NJC

Accepted Manuscript

This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/njc

NJC

RSCPublishing

Perspective

Model foldamers: applications and structures of stable macrocyclic peptides identified using *in vitro* selection.

Cite this: DOI: 10.1039/c3nj00000x

Seino A. K. Jongkees,[a] Christopher J. Hipolito,[a] Joseph M. Rogers[a] and Hiroaki Suga[a]

Received 00th XXXXX 2013,
Accepted 00th XXXXX 2013

DOI: 10.1039/c3nj00000x

www.rsc.org/njc

[a] Department of Chemistry, Graduate School of Science, University of Tokyo. 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033 Japan.

Foldamers are synthetic molecules that seek to mimic the structure-forming propensity of biomolecules, such as proteins. However, on a short oligomer scale, peptides often do not fold in the same manner as large proteins, despite being composed of the same amino acid building blocks. Constraints to available peptide conformations can improve these folding characteristics. One important constraint that leads to an increase in folding behaviour is the formation of a macrocycle, while doing this by means other than disulfide bond formation ensures that this structural constraint persists in all biological settings. Additional non-natural features, such as incorporation of amino acids with unusual side chains, D-amino acids, *N*-alkyl amino acids, and β-hydroxy acids, further mimic the synthetic characteristics of foldamers, giving a class of compound that is intermediate between natural proteins and synthetic foldamers. *In vitro* selection methods, such as phage and mRNA display, allow access to *de novo* peptides based solely on their ability to bind a target, potentially giving access to unique structures and functions. Recently, a series of structures have become available for several such partially synthetic macrocyclic peptides derived from *in vitro* selection. Here we present an overview of the structural features of these stable macrocyclic peptides and their binding to protein targets, as well as some initial indications of their folding behaviour free in solution, and discuss implications for future design and functions of foldamers.

Introduction

In vitro selection is a powerful technique for isolating molecules with novel functions from large libraries. In particular, libraries of ribosomally translated peptides contain a variety of functional groups in highly diverse arrangements within a relatively compact scaffold, and using a variety of techniques, can be readily linked to the corresponding sequence information at the nucleic acid level. The functional and spatial diversity of these peptides can be further increased if non-proteinogenic amino acids are included.[1] Moreover, by incorporation of backbone-modified building blocks, for example *N*-alkylated amino acids, α-hydroxy acids, and β-amino acids, within a predominantly α-amino acid background it should also be possible to access hybrid compounds with functionality and structure not available to natural peptides.

Macrocyclization of peptides made from canonical amino acids often confers benefits over the linear form, by conformationally restricting the peptide backbone. Common benefits include higher affinity binding and/or target selectivity,[2] tunable membrane permeability,[3] and increased resistance to both spontaneous[4] and protease-catalysed[5] hydrolysis. Macrocyclization would likely also confer similar benefits on peptide hybrids and peptidomimetics. However, in order for these benefits to persist in a reducing biological setting the macrocyclization needs to be chemically stable, and so not achieved by easily reversible disulfide bond formation. While many examples exist of disulfide-cyclised peptides, far fewer examples of stable peptide-based macrocycles were known. Recently, structural information has become available for several stable macrocyclic peptides derived from *in vitro* selection, and the structural features exhibited therein will be the topic of this focused review.

Synthetic molecules that aim to mimic biomolecule folding, foldamers, aspire to also mimic their function, such as the binding of protein targets.[6,7] Conformationally constrained peptides, particularly those containing non-proteinogenic amino acids, bridge the gap between natural biomolecules and entirely synthetic foldamers. The structural information their binding provides will give valuable information for future foldamer application and design.

Selection methods

Two selection methods have been used for *in vitro* discovery of tight binding non-reducible macrocyclic peptides,[8] with the difference being in the means of linking sequence information to the displayed peptide. These methods are phage and mRNA display. In the case of phage display the link comes from assembly of a phage particle, while in the case of mRNA display it is a covalent link between the peptide and its encoding mRNA.

Phage display. Phage display, as the older of the two methods, has been reviewed many times,[9,10] and has been used successfully in a variety of directed evolution and selection based applications. In this method, the coding sequence for the peptide or protein of interest is inserted into that of one of the coat proteins that forms the surface of the phage particle (Figure 1). The peptide of interest is thus displayed as a fusion with the viral coat protein. The most commonly used phage is M13, while the most commonly used surface protein is pIII to ensure tight control over copy number and orientation of the fusion protein. In the case of *in vitro* selection, a library with section(s) of random sequence is used for the peptide of interest, resulting in each phage particle potentially displaying a unique peptide on its surface and containing the relevant coding sequence inside. Selection by phage display involves binding of phage, which display the protein or peptide to be selected, to an immobilised target. This is followed by washing away of non-binders, and finally elution and replication of the target-binding phage in a bacterial host. The aforementioned process is referred to as a round and multiple rounds are typically performed for a single selection. Negative selection, which removes phage binding to the support used for target-immobilisation, is also typically carried out. Because phage display involves a replication stage inside a living host it is somewhat limited in its flexibility. Displayed proteins that interfere with the replication process or that are detrimental to the cell may be selected against, despite having good target affinity. For the same reasons, any peptides that enhance phage replication may become enriched despite poorer target binding. Reprogramming the genetic code in phage-displayed peptides (*vide infra*) is also more difficult, as protein expression takes place inside bacterial cells, but is possible.[11]

mRNA display. A more recently developed method, which avoids the use of replication in a living host and thus is completely *in vitro*, is mRNA display.[1,12] In this method, peptide is covalently linked to its mRNA via attachment of the antibiotic puromycin (Figure 2). Puromycin is an aminoacyl-tRNA mimetic with a functionally non-hydrolysable amide bond between the amino acid and nucleoside portions. Because an *in vitro* translation system lacking a release factor(s) can be used in these experiments, the ribosome will stall upon reaching a stop codon. This allows puromycin, which is ligated through its nucleoside portion to the end of the mRNA chain, to enter the ribosomal P-site and stably and covalently conjugate the mRNA to the nascent polypeptide. If the mRNA used contains section(s) of random sequence, a library of varied peptides will result, each covalently attached to its respective coding sequence. The screening process for mRNA display is effectively the same as that of phage display, involving binding to an immobilised target, washing away of non-binding peptides, elution of the nucleic acid that codes for the target-binding peptides, and replication of the genetic material by

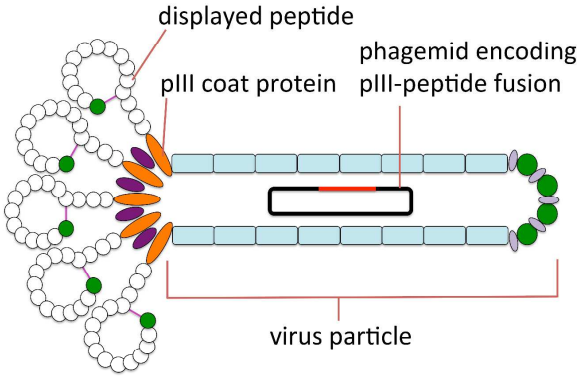

Fig. 1 Cartoon representation of peptides displayed on a phage particle.

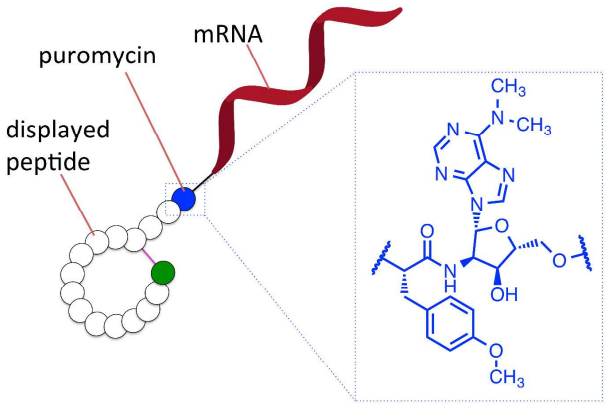

Fig. 2 Cartoon representation of a peptide displayed on its mRNA, with expanded detail of the puromycin linker.

PCR before repeating the process. Appropriate negative selections against target immobilisation media are also carried out for this method.

Ribosome display is a similar but older method. It has not been used for *in vitro* selection of stable macrocyclic peptides, but is certainly capable of being used for this application and so is briefly mentioned here for completeness. This technique also uses a stalled ribosome to link a displayed peptide to its mRNA, in this case by maintaining an intact ribosome•peptide•mRNA non-covalent complex.[13] While it does not require the use of a puromycin linker, the required translation complex is fragile, which limits the conditions under which selection can be carried out.

Chemical diversification of libraries

Additional chemical diversity, and increasing foldamer-like properties, can be incorporated into the library resulting from either display method, by ribosomal incorporation of

nonproteinogenic (some artificial) amino acids and/or by chemical diversification following synthesis (Figure 3). The latter case has typically been for cyclization, although potential exists for other applications.

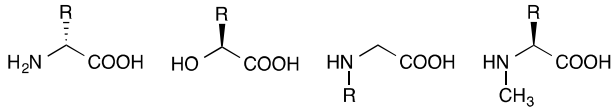

Fig. 3 Example non-standard building blocks able to be incorporated into macrocyclic peptides; D-amino acids, α-hydroxy acids, *N*-alkyl glycines, *N*-methyl amino acids.

Incorporation of non-proteinogenic amino acids is achieved by charging of a tRNA displaying an appropriate anticodon with the desired amino acid.[8] This can be by chemical synthesis of aminoacyl nucleotides followed by ligation to a truncated tRNA, by fortuitous misrecognition of a nonnatural amino acid by an aminoacyl-tRNA synthetase (ARS) or by use of a mutant ARS with altered substrate specificity and subsequent charging of the amino acid to its cognate tRNA, or by the use of tRNA acylating ribozymes, known as flexizymes. This last case, the combination of mRNA display with an amino-acylating ribozyme for genetic code reprogramming, is termed the Random nonstandard Peptide Integrated Discovery (RaPID) system.[14] By use of these systems, a large variety of non-proteinogenic amino acids have been incorporated, including D-amino acids, *N*-alkylated amino acids, α-hydroxy acids, and L-α-amino acids with sidechains having functional groups such as heterocycles, aliphatics, azides, alkynes, selenoethers, and halides.[1]

Macrocyclization of peptide libraries is usually carried out following translation by means of a chemoselective reaction that forms a pre-defined ring architecture. Incorporation of reactive non-natural amino-acids can also be used for macrocyclization, for example in reactions between alkyne and azide, benzylamine and hydroxy-tryptophan, or thiol and α-carbonyl halide.[14] The last of these, usually between a chloroacetylated side-chain or N-terminus and the sulfhydryl group of a cysteine incorporated into the same peptide (Figure 4a),[15,16] occurs spontaneously following translation. This method has been applied to mRNA display-based selections (*vide infra*), avoiding the need for addition of any exogenous cyclising reagents. Also possible is the conjugate addition of a cysteine side chain to dehydroalanine, itself derived from oxidative elimination of a selenoether-containing side-chain (Figure 4b).[17] For stable macrocyclization of peptides containing only canonical amino acids, selective reaction of a bi- (or tri-) functional auxiliary ligand is employed. Examples include reaction of amines (either the N-terminus or a lysine side-chain) with amidating reagents such as di(N-succinimidyl) glutarate (Figure 4c)[18] and reaction of thiols with alkylating agents such as bis(bromomethyl)benezene (Figure 4d).[19]

Applications of *in vitro* selected macrocyclic peptides

Stable macrocyclic peptides have been selected for a variety of functions using the techniques described in the section above. To date most of these applications have been based on inhibition of a target protein's function, which likely reflects the relative ease of disrupting interactions as compared to enhancing or making new interactions. In all of these cases, the inhibitory activity was shown to be optimal only upon closing of the relevant macrocycle. By far the most common target category has been that of enzyme inhibitors. Enzyme reactivities for which selection of macrocyclic peptide inhibitors has been carried out include a NAD^+-dependent histone deacetylase (SIRT2[20,21]), several proteases (kallikrein,[19] urokinase-type plasminogen activator uPa,[22] thrombin[23]), a transamidase (sortase A[17]), a ubiquitin ligase (E6AP[24]), and a kinase (Akt2[25]). In addition to enzyme inhibitors, stable macrocyclic peptide inhibitors have also been reported[26] for disruption of peptide-protein interaction in activation of vascular endothelial growth factor receptor 2 (VEGFR2) by its agonist the peptide hormone vascular endothelial growth factor (VEGF). Finally, inhibitors of two different drug transporters have also been described, from prokaryotic and eukaryotic sources; the bacterial Multidrug And Toxic compound Extrusion transporter from *Pyrococcus furiosus* (PfMATE)[27,28] and eukaryotic P-glycoprotein homologue CmABCB1,[29] a member of the ATP-binding cassette (ABC) superfamily, from the red alga *Cyanidioschyzon merolae*. The latter was developed as a co-crystallisation ligand, with inhibition being a fortuitous function.

X-ray crystal structures have been solved for several of these macrocyclic peptides bound to their targets, while others have been investigated in solution for formation of defined structure using techniques such as NMR and circular dichroism. These experiments show a surprising diversity of binding locations and structures adopted.

Binding locations of inhibitors, and effects on target protein structure

The binding locations for these functional macrocyclic peptides do not appear to be predictable on the basis of sequence or selection strategy, although enzyme inhibitors certainly tend to be biased towards the enzyme active site rather than allosteric sites in the structures solved to date. As an example of active site binding, the SIRT2 deacetylase-inhibiting peptide S2iL5, which contains a trifluoroacetamide mechanism-based inhibition warhead, clearly bound with this warhead situated in the active site for de-acetylation, with the rest of the peptide occupying the substrate binding site (Figure 5a).[21] Binding of the macrocyclic peptide to SIRT2 caused three conformational shifts near the binding pocket; a smaller domain moves closer to a larger domain to form a 'closed' architecture, a SIRT2-specific insertion shifts from the resting-state α-helix form to a loop that interacts with the macrocyclic peptide, while a cofactor (NAD^+) binding loop adopts a partially closed conformation midway between the apo-form and the cofactor

bound form. Overall, this peptide binds in a substrate-like manner, with lack of turnover coming from the trifluoroacetamide warhead.

All reported uPa peptidase-inhibiting bi-cyclic peptides[22,30] also bind in the enzyme's substrate-binding cleft (Figure 5b). In these cases, with the enzyme being a peptidase and the inhibitors being peptides, a means to prevent inhibitor turnover

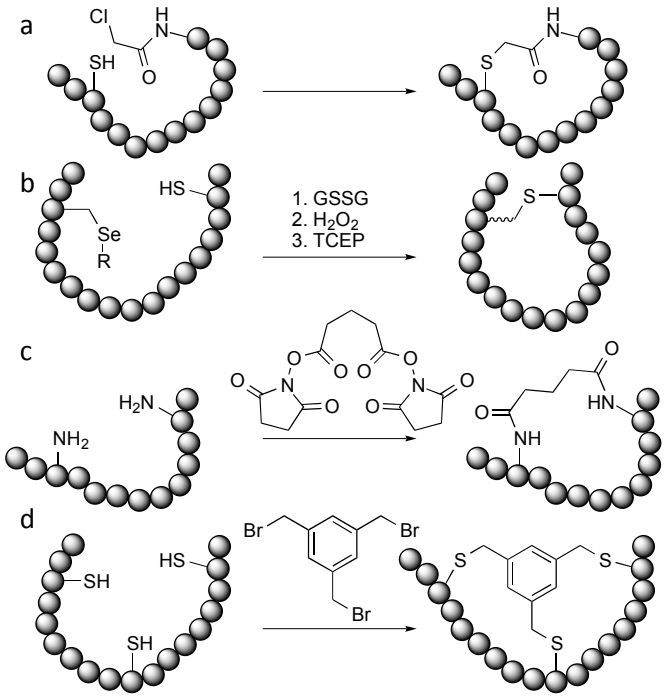

Fig. 4 Example macrocyclization strategies for peptides; thioetherification of a cysteinyl thiol with a *N*-chloroacetylated side chain or N-terminus (a), thioetherification of a cysteinyl thiol with a selenoether by thiol protection - oxidative elimination – thiol deprotection/conjugate addition (b), amidation of multiple side-chain or N-terminal amines with di(*N*-succinimidyl) glutarate (c), and alkylation of multiple cysteinyl thiols by 1,3,5-tris(bromomethyl)benzene (d).

is crucial. In one example,[22] the S1 subsite is occupied by a substrate-like arginine residue, but the macrocycle-induced conformation of the inhibitor peptide chain amide in the position of the normally scissile bond makes cleavage impossible (Figure 5e). By contrast in upain-1 (Figure 5c), a disulfide-containing mono-cyclic peptide selected against this same target,[31] it is the interaction of peptide side-chains with the catalytic residues that prevents hydrolysis (Figure 5f).[32] The indole ring from a tryptophan blocks access of the catalytic nucleophile to the scissile bond, while a carboxyl group from glutamate occupies the oxyanion hole and stabilises the conformation that prevents access of the catalytic nucleophile to the carbonyl group of the scissile bond. In this case, mutation of the sterically blocking tryptophan to an alanine converts the inhibitor to a slowly hydrolysed substrate. Several stable bicyclic peptides bind in a similar mode to upain-1,[30] with turnover prevented by similar interactions with the catalytic residues. One of these, however, is cleaved by the protease but remains bound as an ester through the C-terminus of the cleaved fragment to the active site nucleophile. Why turnover is not completed is not immediately apparent from the crystal structure and was not investigated further. Despite binding as a covalent adduct this compound was not a particularly high-affinity inhibitor of uPa, possibly due to compensating low k_{on} and k_{off} rates.

Macrocyclic peptide inhibitors of prokaryotic[27,28] and eukaryotic[29] drug exporters (PfMATE and CmABCB1, respectively) serve to illustrate several other binding modes. Four different macrocyclic peptides that inhibit PfMATE were crystallised, binding in two different locations.[27,28] The first pair of inhibitors, a larger macrocycle called MaL6 and a lasso-shaped ligand composed of a smaller macrocycle with a linear section called MaD8, bound superficially, between the N- and C-lobes of the extracellular side (Figure 5g). The other two peptides, two lasso-shaped macrocyclic peptides called MaD5 and MaD3S, were observed deeper inside the protein, in the substrate-binding pocket for target xenobiotics, with the C-terminal tail of these lariat-shaped inhibitors extending towards the surface of the protein (Figure 5h). These binding pocket-located peptides were the better of the two types of inhibitors. The tails of the two crystallised lariat structures were poorly ordered in the crystal structures, yet these peptides showed differing levels of inhibition despite containing the same core macrocycle, differing in the disordered tails. The nature of this tail is thus important for effective inhibition despite its interactions being unclear. However, this may also result from a change in cell penetration rather than target interaction. All of these peptides that inhibit PfMATE bind to the 'outward open' conformation of the transporter, with the critical trans-membrane helix 1 in its 'straight' conformation, primed for target export but unable to undergo the required kinking of this helix because of putative high affinity of the bound inhibitors.

By contrast, a single molecule of aCAP, the macrocyclic peptide inhibitor of CmABCB1 transporter, was observed to bind to the surface of one of two monomers of the target protein homodimer (Figure 5i), adjacent to where the outer leaflet lipid head-groups of the membrane would be.[29] This macrocyclic peptide binds to a bundle of transmembrane helices that are required for linking ATP hydrolysis to drug export. Binding of the peptide inhibitor acts as a staple, clamping these helices together and preventing the helix-bundle dissociation required for opening of the extracellular-space-facing gate of the drug-binding site, thus preventing transport.

Structures of bound macrocyclic peptide inhibitors

The macrocyclic peptides binding in these diverse locations are able to adopt higher order structure of their own, either by spontaneously and independently folding in solution or by assuming a conformation that optimises interactions with the target during binding, with the target acting as a template for

folding of the bound peptide. A surprising diversity of conformations has been observed in the small number of structures solved for stable macrocyclic peptide inhibitors. It is worth emphasising that these are all crystal structures, and so there may be a selection bias in play here for structured peptides that are better able to cocrystallise, and these may not be necessarily representative of structures that dominate in solution. However, there is no evidence for this being the case, and with this caveat in mind an examination follows of the peptide structures obtained to date.

Helix- and sheet-containing peptides

A few examples exist of protein-like secondary structure in stable macrocyclic peptide inhibitors. In one case, the macrocyclic peptide inhibitor aCAP, which 'staples' the transmembrane helices of the eukaryotic drug exporter CmABCB1,[29] forms a short section of α-helix with an unstructured loop connecting the two ends of the peptide (Figure 5l). This helix aligns itself pointing away from the target protein, perpendicular to the 'stapled' helices and parallel with the membrane. Target-binding interactions are formed with amino acids from both the helical and loop parts of the inhibitor.

In another example, a section of anti-parallel β-sheet-like backbone interactions is present in MaL6 (Figure 5j), the peptide inhibitor of the prokaryotic drug exporter PfMATE that consists of a large macrocycle.[28] In this case, almost the entire macrocycle is involved in the sheet structure, which is surrounded by target protein upon binding. This macrocyclic peptide is adopting a well-defined protein-like folded structure, making it an excellent example of foldamer-like behaviour in this class of semi-synthetic peptides. Finally, S2iL5, the warhead-containing macrocyclic peptide inhibitor of the deacetylase SIRT2[21] forms a backbone-to-backbone interaction with its target (Figure 5d), similar to that seen in an anti-parallel β-sheet, suggesting the possibility of formation of extended sections of intermolecular secondary structure. Such β-sheet interactions are common in protein-protein interactions as well as strongly interacting aggregates,[33] such as amyloid plaques[34] and silk fibres,[35] suggesting the many hydrogen bonds present in a small area in this motif may allow strong target binding.

S2iL5 also contains an unusual structural motif for self-folding, with an arginine side chain situating itself inside the peptide macrocycle and forming two direct and two water-mediated hydrogen bonds with other residue side chains and backbone amides. While not a traditional protein secondary structural element, this is a compact and well-defined fold, referred to as a 'power button icon' due to adjacent arginine and trifluoroacetyl-lysine side chains extending into the center of the macrocycle and outwards into the target protein active site, respectively. While binding is best as a macrocycle, this folding up around the side chain of an arginine means that linear S2iL5 remains an unusually good inhibitor, as pre-organisation may be able to occur without macrocyclization.

Extended peptides

Based on the structural information available to date, the majority of stable macrocyclic peptide inhibitors derived from *in vitro* selection bind to their targets in extended conformations (as exemplified in Figure 5e). This includes three of the four peptide inhibitors of PfMATE transporter, the SIRT2 deacetylase inhibitor S2iL5, and the bicyclic peptidase inhibitors derived from phage display (inhibiting kallikrein, uPa, and thrombin). Based on the abundance of protein-protein interactions mediated by short loops,[36] large elements of protein-like secondary structure is likely not necessary for strong target binding. Stable macrocyclic peptides that mimic such small loops with non-regular secondary structure should provide excellent scaffolds for strong interactions with proteins. Binding of the stable macrocyclic peptides covered in this review appears to match the general patterns for binding of macrocycles to proteins,[37] where an extended conformation is the predominant form for binding of larger (>600 Da) macrocycles. In 12 of these 22 non-redundant structures available for protein-bound macrocycles, no intramolecular hydrogen bonds were observed, with the remainder showing very few. Some flexibility in the macrocyclic peptide may indeed be beneficial, allowing the inhibitor to move and maintain contacts during dynamics of the target protein, as seen for small molecule transition state analogues.[38]

The bicyclic peptidase inhibitors derived from phage display in particular allow an examination of the effects of conformational constraint by cyclization, as representative bi- and mono-cyclic inhibitor structures are available, as well as their non-cyclised analogues. These compounds were cyclised by reactions of cysteine side-chains to either form disulfides, for the monocyclic compound, or be alkylated with an auxiliary ligand such as 1,3,5-tris(bromomethyl)benzene, for the bicyclic compounds. The bicyclic inhibitor of uPa,[22] UK18, showed potency dependent on the degree of cyclization, with a mono-cyclised form of the bi-cyclic inhibitor showing intermediate effectiveness (383 nM) between the bi-cyclic (53 nM) and linear analogues (17.5 μM). While no structure was solved for the linear or monocyclised analogues, it can be presumed that the weaker binding is the result of less pre-organisation of the ligand but similar potential interactions with the target, thus a change in entropic cost. By contrast upain-1, a disulfide monocyclic inhibitor of the same protein derived from a different selection experiment,[31] showed evidence of more self-folding (*vide infra*),[39] and thus more pre-organisation, yet exhibited a lower target affinity. This suggests that optimisation for stable folding in solution may not necessarily improve affinity or inhibition. This structure, stabilised by 3 β-turns comprised of 9 of its 12 amino acids and a carboxyl side-chain to amide nitrogen interaction, shows a high degree of self-bonding and organisation, yet it has a relatively low target affinity (~500 nM).

The effect of the chemical linker used for closing of the macrocycle structure was also investigated for macrocyclic peptide inhibitors of uPa.[30] Selection for inhibitory peptide

macrocycles was carried out by phage display, with the libraries cyclised using each of three different thiol-reactive compounds – 1,3,5-tris(bromomethyl)benzene/TBMB, 1,3,5-triacryloyl-1,3,5-triazinane/TATA, and *N*,*N'*,*N''*-(benzene- 1,3,5-triyl)-tris(2-bromoacetamide)/TBAB. Inhibition by these peptides was in most cases highly dependent on cyclization with the same linker as was used in the selection, although one

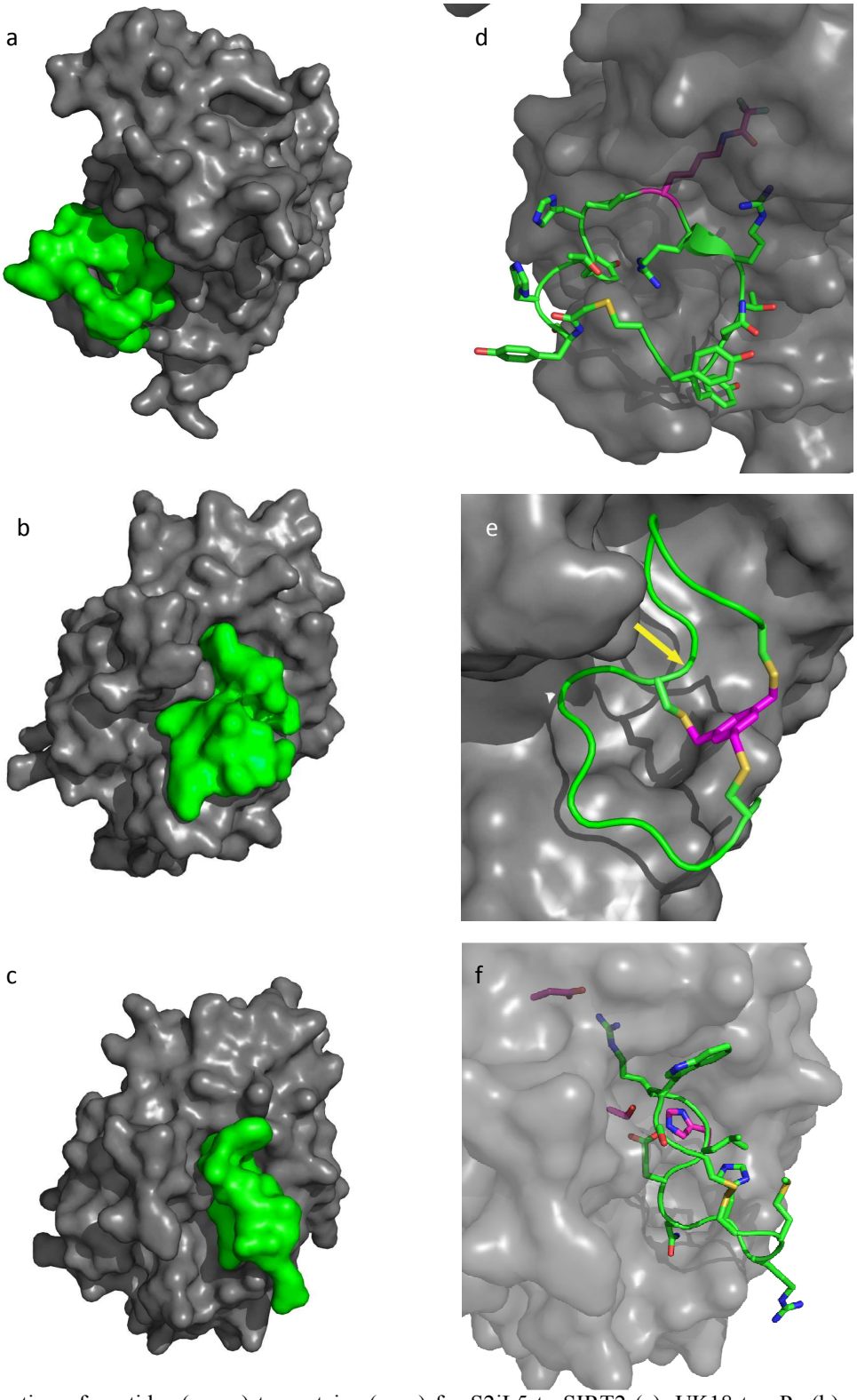

Fig. 5 Binding location of peptides (green) to proteins (gray) for S2iL5 to SIRT2 (a), UK18 to uPa (b), upain-1 to uPa (c). Close up of peptide structure for S2iL5, showing intramolecular hydrogen bonding around the central arginine side-chain (d), UK18, showing folding around the central ligand, with a yellow arrow indicating the site of induced unusual backbone conformation that prevents hydrolysis (e), upain-1, showing side-chain arrangements to prevent peptidase activity by the key catalytic side chains (f). In magenta, the mechanism based N-trifluoroacetamide warhead, the cyclising linker, and the key catalytic residues (respectively). Grey, protein target surface; green, peptide ligand; red, oxygen; blue, nitrogen; yellow, sulfur.

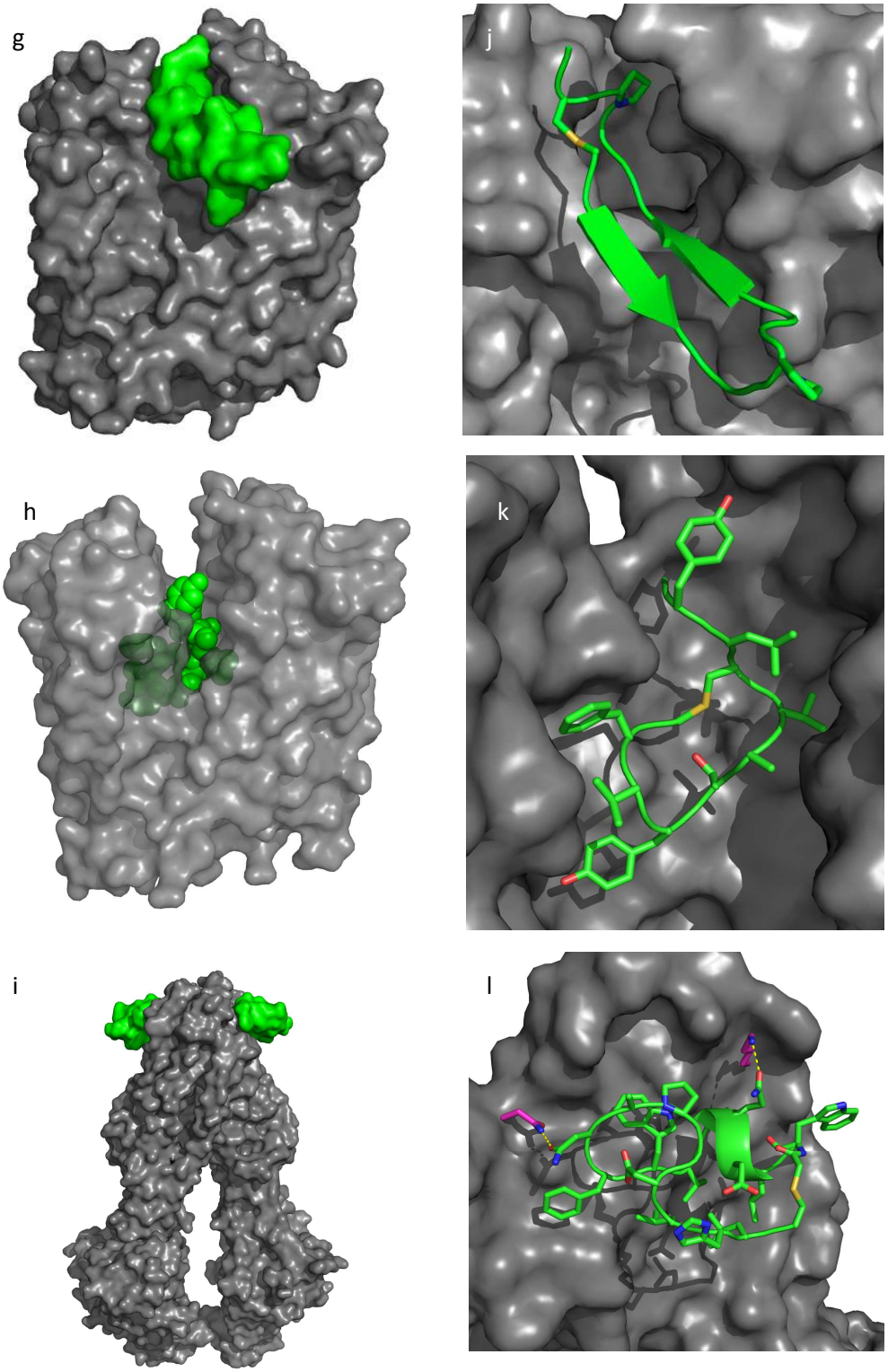

Fig. 5 (cont.) Binding location of peptides (green) to proteins (gray) for MaL6 to PfMATE (g), MaD3S to PfMATE (h), aCAP to CmABCB1 (i). Close up of peptide structure for MaL6, showing regions of beta sheet folding (j), MaD5, showing side-chain interactions (k), aCAP, showing intramolecular hydrogen-bonding to give helical folding (l). In magenta, the key catalytic side chains. Grey, protein target surface; green, peptide ligand; red, oxygen; blue, nitrogen; yellow, sulfur.

consensus motif was found that was still able to inhibit to some degree if either of the other two linkers were used. Four of the resultant macrocyclic peptide inhibitors were successfully co-crystallised with their target protein, with one of these crystal structures, that of peptide UK903, showing multiple peptide-to-cyclising-linker hydrogen bonds. Another macrocyclic peptide arising from the same selection, UK749, had more than twice as many inhibitor-to-target-protein hydrogen bonds but much less intramolecular hydrogen bonding, yet these two inhibitors showed comparable target-affinity. While hydrogen bonding is of course not the sole determinant of target affinity, this gives some indication of the benefits of ligand higher order structure.

It bears emphasising that macrocyclic peptides can be effective high affinity inhibitors of their targets, often with good selectivity, even when binding in these extended conformations. Substantial effort has been directed towards 'stapling' peptides from solid phase synthesis in order to stabilise common secondary structure motifs, and thus improve binding affinity.[40,41] What the peptide structures reviewed here show is that while macrocyclization does have benefits, a protein-like secondary structure is not necessarily best for tight binding. Indeed, stabilisation of a more irregular conformation by macrocyclization has recently been shown to still give substantial benefits to binding affinity, despite not adopting a conventional stable fold.[42]

Structures of unbound macrocyclic peptides

While no complete solution structures are available for any of the stable macrocyclic peptides discussed in this article, two reports of NMR experiments on macrocyclic peptide inhibitors give some interesting insights. The first of these,[19] investigating a set of 17 amino-acid bicyclic peptides selected against the peptidase kallikrein, showed evidence of an extended and likely disordered conformation, with no long range and few medium range NOEs, suggesting no interactions between loops and little pre-organisation of the ligand. A restrained simulated annealing protocol was used to analyse the best inhibitor, PK15, giving a model consistent with the available NOE data. This showed a loose organisation of the two extended peptide loops around the central linker. Despite the apparent flexibility of this prototypical peptide in solution, several of the peptides resulting from this selection exhibited high affinity for the target protein, down to the low nanomolar range.

Further investigation of PK15, the highest-affinity macrocyclic peptide inhibitor isolated from the selection mentioned in the previous paragraph, showed that the degree of self-organisation in solution was highly dependent on the nature of the cyclising linker.[43] Four linkers were tested in this work, of which three reacted well with thiols to give efficient cyclization (see the previous section for ligand names and abbreviations). While inhibition was unsurprisingly much stronger with TBMB, the original linker used in the selection, the solution NMR data with each of the other two linkers showed clear differences. Spectra from the peptide cyclised with TATA showed two different sets of spin systems, suggesting two different backbone conformations in solution. By contrast, spectra from the peptide cyclised by TBAB showed only a single spin system, but with many more long- and medium-range NOEs than either other ligand, suggesting a much more compact and ordered structure. This is consistent with the results, mentioned in the previous section, of co-crystallisation of a TBAB-cyclised peptide, which showed more self-interactions than those selected with other cyclising reagents, and suggests that such macrocyclic peptides are indeed able to pre-organise in solution prior to binding to a target protein.

The structure of upain-1, the disulfide-cyclised inhibitor of uPa, has also been investigated in solution.[39] Counter-intuitively, the core sequence of this peptide, the section bounded by the two cysteines forming the disulfide for cyclisation, was found to be a less effective inhibitor than its fusion protein with the phage pIII coat protein, and NMR solution structures of both the core and a shortened fusion peptide shed some light on this (Figure 6). Both the core and fusion structures adopt a two-turn structure in solution, although the residues that make up the turns are slightly different from one to the other. The extra residues in the fusion, in particular the backbone peptide of the three residues N-terminal to the disulfide-cyclised core, force adoption of a more rigid conformation in solution. Both these solution conformations are different from that exhibited in the target-protein bound crystal structure, which is largely the same for both the core and fusion peptides. In both the fusion and core peptides, binding to the target protease uPa leads to an induced conformational change around the critical tryptophan side chain that prevents inhibitor turnover (*vide supra*), and correspondingly both exhibit the same k_{on}. The increased affinity in the fusion peptide arises from a decrease in k_{off}, which may be a result of the dynamics for more rapid dissociation proceeding through a conformation that is permitted by the more flexible core peptide but not by the constrained fusion peptide.

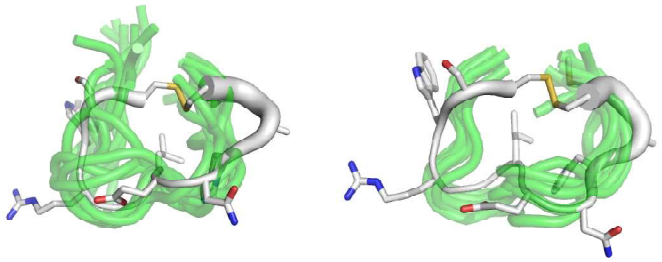

Fig. 6 Overlays of NMR solution structures (green) for unbound upain-1 core protein (left) and upain fusion protein (right) with a crystal structure of the upain fusion protein bound to its target (gray), width of the line represents crystallography B-factors.

In contrast to this induced-fit binding of upain-1, the 14 amino acid trypsin inhibitor SFTI-1 from sunflower seeds presents an example of a disulfide-constrained backbone

macrocyclic peptide that adopts the same rigid but extended fold both free in solution and when bound to its target.[44] Folding of this peptide is almost completely maintained in the case of an acyclic permutant, with a correspondingly small loss in binding affinity. This clearly demonstrates the ability of such a short peptide to fold up in solution, and for this folding to pre-organise it in a manner that facilitates tight binding.

Finally, also worthy of mention here is the work on solution structures of non-biological ATP-binding peptides derived from *in vitro* selection and subsequent evolution.[45,46] While not macrocyclic and substantially longer than the other peptides discussed here, with a minimal sequence of around 60 amino acids, these structures show the ability of a small protein derived from a library of random sequences to exhibit both stable folding and clear function. Through the use of the cyclization methods discussed in this review it seems plausible that a macrocyclic peptide of intermediate size may also be able to arrive at a similarly stable folded state in solution, without requiring the templating effects of a target protein.

Conclusions

Stable, foldamer-like, macrocyclic peptides are a promising class of compounds for many applications, representing a near-ideal balance of screening ability, functionality, stability, synthetic ease, and structural diversity. Through incorporation of non-proteinogenic amino acids and cyclization, followed by selection under appropriate conditions, many of the traditional drawbacks of peptide therapeutics using entirely natural amino acids can potentially be overcome. Macrocyclic structures have been observed to bind to a diverse range of locations on their target proteins, such as in active sites of enzymes, target binding pockets and exit tunnels for transporters, and areas of protein-protein interaction and across helices from adjacent sub-units in complexes. They are also able to adopt a variety of structures when bound to their targets, including short sections of well-defined protein-like secondary structure and, notably, extended conformations with large amounts of target-accessible surface area. There are also indications that macrocycles may be able to adopt well-defined stably folded structures in the absence of their binding partners. However, this does not appear to be an absolute requirement for effective binding. With regards to functional foldamer design, it is clear that mimicking elements of protein secondary structure is not an essential requirement for tight, specific binding. Induced fit binding may also be possible, or even advantageous, for synthetic foldamers. Further, the disorder present in many of the bound structures, exemplified by the PfMATE binders, highlights the potential benefit of foldamer structures that are not entirely rigid.

Acknowledgements

This work was supported by JSPS Postdoctoral Fellowships For Foreign Researchers to SJ and JR (P13026 and P13766).

Notes and references

1. K. Josephson, A. Ricardo, and J. W. Szostak, *Drug Discov. Today*, 2014, **19**, 388–399.
2. P. Li and P. P. Roller, *Curr. Top. Med. Chem.*, 2002, **2**, 325–341.
3. T. Rezai, J. E. Bock, M. V Zhou, C. Kalyanaraman, R. S. Lokey, and M. P. Jacobson, *J. Am. Chem. Soc.*, 2006, **128**, 14073–14080.
4. S. J. Bogdanowich-Knipp, S. Chakrabarti, T. D. Williams, R. K. Dillman, and T. J. Siahaan, *J. Pept. Res.*, 1999, **53**, 530–541.
5. M. Werle and A. Bernkop-Schnürch, *Amino Acids*, 2006, **30**, 351–367.
6. W. S. Horne, M. D. Boersma, M. A. Windsor, and S. H. Gellman, *Angew. Chem. Int. Ed. Engl.*, 2008, **47**, 2853–2856.
7. J. Buratto, C. Colombo, M. Stupfel, S. J. Dawson, C. Dolain, B. Langlois d'Estaintot, L. Fischer, T. Granier, M. Laguerre, B. Gallois, and I. Huc, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 883–887.
8. T. Passioura, T. Katoh, Y. Goto, and H. Suga, *Annu. Rev. Biochem.*, 2014, **83**, 727–752.
9. J. W. Kehoe and B. K. Kay, *Chem. Rev.*, 2005, **105**, 4056–4072.
10. J. Pande, M. M. Szewczyk, and A. K. Grover, *Biotechnol. Adv.*, 2010, **28**, 849–858.
11. J. W. Day, C. H. Kim, V. V Smider, and P. G. Schultz, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 2598–2600.
12. K. Ito, T. Passioura, and H. Suga, *Molecules*, 2013, **18**, 3502–3528.
13. J. A. Douthwaite and R. H. Jackson, Eds., *Ribosome display and related technologies*, Springer, 2012.
14. C. J. Hipolito and H. Suga, *Curr. Opin. Chem. Biol.*, 2012, **16**, 196–203.
15. Y. Goto, A. Ohta, Y. Sako, Y. Yamagishi, H. Murakami, and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 120–129.
16. Y. Sako, Y. Goto, H. Murakami, and H. Suga, *ACS Chem. Biol.*, 2008, **3**, 241–249.
17. F. T. Hofmann, J. W. Szostak, and F. P. Seebeck, *J. Am. Chem. Soc.*, 2012, **134**, 8038–8041.
18. S. W. Millward, T. T. Takahashi, and R. W. Roberts, *J. Am. Chem. Soc.*, 2005, **127**, 14142–14143.
19. C. Heinis, T. Rutherford, S. Freund, and G. Winter, *Nat. Chem. Biol.*, 2009, **5**, 502–507.
20. J. Morimoto, Y. Hayashi, and H. Suga, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 3423–3427.
21. K. Yamagata, Y. Goto, H. Nishimasu, J. Morimoto, R. Ishitani, N. Dohmae, N. Takeda, R. Nagai, I. Komuro, H. Suga, and O. Nureki, *Structure*, 2014, **22**, 345–352.
22. A. Angelini, L. Cendron, S. Chen, J. Touati, G. Winter, G. Zanotti, and C. Heinis, *ACS Chem. Biol.*, 2012, **7**, 817–821.
23. Y. V. G. Schlippe, M. C. T. Hartman, K. Josephson, and J. W. Szostak, *J. Am. Chem. Soc.*, 2012, **134**, 10469–10477.
24. Y. Yamagishi, I. Shoji, S. Miyagawa, T. Kawakami, T. Katoh, Y. Goto, and H. Suga, *Chem. Biol.*, 2011, **18**, 1562–1570.
25. Y. Hayashi, J. Morimoto, and H. Suga, *ACS Chem. Biol.*, 2012, **7**, 607–613.
26. T. Kawakami, T. Ishizawa, T. Fujino, P. C. Reid, H. Suga, and H. Murakami, *ACS Chem. Biol.*, 2013, **8**, 1205–1214.
27. C. J. Hipolito, Y. Tanaka, T. Katoh, O. Nureki, and H. Suga, *Molecules*, 2013, **18**, 10514–10530.
28. Y. Tanaka, C. J. Hipolito, A. D. Maturana, K. Ito, T. Kuroda, T. Higuchi, T. Katoh, H. E. Kato, M. Hattori, K. Kumazaki, T. Tsukazaki, R. Ishitani, H. Suga, and O. Nureki, *Nature*, 2013, **496**, 247–251.
29. A. Kodan, T. Yamaguchi, T. Nakatsu, K. Sakiyama, C. J. Hipolito, A. Fujioka, R. Hirokane, K. Ikeguchi, B. Watanabe, J. Hiratake, Y. Kimura, H. Suga, K. Ueda, and H. Kato, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 4049–4054.
30. S. Chen, D. Bertoldo, A. Angelini, F. Pojer, and C. Heinis, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 1602–1606.
31. M. Hansen, T. Wind, G. E. Blouse, A. Christensen, H. H. Petersen, S. Kjelgaard, L. Mathiasen, T. L. Holtet, and P. A. Andreasen, *J. Biol. Chem.*, 2005, **280**, 38424–38437.
32. G. Zhao, C. Yuan, T. Wind, Z. Huang, P. A. Andreasen, and M. Huang, *J. Struct. Biol.*, 2007, **160**, 1–10.
33. P. Cheng, J. D. Pham, and J. S. Nowick, *J. Am. Chem. Soc.*, 2013, **135**, 5477–5492.

34. L. C. Serpell, *Biochim. Biophys. Acta*, 2000, **1502**, 16–30.
35. A. Simmons, E. Ray, and L. W. Jelinski, *Macromolecules*, 1994, **27**, 5235–5237.
36. J. Gavenonis, B. A. Sheneman, T. R. Siegert, M. R. Eshelman, and J. A. Kritzer, *Nat. Chem. Biol.*, 2014, **10**, 716–722.
37. E. A. Villar, D. Beglov, S. Chennamadhavuni, J. A. P. Jr, D. Kozakov, S. Vajda, and A. Whitty, *Nat. Chem. Biol.*, 2014, **10**, 723–731.
38. J. S. Hirschi, K. Arora, C. L. Brooks, and V. L. Schramm, *J. Phys. Chem. B*, 2010, **114**, 16263–16272.
39. L. Jiang, A. S. P. Svane, H. P. Sørensen, J. K. Jensen, M. Hosseini, Z. Chen, C. Weydert, J. T. Nielsen, A. Christensen, C. Yuan, K. J. Jensen, N. C. Nielsen, A. Malmendal, M. Huang, and P. A. Andreasen, *J. Mol. Biol.*, 2011, **412**, 235–50.
40. J. A. Robinson, *Acc. Chem. Res.*, 2008, **41**, 1278–1288.
41. V. Azzarito, K. Long, N. Murphy, and A. Wilson, *Nat. Chem.*, 2013, **5**, 161–173.
42. A. Glas, D. Bier, G. Hahne, C. Rademacher, C. Ottmann, and T. N. Grossmann, *Angew. Chem. Int. Ed. Engl.*, 2014, **53**, 2489–2493.
43. S. Chen, J. Morales-Sanfrutos, A. Angelini, B. Cutting, and C. Heinis, *ChemBioChem*, 2012, **13**, 1032–1038.
44. M. L. Korsinczky, H. J. Schirra, K. J. Rosengren, J. West, B. A. Condie, L. Otvos, M. A. Anderson, and D. J. Craik, *J. Mol. Biol.*, 2001, **311**, 579–591.
45. S. S. Mansy, J. Zhang, R. Kümmerle, M. Nilsson, J. J. Chou, J. W. Szostak, and J. C. Chaput, *J. Mol. Biol.*, 2007, **371**, 501–513.
46. M. D. Smith, M. A. Rosenow, M. Wang, J. P. Allen, J. W. Szostak, and J. C. Chaput, *PLoS One*, 2007, **2**, e467.