tRNA^{fMet}_{CAU}$ (25 µM) and *N*-δ-Fluoroacetimidoyl ornithine-CBT (**3**, 25 µM, Figures S21 and S22). Ribosomes were then dissociated by addition of EDTA (18 mM final concentration, pH 8) and library mRNA reverse transcribed using MMLV RTase, Rnase H Minus (Promega). Reaction mixtures were buffer exchanged into selection buffer (50 mM HEPES, pH 7.5, 150 mM NaLCl, 2 mM DTT, 10 mM $CaCl_2$) using 1 mL homemade columns containing pre-equilibrated Sephadex resin (Cytiva). Blocking buffer was added (1 mg/mL sheared salmon sperm DNA (Invitrogen), 0.1% acetyl-BSA final (Invitrogen)). Libraries were incubated with negative selection beads (3x30 min, 4 °C). Libraries were then incubated with bead-immobilised bio-His-PADI4 or bio-His-PADI4_C645A (200 nM, rt for 1 h or 4 °C for 15 min) before washing (3 x 1 bead volume selection buffer, 4 °C then 3 x 1 bead volume 5 M guanidinium HCl, 4 °C) and elution of retained mRNA/DNA/peptide hybrids in PCR buffer (95 °C, 5 min). Library recovery was assessed by quantitative real-time PCR relative to a library standard, negative selection and the input DNA library. Recovered library DNA was used as the input library for the subsequent round. Following completion of the selections, double indexed libraries (Nextera XT indices) were prepared and sequenced on a MiSeq platform (Illumina) using a v3 chip as single 151

cycle reads. Sequences were ranked by total read numbers and converted into their corresponding peptides sequences for subsequent analysis (Supplementary File 1-3).

Bead preparation:

For PADI4 immobilisation, bio-His-PADI4 or bio-His-PADI4_C645A were incubated with magnetic streptavidin beads (Invitrogen) (4 °C, 15 min to an immobilisation level of 0.9 pmol/µL beads) immediately before use in the selection. Biotin was added to cap unreacted streptavidin sites (25 µM final, 4 °C, 15 min). Beads were washed 3 x 1 bead volume selection buffer and left on ice for use in the selection. Negative beads were prepared similarly except that only selection buffer or selection buffer plus biotin (25 µM) were added to beads and following washing these two variants were mixed.

COLDER assays

PADI4 citrullination activity was analysed using the COLDER assay[52] in 96-well plates. Peptide dilutions were prepared from a 500 µM stock, to give a 10 times concentrated dilution series (500 µM, 300 µM, 100 µM, 30 µM, 10 µM, 3 µM, 1 µM, 0.3 µM, 0.1 µM and 0.03 µM) in COLDER buffer (50 mM HEPES, 150 mM NaCl and 2 mM DTT, pH 7.5) containing 1% DMSO. In triplicate, each was diluted 10-fold further when mixed with 50 nM His-PADI4, 0.6 mg/mL BSA and 10 mM $CaCl_2$, in COLDER buffer. With or without one hour of incubation, 10 mM N^{α}-Benzoyl-L-arginine ethyl ester hydrochloride (BAEE, Merck) was added to initiate the reaction (50 µL final volume). After 30 min at rt, EDTA (50 mM final concentration) was used to quench the reaction and 200 µL of COLDER solution containing 20 mM Diacetyl monoxime/2,3-butanedione monoxime (Merck), 0.5 mM Thiosemicarbazide (Acros Organics), 2.25 M H_3PO_4, 4.5 M H_2SO_4 and 1.5 mM $NH_4Fe(SO_4)_2.12H_2O$ was added to each well. Samples were incubated for 20 min at 95 °C before measuring absorbance at 540 nm on a CLARIOstar Plus (BMG LABTECH). Data analysis was performed with GraphPad Prism. Data are presented as the average ± standard error of the mean from at least two independent replicates.

Incubation time-dependent potency $IC_{50}(t)$:

To determine K_I and k_{inact} of covalent peptide inhibitors, COLDER assays were used. Peptide dilutions were prepared using 5-fold dilutions from 500 µM and 300 µM to 0.8 µM and 2.4 µM, respectively, at 1 % DMSO in COLDER buffer (50 mM HEPES, 150 mM NaCl and 2 mM DTT, pH 7.5). The 9 peptide dilutions were added to a 96-well plate, alongside a 1 % DMSO control. An equal volume of 10 mM BAEE was added to each well and the solution was homogenised by pipette mixing. This was mixed with 50 nM His-PADI4, 0.6 mg/mL BSA and 10 mM $CaCl_2$, in COLDER buffer, to bring the final volume to 300 µL in each well. At 15-minute intervals, for 5 timepoints, 50 µL of solution was taken from each well and quenched with 10 µL EDTA (300 mM). 200 µL of COLDER solution containing 20 mM Diacetyl

monoxime/2,3-butanedione monoxime (Merck), 0.5 mM Thiosemicarbazide (Acros Organics), 2.25 M H_3PO_4, 4.5 M H_2SO_4 and 1.5 mM $NH_4Fe(SO_4)_2.12H_2O$ was added to each well. Samples were incubated for 20 min at 95 °C before measuring absorbance at 540 nm on a CLARIOstar Plus (BMG LABTECH). IC_{50} values for each time point were determined using non-linear regression with GraphPad Prism. Incubation time–dependent potency $IC_{50}(t)$ against incubation time was fitted to the Krippendorff equation (below) to determine K_I and k_{inact} using a Python script.[53]

$$\mathrm{IC_{50}}(t) = \mathrm{K_I}\left(1 + \frac{\mathrm{S}}{\mathrm{K_M}}\right) \cdot \left(\frac{2 - 2\mathrm{e}^{-\eta_{\mathrm{IC_{50}}} \cdot \mathrm{k_{inact}} \cdot \mathrm{t}}}{\eta_{\mathrm{IC_{50}}} \cdot \mathrm{k_{inact}} \cdot \mathrm{t}} - 1\right)$$

Where $$\eta_{\mathrm{IC_{50}}} = \frac{\mathrm{IC_{50}(t)}}{\mathrm{K_I}\left(1+\frac{\mathrm{S}}{\mathrm{K_M}}\right) + \mathrm{IC_{50}(t)}}$$

Acknowledgements

This work was supported by funding from the UKRI [grant number EP/X020878/1]. This work was also supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (CC2030), the UK Medical Research Council (CC2030), and the Wellcome Trust (CC2030). We would like to thank the Crick Advanced Sequencing Science Technology Platform for assistance with next generation sequencing and Sarah Maslen of the Crick Proteomics Science Technology Platform for performing the intact mass spectrometry for this work. Additional thanks to the Crick Chemical Biology Science Technology Platform for their advice about peptide synthesis, especially Dhira Joshi and to Samrah Bourhan for her help during her BSc project. Illustrative figures were created using BioRender.com.

Author contributions

I.R.M. and E.D.D.C. planned and executed experiments and analysed data; S.K. analysed SPR data; L.J.W conceptualised the project, obtained funding and supervised the work; I.R.M and L.J.W wrote the manuscript with help from all authors.

Conflict of Interest

The authors declare no conflicts of interest.

Data availability

Detailed Supplementary Methods and Supplementary Figures are provided in the Supplementary Information. Sequencing data are provided in Supplementary Data 1-3.

References

1. Bauer, R. A. Covalent inhibitors in drug discovery: from accidental discoveries to avoided liabilities and designed therapies. *Drug Discov. Today* **20**, 1061–1073 (2015).
2. Gehringer, M. & Laufer, S. A. Emerging and Re-Emerging Warheads for Targeted Covalent Inhibitors: Applications in Medicinal Chemistry and Chemical Biology. *J. Med. Chem.* **62**, 5673–5724 (2019).
3. Lanning, B. R. *et al.* A road map to evaluate the proteome-wide selectivity of covalent kinase inhibitors. *Nat. Chem. Biol.* **10**, 760–767 (2014).
4. Singh, J., Petter, R. C., Baillie, T. A. & Whitty, A. The resurgence of covalent drugs. *Nat. Rev. Drug Discov.* **10**, 307–317 (2011).
5. Flanagan, M. E. *et al.* Chemical and Computational Methods for the Characterization of Covalent Reactive Groups for the Prospective Design of Irreversible Inhibitors. *J. Med. Chem.* **57**, 10072–10079 (2014).
6. Glas, A. *et al.* Constrained Peptides with Target-Adapted Cross-Links as Inhibitors of a Pathogenic Protein–Protein Interaction. *Angew. Chem. Int. Ed.* **53**, 2489–2493 (2014).
7. Alleyne, C. *et al.* Series of Novel and Highly Potent Cyclic Peptide PCSK9 Inhibitors Derived from an mRNA Display Screen and Optimized via Structure-Based Design. *J. Med. Chem.* **63**, 13796–13824 (2020).
8. Nomura, K. *et al.* Broadly Applicable and Comprehensive Synthetic Method for N-Alkyl-Rich Drug-like Cyclic Peptides. *J. Med. Chem.* **65**, 13401–13412 (2022).
9. Tanada, M. *et al.* Development of Orally Bioavailable Peptides Targeting an Intracellular Protein: From a Hit to a Clinical KRAS Inhibitor. *J. Am. Chem. Soc.* **145**, 16610–16620 (2023).

10. Dougherty, P. G., Sahni, A. & Pei, D. Understanding Cell Penetration of Cyclic Peptides. *Chem. Rev.* **119**, 10241–10287 (2019).
11. Gentilucci, L., De Marco, R. & Cerisoli, L. Chemical modifications designed to improve peptide stability: incorporation of non-natural amino acids, pseudo-peptide bonds, and cyclization. *Curr. Pharm. Des.* **16**, 3185–3203 (2010).
12. Stebbins, J. L. *et al.* Structure Based Design of Covalent Siah Inhibitors. *Chem. Biol.* **20**, 973–982 (2013).
13. Huhn, A. J., Guerra, R. M., Harvey, E. P., Bird, G. H. & Walensky, L. D. Selective Covalent Targeting of Anti-Apoptotic BFL-1 by Cysteine-Reactive Stapled Peptide Inhibitors. *Cell Chem. Biol.* **23**, 1123–1134 (2016).
14. Yoo, D. Y., Hauser, A. D., Joy, S. T., Bar-Sagi, D. & Arora, P. S. Covalent Targeting of Ras G12C by Rationally Designed Peptidomimetics. *ACS Chem. Biol.* **15**, 1604–1612 (2020).
15. Petri, L. *et al.* A covalent strategy to target intrinsically disordered proteins: Discovery of novel tau aggregation inhibitors. *Eur. J. Med. Chem.* **231**, 114163 (2022).
16. Huang, Y., Wiedmann, M. M. & Suga, H. RNA Display Methods for the Discovery of Bioactive Macrocycles. *Chem. Rev.* **119**, 10360–10391 (2019).
17. Suga, H. Max-Bergmann award lecture:A RaPID way to discover bioactive nonstandard peptides assisted by the flexizyme and FIT systems. *J. Pept. Sci.* **24**, e3055 (2018).
18. Heinis, C. & Winter, G. Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.* **26**, 89–98 (2015).
19. Obexer, R., Walport, L. J. & Suga, H. Exploring sequence space: harnessing chemical and biological diversity towards new peptide leads. *Curr. Opin. Chem. Biol.* **38**, 52–61 (2017).
20. 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).

21. Dai, S. A. *et al.* State-selective modulation of heterotrimeric Gαs signaling with macrocyclic peptides. *Cell* **185**, 3950-3965.e25 (2022).
22. Patel, K. *et al.* Cyclic peptides can engage a single binding pocket through highly divergent modes. *Proc. Natl. Acad. Sci.* **117**, 26728–26738 (2020).
23. Goto, Y. & Suga, H. The RaPID Platform for the Discovery of Pseudo-Natural Macrocyclic Peptides. *Acc. Chem. Res.* **54**, 3604–3617 (2021).
24. Iskandar, S. E. *et al.* Identification of Covalent Cyclic Peptide Inhibitors in mRNA Display. *J. Am. Chem. Soc.* **145**, 15065–15070 (2023).
25. Wu, Y. *et al.* Identification of photocrosslinking peptide ligands by mRNA display. *Commun. Chem.* **6**, 103 (2023).
26. McCarthy, K. A. *et al.* Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **140**, 6137–6145 (2018).
27. Chen, S. *et al.* Identification of highly selective covalent inhibitors by phage display. *Nat. Biotechnol.* **39**, 490–498 (2021).
28. Zheng, M. & Gao, J. Phage Display of Two Distinct Warheads to Inhibit Challenging Proteins. *ACS Chem. Biol.* **18**, 2259–2266 (2023).
29. Zheng, M. *et al.* Lysine-Targeted Reversible Covalent Ligand Discovery for Proteins via Phage Display. *J. Am. Chem. Soc.* **144**, 15885–15893 (2022).
30. Hecht, S. M., Alford, B. L., Kuroda, Y. & Kitano, S. “Chemical aminoacylation” of tRNA’s. *J. Biol. Chem.* **253**, 4517–4520 (1978).
31. Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A General Method for Site-specific Incorporation of Unnatural Amino Acids into Proteins. *Science* **244**, 182–188 (1989).

32. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.* **7**, 28–33 (2000).
33. Murakami, H., Ohta, A., Ashigai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **3**, 357–359 (2006).
34. Goto, Y., Katoh, H & Suga, H. Preparation of materials for flexizyme reactions and genetic code reprogramming. *Nat. Protoc. Exch.* **6**, 779–790 (2011).
35. Yamagishi, Y. *et al.* Natural Product-Like Macrocyclic N-Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. *Chem. Biol.* **18**, 1562–1570 (2011).
36. Goto, Y. *et al.* Reprogramming the Translation Initiation for the Synthesis of Physiologically Stable Cyclic Peptides. *ACS Chem. Biol.* **3**, 120–129 (2008).
37. Saito, H., Kourouklis, D. & Suga, H. An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20**, 1797–1806 (2001).
38. Vossenaar, E. R., Zendman, A. J. W., van Venrooij, W. J. & Pruijn, G. J. M. PAD, a growing family of citrullinating enzymes: genes, features and involvement in disease. *BioEssays* **25**, 1106–1118 (2003).
39. Li, P. *et al.* Regulation of p53 Target Gene Expression by Peptidylarginine Deiminase 4. *Mol. Cell. Biol.* (2008) doi:10.1128/MCB.01747-07.
40. Stadler, S. C. *et al.* Dysregulation of PAD4-mediated citrullination of nuclear GSK3β activates TGF-β signaling and induces epithelial-to-mesenchymal transition in breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11851–11856 (2013).
41. Wang, S. & Wang, Y. Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis. *Biochim. Biophys. Acta* **1829**, 1126 (2013).

42. Kolodziej, S. *et al.* PADI4 acts as a coactivator of Tal1 by counteracting repressive histone arginine methylation. *Nat. Commun.* **5**, 3995 (2014).
43. Jones, J., Causey, C., Knuckley, B., Slack-Noyes, J. L. & Thompson, P. R. Protein arginine deiminase 4 (PAD4): current understanding and future therapeutic potential. *Curr. Opin. Drug Discov. Devel.* **12**, 616–627 (2009).
44. Witalison, E. E., Thompson, P. R. & Hofseth, L. J. Protein Arginine Deiminases and Associated Citrullination: Physiological Functions and Diseases Associated with Dysregulation. *Curr. Drug Targets* **16**, 700–710 (2015).
45. Suzuki, A. *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
46. Luo, Y. *et al.* Inhibitors and Inactivators of Protein Arginine Deiminase 4: Functional and Structural Characterization. *Biochemistry* **45**, 11727–11736 (2006).
47. Luo, Y., Knuckley, B., Lee, Y.-H., Stallcup, M. R. & Thompson, P. R. A fluoroacetamidine-based inactivator of protein arginine deiminase 4: design, synthesis, and in vitro and in vivo evaluation. *J. Am. Chem. Soc.* **128**, 1092–1093 (2006).
48. Knuckley, B. *et al.* Substrate specificity and kinetic studies of PADs 1, 3, and 4 identify potent and selective inhibitors of Protein Arginine Deiminase 3. *Biochemistry* **49**, 4852–4863 (2010).
49. Jones, J. E. *et al.* Synthesis and Screening of a Haloacetamidine Containing Library To Identify PAD4 Selective Inhibitors. *ACS Chem. Biol.* **7**, 160–165 (2012).
50. Iwane, Y., Kimura, H., Katoh, T. & Suga, H. Uniform affinity-tuning of N-methyl-aminoacyl-tRNAs to EF-Tu enhances their multiple incorporation. *Nucleic Acids Res.* **49**, 10807–10817 (2021).

547 51. Bertran, M. T. *et al.* A cyclic peptide toolkit reveals mechanistic principles of
548 peptidylarginine deiminase IV (PADI4) regulation. 2023.12.12.571217 Preprint at
549 https://doi.org/10.1101/2023.12.12.571217 (2023).
550 52. Knipp, M. & Vašák, M. A Colorimetric 96-Well Microtiter Plate Assay for the
551 Determination of Enzymatically Formed Citrulline. *Anal. Biochem.* **286**, 257–264 (2000).
552 53. Krippendorff, B.-F., Neuhaus, R., Lienau, P., Reichel, A. & Huisinga, W. Mechanism-Based
553 Inhibition: Deriving KI and kinact Directly from Time-Dependent IC50 Values. *J. Biomol.*
554 *Screen.* **14**, 913–923 (2009).
555 54. Jones, J. E., Dreyton, C. J., Flick, H., Causey, C. P. & Thompson, P. R. Mechanistic Studies
556 of Agmatine Deiminase from Multiple Bacterial Species. *Biochemistry* **49**, 9413–9423
557 (2010).
558 55. Stone, E. M., Schaller, T. H., Bianchi, H., Person, M. D. & Fast, W. Inactivation of Two
559 Diverse Enzymes in the Amidinotransferase Superfamily by 2-Chloroacetamidine:
560 Dimethylargininase and Peptidylarginine Deiminase. *Biochemistry* **44**, 13744–13752
561 (2005).
562 56. Humphries, F. *et al.* Targeting STING oligomerization with small-molecule inhibitors.
563 *Proc. Natl. Acad. Sci.* **120**, e2305420120 (2023).
564