Structural diversity in de novo cyclic peptide ligands from genetically encoded library technologies

Tom E. McAllister¹  |  Oliver D. Coleman¹  |  Grace Roper¹,²  |  Akane Kawamura¹,²

¹Chemistry – School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, UK
²Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Oxford, UK

Correspondence
Tom E. McAllister and Akane Kawamura, Chemistry – School of Natural and Environmental Sciences, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK.
Email: tom.mcallister@newcastle.ac.uk (T. E. M.) and akane.kawamura@newcastle.ac.uk (A. K.)

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/R512333/1; H2020 European Research Council, Grant/Award Number: Starting Grant / EPITOOLS-679479

Abstract
Cyclic peptides discovered by genetically encoded library technologies have emerged as a class of promising molecules in chemical biology and drug discovery. Here we review the cyclic peptides identified through these techniques reported in the period 2015 to 2019, with a particular focus on the three-dimensional structures that peptides adopt when binding to their targets. A range of different structures have been revealed through co-crystal structures, highlighting how versatile and adaptable these molecules are in binding to diverse protein targets, such as enzymes and receptors, or challenging shallow surfaces involved in protein-protein interfaces. Analysis of the properties of the peptides reported shows some interesting trends, with further insight for those with structural information suggestive that larger peptides are more likely to adopt secondary structure. We highlight examples where co-crystal structures have informed the key interactions that promote high affinity and selectivity of cyclic peptides against their targets, identified novel inhibitor binding sites, and provided new insights into the biology of their targets. The structure-guided modifications have also aided the design of cyclic peptides with improved activity and physicochemical properties. These examples highlight the importance of crystallography in future cyclic peptide drug discovery initiatives.

KEYWORDS
chemical biology, crystal structures, cyclic peptides, display technologies, drug discovery

1 | INTRODUCTION

Cyclic peptides (CPs) are an emerging class of molecules that occupy the ‘Goldilocks space between small molecules and large biologics’,¹ with the potential to combine the best attributes of antibodies (high specificity and affinity) and small molecules (bioavailability and pharmacokinetics). Thus CPs have recently been attracting significant attention for use in a therapeutic setting²,³ and as tools for chemical biology owing to the high specificity and high affinities that they can achieve against a wide range of targets. One of the particularly intriguing properties of CPs is the diversity of the chemical space they can cover and three-dimensional structures they are able to adopt in order to achieve binding to their target protein. Unlike linear peptides, which are conformationally flexible and highly dynamic, the constraint of cyclisation in CPs restricts the number of conformations the peptide chain can adopt, thus reducing the entropic penalty for binding to its target. Small molecules, on the other hand, are typically very conformationally constrained, making them ideal for targeting pockets and presenting binding groups in a defined arrangement, but are limited when targeting large, shallow protein surfaces often considered to be ‘undruggable’ biological interaction space, such as protein-protein interactions (PPIs).
There are many examples of bioactive CPs found in nature; the fungal natural product cyclosporine is used as an immunosuppressant while phalloidin and α-amanitin are highly toxic to humans. Defined biological activities for these peptides have been elucidated, but rational de novo design of such molecules with desirable properties against targets of interest is highly challenging. It is thus not surprising, that the richest source of CPs (aside from those found naturally occurring) is through genetically encoded library technologies, where diverse starting pools of combinatorial peptides that cover vast chemical space can rapidly be generated and screened to identify CPs that efficiently bind to a target of interest.

This review will survey CPs that have been identified through genetically encoded peptide selection techniques, with particular focus on the CPs where three-dimensional structures in complex with their targets have been determined. We provide an update on previous reviews on de novo CP structures focusing on the last 5 years (2015-2019) (see Malde et al. for a meta-analysis of crystal structures of bound CPs within the protein data bank [PDB]). We will not cover peptides that are designed to be α-helical, such as stapled peptides, since, by definition, these will form an α-helix and thus have minimal variation in their structures.

2 | METHODS

Identifying Relevant Journal Articles: An initial list of articles was obtained by searching www.ncbi.nlm.nih.gov/pmc with the following term: 'cyclic peptide'[Body - All Words] AND display[Body - All Words] AND selection[Body - All Words] AND (January 1, 2015)[PubDate]: December 31, 2019 [PubDate]). Articles were further manually filtered on the basis of being a CP discovered by a genetically encoded library technique against a protein target; iterations or previous peptides were not included unless either a selection process was used or the parent sequence was also first reported within the same period, that is, 2015 to 2019. Further articles were subsequently included, identified from works within articles returned by the search terms or serendipitous discovery.

Calculating Cyclic Peptide Properties: Peptide sequences from those listed in Table 1 were extracted from published articles and redrawn in ChemDraw (v19) and used as the input for DataWarrior to calculate cLogP, the number of H-bond acceptors and donors, total solvent-accessible surface area (SASA) and polar surface area. A selection of natural product (like) peptide drugs used in this analysis were: cyclosporin, octreotide, pasireotide, linoleotide, romidepsin, and lanreotide.

Binding constants (Kd) were converted to pKd (−log(Kd)) where data was available. The binding efficiency index (BEI) was calculated by dividing the pKd of a CP by its molecular weight in kDa. Graphs were generated using GraphPad Prism v8.

Data extracted from PDB files: To determine the interacting area of the CPs with their respective proteins from crystal structures, first the solvent-accessible surface area (SASA) was calculated for the CP-protein complex in PyMOL using the get_area command with dot_solv set to 1 and all 'ignore' flags removed (using the command 'flag ignore, all, clear'). SASA was calculated for each of the interacting species separately that is, the CP and the protein, with the interacting surface area calculated using Equation (1). The CP and protein interface was assumed to be complementary.

$$\text{Interacting area} = \frac{\text{SASA}_{\text{CP}} + \text{SASA}_{\text{rot}} - \text{SASA}_{\text{complex}}}{2}$$ (1)

The number of hydrogen bonds was counted manually from the find polar contacts function in PyMOL. The percentage of peptide containing secondary structure was determined by the flags associated to the atoms in the PyMOL object based on the following commands:

```
> ss = [i.ss for i in cmd.get_model('object').atom]
> print('Helix content: %5.2f%% %%(100.0* ss.count('H')/len(ss))
> print('Sheet content: %5.2f%% %%(100.0* ss.count('S')/len(ss))
```

3 | CYCLIC PEPTIDE SELECTION: GENETICALLY ENCODED LIBRARY TECHNOLOGIES

The three most common selection methods for discovering de novo CPs against protein targets of interest are: (a) phage display, (b) mRNA display, and (c) split-intein circular ligation of peptides and proteins (SICLOPPS). In all cases, a diverse library of combinatorial peptides are ribosomally synthesised from a highly degenerate DNA/RNA template, post-translationally cyclised and screened against a target protein (or PPI). After rounds of selections, enriched peptide sequences are decoded via sequencing of their associated genetic tags. CP hits are then chemically synthesised, and their affinity for their target confirmed via biochemical and biophysical assays. Here we provide a very brief overview of each CP selection methods (see for a detailed review).

In phage display, a library of DNA sequences is ligated into a phagemid and transformed into bacteria, which produce phage displaying a single member of the library on its surface as fusion to a coat protein. Peptides are (almost always) cyclised by cysteine disulfides or chemical modification of cysteine residues with linkers. The pool of phage is then incubated with an immobilised target protein, and binding phage are recovered and used as input for a further round of enrichment. Starting library sizes are typically around 10^6 sequences, a limitation imposed by the efficiency of transformation.

In mRNA display, a puromycin-linked mRNA library is translated in vitro to form mRNA-peptide fusions with covalent linkage of the puromycin to the nascent peptide. The mixture is then applied to an immobilised target protein to recover binding sequences. The mRNA is reverse transcribed into DNA for PCR amplification. The recovered DNA is transcribed to yield an enriched mRNA library, which is further used as input for the next round of enrichment. As
this method is in vitro and cell-free, there is no upper limit on the starting library size, with initial diversity >10^12 routinely used (with the only potential limitation being the number of unique molecules of library DNA that can be practically used). mRNA display was further advanced by codon reprogramming and incorporation of non-canonical amino acids (ncAA) into peptides. This has been achieved through engineered aminoacyl-tRNA synthetases or by using a set of ribozymes (termed flexizymes). An integrated system using flexizymes coupled to mRNA-display is referred to as the Random nonstandard Peptide Integrated Discovery (RaPID) system, which has been widely used to incorporate ncAA for peptide cyclisation, as well as for active-site targeting warheads, backbone modifications and exotic side-chain modifications.

In SICLOPPS, a DNA library is transformed into a host cell (either prokaryotic or eukaryotic) where the peptide is initially synthesised as part of a split-intein fusion protein, further processed when the intein fragment forms the active intein and released as a head-to-tail part of a split-intein fusion protein, further processed when the intein prokaryotic or eukaryotic) where the peptide is initially synthesised as

Interestingly, the molecular weights (MW) of cyclic peptides (CPs) identified during the reporting period cover a relatively wide range, from 443 to 2717 Da (Figure 1A). CPs from different selection methods appear to cluster together, with the median MW for phage-display, mRNA-display and SICLOPPS at 1430, 1767, and 769 Da, respectively, but this likely reflects the library design rather than the method used. For example, many phage-display derived peptides are from a common disulfide-linked library (CX-C) and cluster in the 1100 to 1500 Da region. Similarly, SICLOPPS derived peptides cluster at 500 to 1000 Da for head-to-tail cyclised hexamers, although an outlier is found at ~2500 Da for a significantly larger lanthipeptide scaffold. We are not aware of any comparative discovery efforts using the same library design with different techniques.

The DataWarrior software suite was used to calculate physicochemical properties, as it is an open-source data analysis program and widely accessible, although some functions are not optimised for use with large complex molecules such as CPs. Perhaps unsurprisingly, the calculated total surface area (cPSA) and total number of hydrogen bond donors (HBD) and acceptors (HBA) correlate with MW (Figure 1B,C), whilst the calculated lipophilicities of the peptides cover a wide range (Figure 1D). It should also be noted that these parameters may not tell the whole story as CPs can have unexpected behaviours such as so-called chameleonic peptides and perhaps the perceived drive to increase lipophilicity for better passive membrane diffusion must be balanced against lipophilic liabilities. The calculated properties of some CPs are comparable to calculated physicochemical properties for established peptide drugs (cyclosporin, romidepsin, octreotide, pasireotide, linaclotide and lanreotide; Figure 1A-D) and the recently reviewed physicochemical properties for oral peptide-based drugs (MW ≤ 1200 Da, −5 ≤ LogP > 8, up to 50 HBA and 25 HBD, TPSA ≤ 400 Å²). Additionally, the affinities of the surveyed CPs (where data is reported) range from pKD 3.8 to 9.6 (Figure 1E) and, whilst low affinities (pKD < 5) are observed for low molecular weight CPs (MW < 1000), no correlation is obvious for higher MWs (1000-2500) and their affinities (Figure 1F). The binding efficiency index (BEI, given as a binding affinity relative to MW) is however relatively similar across the selection methods in this dataset, suggesting comparable binding efficiencies can be obtained using these methodologies (Figure 1G). No clear trends were observed for cLogP as a function of molecular weight, with most clustered in the +5 to −10 region (and outliers at >2500 Da either FLAG-tagged or modified with additional arginine residues; Figure 1J), and for cPSA and cLogP as a function of pKD (Figure 1KL).

4 | REPORTED DE NOVO CYCLIC PEPTIDES AGAINST PROTEIN TARGETS

In the last 5 years (2015-2019), there have been over 40 publications describing the identification of de novo CP binding sequences using various display technologies against protein targets of interest (Table 1). Target classes include enzymes (e.g., proteases, epigenetic proteins and metabolic enzymes), transcription factors, signalling ligands and receptors (or soluble portions thereof), with a wide range of ligands, including orthosteric/allosteric inhibitors and binders, demonstrating the breadth and diversity of the targets and sites that can be accessed through CP encoded library technologies. An analysis of the properties calculated based on the chemical structures and co-crystal structures with their target proteins provide some insight into the overall trends (or lack of) for the reported de novo CPs from 2015 to 2019.

4.1 | Properties of de novo cyclic peptides

Interestingly, the molecular weights (MW) of de novo CPs identified during the reporting period cover a relatively wide range, from 443 to 2717 Da (Figure 1A). CPs from different selection methods appear to cluster together, with the median MW for phage-display, mRNA-display and SICLOPPS at 1430, 1767, and 769 Da, respectively, but this likely reflects the library design rather than the method used. For example, many phage-display derived peptides are from a common

4.2 | Trends in cyclic peptide-protein complex structures

A total of 10 X-ray crystal structures of de novo CPs in complex with their target proteins have been reported in the period 2015 to 2019, with 9 more prior to this which have been included in the subsequent analysis. These co-crystal structures provide important insight into how CPs bind to their targets and various properties based on their crystal structures are summarised in Table 2. Of the 19 X-ray co-crystal structures from the PDB, 8 are from phage-display and 11 from mRNA-display against 4 and 7 different protein targets respectively (no de novo CP complex crystal structures have been reported using other display methods). Six CPs contain some degree of helical secondary structure, 6 contain sheets and 7 have neither, with the average proportion of the peptide adopting each secondary structure (for
| Target | Abbreviation | Display method | Hit peptide(s) and derivatives | Types of ligands identified | Reference(s) |
|--------|--------------|----------------|-------------------------------|-----------------------------|--------------|
| αvβ3 integrin | yeast | lactacin 481-HVRGDN | Binder | [26] |
| Amyloid-β peptide | Aβ42 | SICLOPPS | AβC5-34, AβC5-116, AβC7-1, AβC7-14 | Inhibitors of Aβ42 aggregation | [27,28] |
| B-Cell lymphoma 6 transcription factor | BCL6 | SICLOPPS | cyclo-CIYYCV | PPI inhibitor | [29] |
| Cofactor independent phosphoglycerate mutase | iPGM | mRNA | Ce-2d | Allosteric inhibitor | [30] |
| Human Cu/Zn superoxide dismutase (A4V) | SOD1(A4V) | SICLOPPS | SOD1C5-4 | Inhibitor of misfolding and aggregation | [27] |
| Factor XII | FXII | phage | FXII801 | Active site inhibitor | [31] |
| Fibroblast growth factor receptor 1 | FGFR1 | phage | F8, G10 | Signalling axis inhibitor | [32] |
| Glucagon-like peptide-1 receptor | GLP-1R | Cell surface display | P1, P10 | Agonist | [33] |
| Hepatocyte growth factor | HGF | mRNA | HIP-8 | Allosteric inhibitor | [34] |
| Hepatocyte growth factor receptor | HGF/Met | mRNA | aMD4 | Agonist | [35] |
| Histone deacetylase 8 | HDAC8 | phage | CycH8a | Active site inhibitor | [36] |
| Histone demethylase 4A | KDM4A | mRNA | CP2, CP2.3, CP2r | Active site inhibitor | [37-39] |
| Human epidermal growth factor receptor | hEGFR | mRNA | CbaP5 | Binder | [40] |
| Human pancreatic α-amylase | HPA | mRNA | pHA-Dm, pHA-L5(d10Y) | Active site inhibitor | [41-43] |
| Human tumour necrosis factor-α | hTNF-α | phage | M21 | Allosteric PPI | [44] |
| Inducible degrader of low density lipoprotein receptor | IDOL | SICLOPPS | RINGpep2 | PPI inhibitor | [45] |
| Insulin-degrading enzyme | IDE | phage | P12-3A, C7C-1 | Inhibitor | [46] |
| Interleukin 6 receptor | IL6R | mRNA | ul66R-1 | Binder | [47] |
| Intestinal alpha-glucosidase | phase | C1, C2 | Binders, C2-Inhibitor | [48] |
| K48-linked Ubiquitin | K48Ub2 | mRNA | Ub4ix | Binder | [49] |
| L-glutamine-binding periplasmic protein QBP | QBP | phage and yeast | D6, F5 | Binder | [50] |
| Matrix metalloproteinase 2 | MMP2 | phage | M219α | Active site inhibitor | [51] |
| Midkine | MK | phage | MK-P1 | Binder | [52] |
| Plexin B1 | PlxnB1 | mRNA | PB1m6 | Allosteric PPI | [53,54] |
| Prolyl-hydroxylase domain-containing protein 2 | PHD2 | mRNA | PHD2_3C | Allosteric | [55] |
| Rabbit anti-goat IgG (H + L) superclonal secondary antibody | phase | AB1-18 | Binder | [56] |
| Streptavidin | mRNA | Flag-cyclo(Strep2), Flag-cyclo(Strep3) | Binder | [24] |
| Ten-eleven translocation 1 | TET1 | mRNA | TIP1 | Active site inhibitor | [57] |
| Tobacco Etch Virus protease | TEV | phage | CycTev1/2 | Binder | [34] |
| Transforming growth factor-β 1-stimulated clone 22 homologous gene-1 | THG-1 | mRNA | R4-1B | Binder | [58] |
| Tumour suppressor gene 101 | TSG101 | UEV | SICLOPPS | XY3-3, KRL74 | PPI inhibitor | [59,60] |
| Urokinase-type plasminogen activator | uPA | phase | IG1,IG2, PEP35 | Inhibitor | [61,62] |
| V-Ki-ras2 Kirsten rat sarcoma viral oncogene homologue G12D | K-Ras(G12D) | phase | KRpep-2d | Allosteric PPI | [25,63,64] |
| Zika virus NSB2-NS3 protease | ZIKVpro | mRNA | Compound 2 | Allosteric inhibitor | [65] |

**Note:** A selection of de novo CPs reported between 2015 and 2019 are listed in an alphabetical order of the targets. In cases with multiple peptides were identified, the most well characterised peptides or tight binders/inhibitors have been given as representative examples where possible.
those that have any) 53.8 ± 21.3% for helices and 41.8 ± 17.5% for sheets. As highlighted in the previous section, while peptides have been grouped by selection method, other factors may be responsible for the differences observed - for example library size, cyclisation method or target type.

Extracting information from their PDB files, overall, both the solvent accessible surface area (SASA) and the interaction area between CP and protein are very consistent across the two selection methods. The average SASAs are 1611 ± 318 Å² and 1603 ± 310 Å² for phage and mRNA respectively, with corresponding interaction areas of 758 ± 91 Å² (49 ± 13% of CP SASA) and 750 ± 136 Å² (49 ± 13%) (Figure 2A,B) despite a large range within the dataset. For example, for mRNA-display the percentage of the CP SASA interacting with target varies from 25-70% (414-923 Å², Figure 2). The crystallised CPs (Figure 1A-E, open symbols) do not appear to have any calculated physicochemical traits that are distinct from the larger data set of all reported CPs (Figure 1A-D); however, pK_D analysis (where data is available) shows CPs that have crystallised between 2015 and 2019 have pK_D > 7, (Figure 1E). This suggests high target affinity may aid CP-protein crystallization.

The molecular weights of the CPs differ significantly between the two selection methods, with CPs identified through mRNA-display typically 22% larger: 1918 ± 272 Da for mRNA-displayed CPs and 1571 ± 415 Da for phage-displayed CPs, reflecting the differences in the library size (Figures 1A and 2C). Similar SASA with different MW suggests mRNA CPs are likely to be more ‘folded’ i.e., that a larger proportion of the molecule is not accessible to solvent in the conformation found in the crystal structure of the complex. A ‘compactness’ score, normalizing molecular weight by SASA (so molecules with large molecular weight but low SASA would give high ‘compactness’ scores), gives phage CPs a score of 0.98 ± 0.33 and mRNA CPs 1.20 ± 0.24. This notion of compactness/folding is supported by the analysis of secondary structural elements within the CPs; 3 out of 8 (37.5%) phage CPs have secondary structure (2 helix, 1 sheet) while

![FIGURE 1](image-url)
for mRNA CPs it is 9 out of 11 (82%; 4 helix, 5 sheet). Additionally, when considering the relationship between some of the calculated properties relating to ‘size’ (total surface area and polar surface) and the observed SASA from the crystal structures, there is a strong linear relationship for the phage ($R^2 = 0.94$ and 0.89; Figure 2D,E), whereas the mRNA CPs show poor correlations ($R^2 = 0.47$ and 0.54 respectively). Since the property calculations we have employed are not optimised for application to large, complex molecules like CPs which may have chameleonic behaviours linked to intramolecular interactions, these deviations are to be expected: the phage CPs do not tend to be as compact and if they form more extended, unstructured conformations the calculated properties would be expected to correlate well. Since more of the mRNA CPs have secondary structures (82% vs 37.5%) and a higher ‘compactness’ score (1.20 vs 0.94), more involved calculations/simulations are likely required to determine representative properties from the chemical structures alone. While CPs can form secondary structures upon binding, folding in solution prior to target binding may occur,[42] in particular for CP libraries composed of longer peptides given the greater propensity for adopting secondary structure. Determining the structure of these CPs in solution would aid in this analysis.

5 | STRUCTURES OF CYCLIC PEPTIDES BOUND TO PROTEIN TARGETS

In the following section, we will focus on the insights gained from the CP:protein co-crystal structures published between 2015 and 2019. Throughout this review, protein residues are referred to with 3-letter codes while CP residues are referred to with 1-letter codes.

5.1 | Cyclic peptides targeting enzymes

Enzymes represent the most common target class for CPs, and indeed over half of the targets reported in the last 5 years are enzymes (Table 2). Many CPs developed are inhibitors, typically targeting the deep active site pocket/cleft, which allows for extensive interactions...
to be made with the protein of interest. With the larger size and
greater number of interactions that CPs make with their target of
interest, CPs typically can achieve higher potency and selectivity
within the target enzyme family, relative to small molecules. CPs
can interact with active-site residues and thus inactivate the cata-
lytic activity of enzymes (e.g., HPA\textsuperscript{[41]}), or act as substrate competi-
tive inhibitors, directly displacing the target substrate (e.g.,
KDM4A\textsuperscript{[37]}; in some cases they bind to allosteric sites to induce
conformational change to inactivate the enzyme function (e.g.,
iPGM\textsuperscript{[30]}). Another type of ligand recently developed is the
\textit{‘silent’} allosteric CPs, where tight binding CPs that do not affect enzyme
activity were used as capture probes for an enzyme-substrate com-
plex (HIF1:PHD2 complex).\textsuperscript{[55]} The recent availability of co-crystal
structures of CPs bound to different enzymes have illuminated the
CP:protein interactions and the mechanism of inhibition at the
molecular level, aided further design, and in some cases, informed
new biological insight of their targets. In this section, we will focus
on structural knowledge gained from recent examples of active-site
targeting CPs and allosteric site targeting CP inhibitors for different
enzyme families.

Different CP scaffolds target the catalytic site of
Human pancreatic \(\alpha\)-amylase

Human pancreatic \(\alpha\)-amylase (HPA) degrades starch into malto-oligosac-
charides within the gut, which are subsequently broken down into glu-
cose. HPA is thus an attractive therapeutic target for the treatment of
diabetes and obesity.\textsuperscript{[72]} Miglitol and acarbose, inhibitors of HPA and
other \(\alpha\)-glucosidase, are currently in clinical use for diabetes; however
their use is limited due to side effects, likely arising from off-target inhibi-
tion of non-digestive \(\alpha\)-glucosidase. Jongkees \textit{et al}. carried out RaPID
mRNA-display selection with \(YZ\) \textit{initiated library to identify selective
CPs for HPA (Table 2).\textsuperscript{[41]} Inspired by the selective inhibitor montbretin
A\textsuperscript{[74,75]} which contains multiple phenolic groups, phenol-containing
nCAAs (DOPA and resorcinol) were incorporated into the library in an
attempt to bias the discovery of active site binding peptides. The \(Z\)
library led to enrichment of consensus sequence \(YPYS\textsubscript{2}\\text{Wx}RH\) contain-
ting two tyrosines in a small five-residue macrocycle, with a four-
residue tail. piHA-Dm, a truncated peptide containing the consensus
sequence (Ac\textsubscript{N}YPYS\textsubscript{2}\\text{WVHR-NH\textsubscript{2}}), was shown to be a substrate com-
petitive inhibitor (\(K_i = 7\) nM), with only modest improvements in potency

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Properties of cyclic peptides co-crystallised with their target. Peptide interacting areas were extracted using PyMOL from their respective PDB files. Co-crystallised peptides generated using either mRNA display (blue circles) or phage display (red diamonds) show remarkable consistency in, A, the total peptide interacting area and B, the percentage of the total peptide area that interacts with the target protein, while, C, the molecular weights of CPs identified through phage-display are generally lower than peptides from mRNA-display. Comparison of calculated size properties for these CPs and the measured SASA from the crystal structures, D-G, showed good correlation for phage-display CPs but very poor for mRNA-display CPs.}
\end{figure}
obtained when longer tail sequences were introduced. The crystal structure of piHA-Dm in complex with HPA (PDB: 5KEZ) revealed that the peptide has a very compact structure with most of the sidechains forming interactions with HPA (Figure 3A,B). The five amino acid macrocycle makes all of the interactions with the catalytic residues with a network of ordered water molecules mediating interactions between HPA and piHA-Dm. The tail forms a 3_10-helix with the helical axis perpendicular to the plane of the macrocycle, and a central water molecule forming three hydrogen bonds stabilizing the structure. Interestingly, catechol or resorcinol moiety containing sequences were not identified from the DY library selection,[41] but replacement of Y3 in piHA-Dm with LDOPA, increased potency 10-fold. Thus, while LDOPA can enhance affinity of CPs to the active site of HPA, competing factors in the mRNA display selection, such as reduced translation efficiency of unnatural amino acids compared to the natural amino acids, may have contributed to the initial lack of enrichment of 1DOPA containing CPs.[43]

The 1Y initiated library showed little conservation amongst the enriched sequences,[41] but identified piHA-L5(d10Y) (Ac-YGHSHIRFGYSYSHVSYCG-NH$_2$), a distinct sequence from piHA-Dm,[42] with $K_i = 14$ nM for HPA (note that position 10 was 1DOPA, abbreviated to ‘d’ in the original manuscript. The d10Y mutant was only 5-fold less potent and was used in the majority of subsequent experiments. In the crystal structure (PDB: 5VA9) the majority of the peptide (F8 to G18) forms an α-helix and the remaining residues adopt an extended loop conformation at one end of the helix (Figure 3C,D).

In this structure, the catalytic residues Asp197 and Glu233 both form hydrogen bonds with the guanidinium group of R7, while Asp300 forms hydrogen bonds with the backbone nitrogens on F8 and G9 via a bridging water molecule.[42] piHA-Dm is half the size of piHA-L5(d10Y) (9 vs 18 residues) but is a more potent inhibitor. Both CPs contain helices which partially occupy the same space adjacent to the catalytic site of HPA, but the orientation of the helices differs between the two (Figure 3E). At the active site, piHA-Dm forms a single direct hydrogen bond between Y3 and Asp197, and all the other interactions with the catalytic residues are mediated by a network of water molecules, whereas piHA-L5(d10Y) interacts directly through R7. Hence, CPs targeting the same catalytic site/residues can adopt very different structures and interactions to achieve catalytic inhibition; the relatively open/accessible active site of HPA may better accommodate peptides with diverse conformations.

**Substrate competitive CP inhibitors of histone lysine demethylase (KDM4A)**

The histone lysine demethylases (KDMs) remove methyl groups from the sidechains of lysine residues in the N-terminal ‘tails’ of histone...
These methyl groups are part of a complex set of post-translational modifications (PTMs) that are found on histone proteins to control eukaryotic gene expression. Mis-regulation of PTMs is often the hallmark of diseases such as cancer. In particular, the KDM4 subfamily which demethylates tri-/di-methyllysines (Kme3/Kme2) on histone H3 at K9 and K36, have been identified as potential therapeutic targets for multiple cancers. KDM4s belong to a large family of Fe(II) and 2-oxoglutarate (2-OG) dependent oxygenases (>60 enzymes) with highly conserved active site architecture, making the development of potent and selective small molecule inhibitors challenging. To identify novel and selective scaffolds for KDM4 inhibition, a Rapid mRNA-display selection was carried out against KDM4A. Several hits containing an ‘RSG’ motif were identified, including CP2 (Ac-DYVYNTRSGWRWYT-C-NH2) which was found to have high potency (IC50 < 50 nM) for KDM4A/B/C and selectivity (>150-fold) over other KDM subfamilies and 2OG oxygenases. An X-ray crystal structure of KDM4A in complex with CP2 (PDB: 5LY1) revealed that CP2 inserts into the substrate binding pocket by adopting a distorted anti-parallel β-sheet conformation (Figure 4A,B). R6 forms part of a type-I β-turn near the active site metal to occupy the same sub-pocket as the trimethylated lysine residue in histone substrates, and the guanidinium group forms hydrogen bonds with Tyr177, Ser288, Asn290. SAR analysis revealed that positive charge on R6 is crucial for its potency, and replacing R6 with Kme3 (CP2R6[Kme3]) or Rme2 converted them to substrates, despite sharing no sequence homology with histone substrates, and confirming novel arginine demethylase activity. The co-crystal structure of KDM4A and CP2R6[Kme3] (PDB: 5LY2) confirmed its productive binding mode, orienting the Kme3 residue in the same manner to the H3K9me3 and H3K36me3 substrates (Figure 4B). Structure guided design to improve the cellular stability and permeability of CP2 (Figure 4C) led to CP2.3, which features five amino acid modifications from CP2 including N-methylations: Ac-DYynVYNTRS-GAWRWAF-PFMeC-NH2 (Figure 4D). CP2.3 maintained potency in vitro (IC50 = 110 nM) and 10-fold improvement in cellular target engagement (EC50 100 nM for CP2.3), with correlating increase in global H3K9me3 levels. To identify CPs with improved cell permeability, a focused library selection was run with the ‘RSG’ motif centrally fixed and the remaining variable region randomised using NKN codons to bias the selection towards positively charged and hydrophobic amino acids. The selection yielded CPs such as CP2f-3 (YIRRGWYLW) and CP2f-7 (YTRFRSGVFFY) with improved in vitro potency (IC50 = 6 nM) (Figure 4D). CP2 was further investigated by deep mutational scanning (DMS), an approach where an mRNA-display library of CP2 derivatives with each residue varied was used for a single round of enrichment and enrichment factors for each peptide sequence position were calculated.
calculated. At each position 40 different amino acids were used: 19 proteinogenic amino acids (excluding methionine) and 21 non-proteinogenic amino acids including N-methylated, aliphatic, aromatic and D-amino acids. A CP2 derivative, rCP2, was synthesised incorporating four identified mutations to reduce the polarity or steric bulk without reducing affinity: N4meA, T5A, W9Bzt and R10Nva (Figure 4D). rCP2 retained the same affinity for KDM4A as CP2 ($K_D$ [SPR] = 7 nM). The results for G8 were further investigated as substitution to any other residue except D4A was detrimental to binding. Inspection of the $\psi$ and $\phi$ angles for G8 in the co-crystal structure revealed that G8 is in a region disallowed for L- but permitted for D-amino acids so it is likely that substitution to an L-amino acid disrupts the $\beta$-turn and may prevent anti-parallel $\beta$-sheet formation. This effect may be a wider phenomenon; since glycine does not have a side chain the backbone geometries it can adopt may be the main selection pressure for its presence in hit sequences. Overall, the results from DMS supported the observations made from the co-crystal structure, providing valuable insight into the importance/mutability of each residue going beyond conventional Ala-scan, including with non-proteinogenic amino acids that may aid with other properties such as membrane permeability or metabolic stability; this is particularly important when structural information is not available.

**CP targeting the substrate binding pocket of Urokinase-type plasminogen activator**

Urokinase-type plasminogen activator (uPA) is a trypsin-like serine protease, and the main enzyme responsible for plasminogen activation in the extracellular space. Serine proteases have highly conserved active sites and as such present challenging targets for the development of selective small molecule inhibitors. A disulfide-linked macrocyclic peptide mupain-1 (CPAYSRYLDLC) was initially identified from a phage display selection ($K_D$ = 400 nM). Recent optimisation work, using a selection based on back-flip library of peptide-protease fusions based on mupain-1, yielded IG1 and IG2 (differing only in the N- and C-termini which were amide/amide and amine/amid acid respectively). IG2 showed particularly high specificity and 100-fold higher affinity for uPA than mupain-1. The co-crystal structure of IG2 (CPAYSRYIGC) with uPA (PDB: 6A8N) shows that the CP does not contain any specific secondary structural elements when bound to the protein, instead adopting an extended conformation covering a large patch of the protein surface making multiple interactions away from the active site binding pocket, furnishing the peptide with excellent selectivity (Figure 5A). Interestingly, IG2 does not interact with any of the residues in the catalytic triad, Asp102-His57-Ser195, yet does form extensive contacts within the S1 substrate recognition pocket - a highly conserved region within serine proteases. R6, the only positively charged residue in IG2, is positioned at the P1 residue site of uPA, yet is not hydrolysed presumably as the amide bond is held in an orientation which does not allow nucleophilic attack by Ser195. This is a common feature with many other CP inhibitors of uPA (Figure 5B,C), which all have an arginine residue in the S1 pocket and occupy varying portions of the surrounding binding surface.

**Allosteric CP inhibitors for isomerase: conformational lock for cofactor independent phosphoglycerate mutase**

Cofactor-independent phosphoglycerate mutase (iPGM) is the sole enzyme responsible for the interconversion of the key metabolic intermediates 2- and 3-phosphoglycerate in nematode worms. iPGM shares no sequence similarity with the equivalent human enzyme, cofactor-dependent PGM. Silencing of ipgm gene in Caenorhabditis elegans and Brugia malayi leads to nematode death, making iPGM an attractive target for anthelmintics. A RaPID mRNA display was carried out using iPGM from C. elegans, yielding peptides with small macrocycles and C-terminal ‘tails’ with very high affinity, in particular Ce-2 (Ac-C$^3$YDYPGDCYLYGTCG-NH$_2$), with $K_D$ of 73 pM for its target. The free cysteine thiol was shown to be important for activity (C14S mutant had 100-fold lower inhibition) and linearised peptides...
showed little/no inhibition. Peptides corresponding to both the macrocycle only (residues 1-8) and tail only (9-14) in isolation demonstrated no inhibition, while an intermediate truncation (residues 1-11) containing part of the tail, Ce-2d (Ac-DYDPGDYCYLGNH2), retained sub nM inhibition. The crystal structure of Ce-2d bound to C. elegans iPGM (PDB: 5KGN) supports the structure activity relationships determined through biochemical assays. Residues 6 to 11 of Ce-2d form an α-helix stabilised by the macrocycle formed by residues 1 to 8, which binds in the hinge region of iPGM (Figure 6), a cleft formed between the transferase and phosphatase domains. The tight binding of Ce-2d can be explained by the extensive inter-molecular interactions that it forms with both domains. While Ce-2d binds near the active site phosphoglycerate binding pocket of the transferase domain (residues highlighted in pink, Figure 6A), no interactions with these residues (either direct or water-mediated) are observed. Ce-2d acts as a non-competitive allosteric inhibitor, inducing iPGM to adopt an open, inactive conformation upon binding. Unusually, most of the intermolecular hydrogen bonds are formed by the backbone of Ce-2d; only D2 and Y3 have side chains that interact with iPGM through several water-mediated hydrogen bonds, and the rest of the amino acids form intramolecular hydrogen bonds, which fold Ce-2d into a precise 3D orientation to allow the backbone atoms to interact with the protein. Residue D6 likely acts to stabilise the macrocycle and helix as it interacts with D2, Y3, C8, and Y9 (Figure 6B); a comprehensive analysis of intermolecular interactions is detailed in Malde et al. The authors note that the C-terminus of Y11 points towards protein bound Zn²⁺ and Mn²⁺ ions which would be within range for chelation by the thiol of C14 in the longer peptide(s), such as Ce-2.³⁰

5.2 | Targeting protein-protein interactions

CPs have chemical diversity manifested via combinations of main and side-chains, that offer different binding interactions (hydrophobic, van der Waals, polar and hydrogen bonding) with their targets. With their increased size, CPs are well suited to targeting protein surfaces and PPI interfaces which are typically relatively large, shallow and lack 'pockets' – and challenging to target using traditional small-molecules.

FIGURE 6  X-ray crystal structure of Ce-2d in complex with iPGM. A, A cartoon representation of Ce-2d (orange) in the hinge cleft of iPGM (teal with active site residues [interacting with the glycerol [pale green]: His147, Arg177, Asp178, Arg210, Arg216 and Arg287] highlighted in pink. B, Aspartate at position 6 (purple) in Ce-2d forms extensive hydrogen bonds to other residues, stabilizing the loop at the base of the helix. PDB: 5KGN

FIGURE 7  Co-crystal structure of 1:2 binding mode of M21 to TNFα. A, Structural representation of bicyclic M21 showing sidechains and central bridging aromatic core. B, M21 (orange) interacts with two TNFα monomers (teal and pale green) in the co-crystal structure. C, Cartoon representation of M21 showing side-chains; the core and loop1 form a large (419 Å²) area, almost entirely comprised of hydrophobic interactions with the blue monomer while loop two interacts with the green monomer through a smaller contact area 238 Å²
In addition to enzymes as described above, CPs have been successfully used to target a wide range of proteins in the last 5 years, including receptors (HGFr,[35] IDOL,[45] and hEGFR[40]), transcription factors (BCL6)[29] and small signalling proteins (hTNF-α[44] and HGF[34]) (Table 1). Many of these proteins are involved in complex biological network/signalling pathways, and CP ligands have efficiently been generated to bind and modulate the protein-protein interactions of these targets. We highlight recent examples of CPs that disrupt PPIs, where co-crystal structures with the target have informed their site of binding, new mechanism of PPI disruption leading to inhibition of downstream signalling pathways.

**TNFα: disassembly and inhibition**

Tumour necrosis factor-alpha (TNFα) is a proinflammatory cytokine, critical for mediation of the normal inflammatory response but overproduction leads to tissue damage associated with various diseases, including rheumatoid arthritis, psoriasis and ankylosing spondylitis.[85] A bicyclic peptide phage display screen was carried out to identify binders of TNFα with an initial library of $\text{CX}_n\text{CX}_m\text{C}$ ($n,m = 2-6$) cyclised through the cysteines by an aromatic core to generate bicyclic peptides.[44] Initial enrichment identified two similar peptides (both with a small CPPC motif in the first loop) and a subsequent affinity maturation procedure starting from the consensus sequence, returned M21 $\text{ACPPC}_6\text{CLWQLC}$ cyclised with a 2,4,6-Tris(bromomethyl)mesitylene (TBMB-methyl) core (Figure 7A). This was shown to bind to TNFα with $K_D(\text{FP})$ around 30 nM for a fluorescently labelled derivative. In cellular assays, a strong time-dependent effect was observed with much more potent inhibition occurring after prolonged incubation of M21 with TNFα, which was shown to be due to disassembly of the TNFα trimers into a mixture of dimers and monomers, with M21 binding to the dimers. The mode of inhibition was analysed using a combination of biophysical techniques (mass spectrometry, analytical ultracentrifugation and multi-angle light-scattering) and confirmed by co-crystal structure of M21-TNFα complex (PDB: 4TWT). M21 interacts with both TNFα monomers (Figure 7B), one through a largely hydrophobic interaction involving the TBMB-methyl core and first loop (A1- C5) with interaction area 419 Å², and the other through an $\alpha$-helix (W7-G12) in the second, larger loop, which has a smaller interaction area of 238 Å² (Figure 7C). In this selection, a range of cyclising 'cores' were used and M21 was shown to be inactive with alternative cores, demonstrating that while often considered to be almost irrelevant (on the grounds that it will be consistent between all the peptides in the pool), the cyclisation method can have profound effects on the CPs discovered and even be involved in binding to the target (a unique feature for this CP in comparison to the others covered here).

**K Ras signalling pathway: CP that selectively disrupts K-Ras(G12D) PPI**

V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (K-Ras) is a eukaryotic signalling protein involved in the pathways for cell growth, proliferation and differentiation.[86] K-Ras binds GTP in its active state and has weak GTPase activity, which can be increased by binding GTPase-activating proteins (GAPs), resulting in hydrolysis of the terminal phosphate and conversion of GTP to GDP. The GDP-bound...
form of K-Ras binds to guanine nucleotide exchange (GNE) factors such as son of sevenless 1 (SOS1), which eject the GDP molecule allowing K-Ras to become reactivated by binding to GTP in the cytosol. Mutations at Gly12 in K-Ras prevent efficient binding of GAPs thereby fixing K-Ras in the GTP-bound, active, form leading to aberrant cell growth. K-Ras Gly12 mutations are observed in many different cancers[87] and are well validated targets, with significant drug discovery effort invested over the last >30 years (see[88] for a recent review on RAS targeting peptides).

A phage display approach was used to target immobilised GDP-bound form of K-Ras(G12D) in the presence of free wild-type K-Ras, to identify peptides capable of specific binding to the mutant form. Three enriched groups of similar peptides were identified following DNA sequencing. The consensus sequence for each group was synthesised chemically and used in SPR binding experiments. Following optimisation of an initial hit, KRpep-2d (Ac-RRRRCPLYISYPVCRRRRR-NH2) with two additional arginine residues at each terminus, was identified, which showed better potency and similar selectivity (IC50 values of 1.6 nM, 18 nM, and 42 nM against K-Ras(G12D), K-Ras(G12C), and wild type K-Ras, respectively). Cellular assays with KRpep-2d revealed dose-dependent reduction in extracellular signal regulated kinase (ERK) 1/2 phosphorylation (a downstream signal of K-Ras) and suppression of cell proliferation only in cells expressing the G12D mutant. An X-ray co-crystal structure (PDB: 5XCO) revealed that the central core (C5-C14) of KRpep-2d adopts an α-helix stabilised by the disulfide with a large loop, and occupies a cleft between α2 and α3 at an allosteric site in KRas(G12D) (Figure 8A-C). The four N-terminal (R1-4) and two C-terminal (R18 and R19) arginine residues are largely solvent exposed and do not form specific interactions with the protein or have any secondary structure (although there is an extensive network of water-mediated hydrogen bonds between the arginines). Residues R16 and R17, however, form part of the α-helical region DPVCRR (D12-R17) and may add stability by extending it for an extra half turn. Comparison with existing structures of other K-Ras complexes identified a hydrogen bond between Asp12 and Gin61 in KRas(G12D) mutant that may stabilise the required Switch II conformation to enable KR-pep2d to bind (Figure 8C). L7 and I9 binding sites are formed predominantly by residues in Switch II and are likely key regions, as an earlier alanine scan highlighted L7, I9, and D12 as key residues in KR-pep2d for inhibition[89]. D12 sits at the base of the α-helix and the carboxylate sidechain forms hydrogen bonds to the sidechain of Gin99 and Arg102. KRpep-2 can allosterically block KRas(G12D) interaction with GNE, a key interaction that activates K-Ras; thus this novel allosteric site that selectively disrupts key PPI for downstream signalling could be targeted by future drug discovery efforts to gain selectivity for the G12D over the wild-type K-Ras (and other mutants, for example, G12C).

Inhibition of semaphorin signalling: Plexin B1

Semaphorins are a class of 20 proteins that act as signalling molecules in a range of processes typically constituting a short-range inhibitory signal. Plexins are a family of 9 transmembrane proteins that function as receptors for semaphorins with extracellular sema, plexin-semaphorin-integrin (PSI) and immunoglobulin-like-plexin-transcription factor (IPT) domains and a GAP domain.[89] The Semaphorin 4D-plexin B1 (Sema4D-PlxnB1) interaction regulates osteoblast differentiation so disruption of this PPI would constitute a treatment for osteoporosis; however the large relatively featureless interface makes targeting with traditional small molecule approaches challenging.[90]

An mRNA-display selection was carried out targeting part of the N-terminal extracellular region of PlxnB1 containing the Sema 4D-binding region (hPlxnB1SP residues 20-535).[54] After 6 rounds of selection, one sequence (and derivatives thereof) was highly enriched, making up over 80% of the identified peptides. Binding affinities for representative peptides were determined by SPR, with the top hit PB1m6 (Ac-DWRPRVARWTGQIYCS-NH2) having the highest affinity (Kd = 3.5 nM). Subsequent assays indicated that PB1m6 modulated the binding of Sema4D and could selectively prevent Sema4D-induced cell collapse. Intriguingly, the X-ray co-crystal structure of PB1m6:hPlxnB1SP (PDB: 5B4W) revealed that the PB1m6 binding site does not overlap with the Sema4D binding site (Figure 9A), and no major conformational changes to hPlxnB1SP were observed when compared to other structures. The authors suggest that PB1m6 may affect the dynamics of the loops involved in Sema4D binding and that this change would not be observable from a crystal structure. PB1m6 has an antiparallel β-sheet-like structure (Figure 9B) and binds in a region between blades 5 and 6 of the 7-bladed β-propeller sema domain, forming interactions around the fourth strand of blade 5 and the loops between β-strands 6A/6B (412-417) and 6C/6D (433-439) (Figure 9C/D).[91] There are also interactions with blade 4 of another molecule of hPlxnB1 within the crystal structure, although this is likely induced by crystal packing rather than a dimeric interaction. Thus, it is important to consider how this affects the peptide conformation in crystallo as it may not be representative of the structure in solution.

The maximal level of disruption of the hPlxnB1-Sema4D interaction achievable was 65% with PB1m6, which was suggested to be due to its lower affinity for full-length hPlxnB1 over hPlxnB1SP. Dimeric versions of PB1m6 were synthesised with two copies of the CP attached via linker regions of various length to gain avidity through multivalency. All the linkers tested led to molecules with at least 80-fold tighter binding than PB1m6 as measured by SPR (Kd = 30 pm), and near total inhibition of Sema4D-mediated cell collapse.[53] This exemplifies a conceptually simple but effective bivalent approach, analogous to antibodies, to enhance the binding of CPs against their targets of interest.

6 | DISCUSSION AND FUTURE PERSPECTIVE

Since the first CP phage display was described over 25 years ago,[52] the CP genetically encoded library technologies have transformed the way ligands are generated for protein targets of interest. The
Wide-ranging properties and functionalities of the amino acid building blocks (including nCAAs) and their combinations, together with their ability to form complex 3D structures mean CP libraries have diversity and chemical space unmatched by many other screening libraries. CPs have successfully been used to target a wide range of different classes of proteins, including 'difficult' PPIs. While beyond the scope of this review, non-protein biomolecules and more complex systems can also be targeted by CPs (e.g., lipids, cells), demonstrating the power and versatility of CP genetically encoded library technologies.

FIGURE 9  Structure of PB1m6 bound to hPlxnB1. A, Two molecules of CP PB1m6 (labelled I (orange) and II (purple)) bind at the dimerisation interface of two hPlxnB1 proteins (labelled I (teal) and II (light green)). The Sema4D binding interface is highlighted on both protomers in pink. B, Cartoon representation of PB1m6 I on the surfaces of hPlxnB1 I and II. C, Peptide only in an alternative orientation. D, Stick representation of hPlxnB1 I with the fifth and sixth blades of the 7-bladed β propeller highlighted in purple and green (hPlxnB1 II is omitted for clarity). Protein residues making hydrogen bonds are shown as sticks with hydrogen bonds as black dashed lines. PDB:5B4W

The structural information that has become available for de novo CPs over the past decade provides significant insight into their unique properties. Surveying the co-crystal structures of de novo CPs published in the last 5 years alone, it is apparent how adaptable CPs can bind to a variety of protein topographies, from shallow surface grooves to deep enzyme active site pockets. Secondary structure motifs, such as α-helices, β-sheets, 3_10-helices, and β-turns, allow the CPs to adopt compact, globular structures binding through precisely oriented backbone hydrogen bonds or more extended backbone conformations, scaffolding sidechains to make key non-polar and...
electrostatic interactions. We have surveyed calculated physicochemical properties and affinities between CPs identified through different selection methods in the last 5 years (section 3). While some trends were observed within this limited dataset, deconvolution from other factors that may be responsible for the differences is challenging. As discussed in section 3, both phage- and mRNA-display rely on binding, enriching for higher affinity binders by through retention of species with slow $k_{\text{off}}$ rates ($N.B. K_D = k_{\text{on}}/k_{\text{off}}$), whereas the SICLOPPS technique, for example, employs a functional read-out that does not explicitly enrich for high-affinity target binding. One confounding factor is the library size (i.e., length of peptides), which is typically not the same across different methods and we are not aware of any examples using the same library design with different selection techniques. Cyclisation methods are also different; typically phage selections employ sidechain-to-sidechain cyclisation (commonly Cys-Cys through a disulfide or via a linker/core), mRNA head-to-sidechain and SICLOPPS head-to-tail. Another factor is target type, which is entirely independent of the selection method; conceptually, at least, any technique could be applied to any target. In addition, the calculated physicochemical properties are based on algorithms designed for small molecules and likely have larger errors for larger peptides, where the intra-molecular H-bonding or secondary structures arise; thus, experimental validation is needed before conclusions are drawn. Rational design of these peptides is beyond what is achievable from the knowledge we have currently, yet using selection procedures from massively diverse pools, peptides with high affinity and selectivity can be generated in a relatively short space of time making these selection techniques very powerful tools.

The CP co-crystal structures are also invaluable to drug discovery efforts. They can reveal new ways of targeting proteins, such as different modes of active site or allosteric site inhibition (e.g., for well-established targets such as K-Ras$^{[64]}$) and identify new protein hot-spots, or even inform on substrates that may lead to new biological insight. Combined with activity-based SAR, co-crystal structures can also aid in medicinal chemistry design to further refine CPs for biological/therapeutic applications, or to inspire peptidomimetic/small molecule design. One of the major hurdles for CPs is reliable cell permeability, a challenge faced when targeting intracellular proteins; as demonstrated for KDM4A,$^{[37]}$ crystal structure-guided modifications can support the design and engineering of cell-permeable and stability of CPs. While we have not covered in this review, CPs can also act as co-crystallization chaperones$^{[68,95]}$ and stabilise protein conformation, thus providing new structural understanding of the target protein. However, in some cases, in particular when CPs are found at the interface of proteins, it is not trivial to distinguish between biologically-relevant interfaces and non-specific interfaces due to crystallographic packing. It is therefore important to consider crystallographic data in combination with protein:CP interaction data in solution (e.g., NMR, biochemical assays).

The reporting period has seen a substantial increase in publications on de novo CPs, several significant advances in CP technologies,$^{[12,18]}$ four companies founded based on CP drug discovery, and several CPs entering clinical trials.$^{[12,18]}$ CPs are exciting and attractive modality with enormous potential, not just in therapeutic applications, but also in other applications such as imaging, diagnostics as well as in chemical biology and basic science, and we anticipate significant growth and impact in these areas.

ACKNOWLEDGEMENT

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 679479 EPITOOLS). G. R. is supported by the Oxford-GSK-Crick Chemical Biology Centre for Doctoral Training Programme via the Systems Approaches to Biomedical Science (EP/R512333/1), EPSRC and GSK. We apologise for the incomplete citations and for the research that we were not able to cite due to space constraint.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

Data sets generated from the peptides listed in Tables 1 and 2 will be made publicly available through Newcastle’s ‘data.ncl’ open data repository and the Oxford University Research Archive.

ORCID

Tom E. McAllister https://orcid.org/0000-0001-6906-4510
Oliver D. Coleman https://orcid.org/0000-0002-6205-9636
Grace Roper https://orcid.org/0000-0002-7198-0267
Akane Kawamura https://orcid.org/0000-0003-1169-5082

REFERENCES

[1] C. Morrison, Nat. Rev. Drug Discov. 2018, 17, 531.
[2] A. A. Vinogradov, Y. Yin, H. Suga, J. Am. Chem. Soc. 2019, 141, 4167.
[3] T. Passiourea, Biochemistry 2020, 59, 139.
[4] A. F. B. Räder, M. Weinmüller, F. Reichart, A. Schumacher-Klinger, S. Merzbach, C. Gilon, A. Hoffmann, H. Kessler, Angew. Chem. Int. Ed. 2018, 57, 14414.
[5] S. A. K. Jongkees, C. J. Hipolito, J. M. Rogers, H. Suga, New J. Chem. 2015, 39, 3197.
[6] A. K. Malde, T. A. Hill, A. Iyer, D. P. Fairlie, Chem. Rev. 2019, 119, 9861.
[7] T. Sander, J. Freyss, M. Von Korff, C. Rufener, J. Chem. Inf. Model. 2015, 55, 460.
[8] Wassenaar, T. A. [PyMOL] Calculating % secondary structure https://sourceforge.net/p/pymol/mailman/message/25689467/ (accessed September 4, 2020).
[9] G. P. Smith, Science 1985, 228, 1315.
[10] R. W. Roberts, J. W. Szostak, Proc. Natl. Acad. Sci. U. S. A. 1997, 94, 12297.
[11] C. P. Scott, E. Abel-Santos, M. Wall, D. C. Wahnon, S. J. Benkovic, Proc. Natl. Acad. Sci. U. S. A. 1999, 96, 13638.
[12] C. Sohrabi, A. Foster, A. Tavassoli, Nat. Rev. Chem. 2020, 4, 90.
[13] K. T. O’Neill, R. H. Hoess, S. A. Jackson, N. S. Ramachandran, S. A. Mousa, W. F. DeGrado, Protein Struct. Funct. Bioinforma 1992, 14, 509.
[14] A. Frankel, S. W. Millward, R. W. Roberts, Chem. Biol. 2003, 10, 1043.
[15] L. Wang, A. Brock, B. Herberich, P. G. Schultz, Science 2001, 292, 498.

[16] Y. Goto, T. Katoh, H. Suga, Nat. Protoc. 2011, 6, 779.
AUTHOR BIOGRAPHIES

Tom E. McAllister completed his PhD at the University of Leeds (UK) with Mike Webb in 2012 investigating phosphohistidine. He subsequently held a postdoctoral position with Bruce Turnbull (also in Leeds) working on virus-like particles. Tom joined the Kawamura group in July 2014 and since then his research has centred around the use and development of the mRNA-display platform for identifying cyclic peptide binders of protein targets of interest, predominantly focusing on epigenetic enzymes.

Akane Kawamura is a professor of chemical biology at Newcastle University (UK) and holds a joint position with the University of Oxford (UK).
She obtained her MChem and DPhil from Oxford, then spent time in the biotech sector before returning to academia in 2009. She worked with Prof. Chris Schofield, FRS before establishing her own group at Oxford in 2012, and joined Newcastle University in 2019. Research in her group focuses on the use of chemical biology approaches to study biomedically important processes such as epigenetics.