








Cite this: *Chem. Sci.*, 2022, 13, 9079

All publication charges for this article have been paid for by the Royal Society of Chemistry

## Identification of novel functional mini-receptors by combinatorial screening of split-WW domains†

Hermann Neitz, <sup>‡c</sup> Niels Benjamin Paul, <sup>‡d</sup> Florian R. Häge, <sup>ae</sup> Christina Lindner, <sup>ae</sup> Roman Graebner, <sup>ae</sup> Michael Kovermann <sup>\*b</sup> and Franziska Thomas <sup>\*ae</sup>

$\beta$ -Sheet motifs such as the WW domain are increasingly being explored as building blocks for synthetic biological applications. Since the sequence-structure relationships of  $\beta$ -sheet motifs are generally complex compared to the well-studied  $\alpha$ -helical coiled coil (CC), other approaches such as combinatorial screening should be included to vary the function of the peptide. In this study, we present a combinatorial approach to identify novel functional mini-proteins based on the WW-domain scaffold, which takes advantage of the successful reconstitution of the fragmented WW domain of hPin1 (hPin1<sub>WW</sub>) by CC association. Fragmentation of hPin1<sub>WW</sub> was performed in both loop 1 (CC-hPin1<sub>WW</sub>-L1) and loop 2 (CC-hPin1<sub>WW</sub>-L2), and the respective fragments were linked to the strands of an antiparallel heterodimeric CC. Structural analysis by CD and NMR spectroscopy revealed structural reconstitution of the WW-domain scaffold only in CC-hPin1<sub>WW</sub>-L1, but not in CC-hPin1<sub>WW</sub>-L2. Furthermore, by using <sup>1</sup>H-<sup>15</sup>N HSQC NMR, fluorescence and CD spectroscopy, we demonstrated that binding properties of fragmented hPin1<sub>WW</sub> in CC-hPin1<sub>WW</sub>-L1 were fully restored by CC association. To demonstrate the power of this approach as a combinatorial screening platform, we synthesized a four-by-six library of N- and C-terminal hPin1<sub>WW</sub>-CC peptide fragments that was screened for a WW domain that preferentially binds to ATP over cAMP, phosphocholine, or IP6. Using this screening platform, we identified one WW domain, which specifically binds ATP, and a phosphorylcholine-specific WW-based mini-receptor, both having binding dissociation constants in the lower micromolar range.

Received 20th February 2022

Accepted 14th July 2022

DOI: 10.1039/d2sc01078j

rsc.li/chemical-science

## Introduction

Small, independently folding protein motifs are commonly used as scaffolds to mimic complex biological systems.<sup>1,2</sup> Among the protein folding motifs available, the coiled coil (CC) has probably been most frequently used in such endeavours to date, as the sequence-structure relationship is well understood.<sup>3</sup> Consequently, CCs have been applied in various synthetic biological designs including biomimetic receptors,<sup>4</sup> catalysts,<sup>5</sup>

interaction domains,<sup>6</sup> or vesicle fusion proteins.<sup>7</sup> Despite the obvious advantages of CCs in the design of biomimetic systems, the disadvantages should not be overlooked and illustrate why further small protein folding motifs are increasingly being explored for applications in synthetic biology. For instance, the rigidity of CC assemblies might impede the effective design of biomimetic receptors and enzymes, since flexibility of the binding pocket is an important factor in ligand or substrate binding.<sup>2,8</sup> Another obstacle is the frequent use of self-assembling CCs, very often homomers, which limits the engineerability and strand exchange is a significant problem.<sup>9</sup>

Among others, WW domains are a good complement to the toolbox of small protein motifs used for synthetic biological applications. WW domains are stable antiparallel  $\beta$ -sheet peptides with a length of about 35 amino acids, which are protein-protein interaction modules and recognize proline-rich amino acid sequences.<sup>10</sup> The name is derived from two characteristic and highly conserved tryptophan residues, which are crucial for structural stability and function.<sup>11,12</sup> Calculations have shown that WW domains cover a very high sequence space, which makes this motif ideal for design and engineering purposes.<sup>13</sup> This has been experimentally proven, as based on sequence alignments, artificial WW domains have been

<sup>a</sup>Institute of Organic Chemistry, Heidelberg University, Im Neuenheimer Feld 270, Heidelberg 69120, Germany. E-mail: franziska.thomas@oci.uni-heidelberg.de

<sup>b</sup>Department of Chemistry, University of Konstanz, Universitätsstraße 10, Konstanz 78457, Germany. E-mail: Michael.Kovermann@Uni-Konstanz.de

<sup>c</sup>Institute of Organic Chemistry, University of Würzburg, Am Hubland, Würzburg 97074, Germany

<sup>d</sup>Institute of Organic and Biomolecular Chemistry, University of Göttingen, Tammannstr. 2, Göttingen 37077, Germany

<sup>e</sup>Centre for Advanced Materials, Heidelberg University, Im Neuenheimer Feld 225, Heidelberg 69120, Germany

† Electronic supplementary information (ESI) available: The Supporting Information contains the complete description of the experimental procedures, additional tables and figures as well as the full characterization of the peptides investigated in this study. See <https://doi.org/10.1039/d2sc01078j>

‡ These authors have contributed equally.



designed or redesigned to modulate folding properties.<sup>14</sup> However, there are only a few examples showing successful engineering of WW domain function. In these cases, active sites located on the  $\beta$ -sheet surface of a natural protein were transferred more or less one-to-one to the  $\beta$ -sheet surface of a WW domain. For example, the WW domain of FBP11 was modified according to the natural model heat shock protein such that it binds specifically to single-stranded DNA,<sup>15</sup> and a zinc-binding WW domain was obtained by transferring the His<sub>3</sub>-site of carbonic anhydrase II to the surface of the hPin1 WW domain and was further engineered to obtain a Zn (II) sensor.<sup>16</sup>

However, intuitive design of new functional WW domains becomes almost impossible if there is no natural model with similar structural features for the intended new function. Therefore, we considered a combinatorial approach that allows screening of multiple WW-domain variants. In the past, it has been shown that the function of fragmented proteins can be reconstituted if their fragments are forced into close proximity.<sup>17</sup> For example, antiparallel CCs have been successfully used in the reconstitution of the green fluorescent protein or the enzyme luciferase.<sup>18</sup> Recently, in-cell proteolysis-based signalling and logic circuits have been developed, in which split proteases are attached to orthogonal CC dimerization domains.<sup>19</sup> Inspired by these reports and encouraged by a circularly permuted hyperstable WW domain,<sup>20</sup> we identified a split-WW-domain approach as a potential screening platform for exploring engineered WW domains with novel functions, in which a fragmented WW domain is reconstituted by association

of an antiparallel CC (A<sub>CC</sub>B<sub>CC</sub>, Fig. 1). This split approach would, if successful, facilitate the diversification of WW-domain sequences and thus the synthesis of WW-domain libraries. However, compared with previously reported systems, the WW domain is a small protein folding motif, and reconstitution of both structure and function in such a system requires thorough initial testing.

In this report, we present the successful reconstitution of a model WW domain, the WW domain of hPin1 (hPin1<sub>WW</sub>, hPin1(6-39)), by CC association. We have combined circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy to prove at first the proper structural reconstitution of hPin1<sub>WW</sub>. Then, NMR, fluorescence and CD spectroscopy was applied to investigate the functional binding of hPin1<sub>WW</sub> to the C-terminal domain peptide (CTD) of RNA polymerase II in a second step. As a result, our data show complete structural and functional reconstitution of fragmented hPin1<sub>WW</sub> in a CC-split-hPin1<sub>WW</sub> system. Based on this, we further developed the CC-split-hPin1<sub>WW</sub> concept into a combinatorial screening platform. As a proof-of-principle, we synthesized a four-by-six library of N- and C-terminal hPin1<sub>WW</sub>-CC peptide fragments that was successfully screened for WW domains possessing binding properties to organophosphates.

## Results and discussion

### Concept design and peptide synthesis

To demonstrate the concept of proximity-induced reconstitution of a fragmented WW domain based on CC association that could be used as a combinatorial screening platform, a suitable model WW domain had to be identified. It has been shown that a circularly permuted variant of hPin1<sub>WW</sub> reliably folds into a WW-domain scaffold.<sup>20</sup> However, the amino acid sequence of this variant was significantly changed to achieve structural stability. For instance,  $\beta$ -caps were introduced at the permuted N- and C-termini and a tryptophan cage was required for structural stabilization. Since such severe structural changes contradict the design of a screening platform for functional WW domains and significantly limit the accessible sequence space, we relied on wild type hPin1<sub>WW</sub> as the target peptide in our studies. If a moderately stable WW domain such as hPin1<sub>WW</sub> does not fold in a proximity-induced manner after fragmentation, the envisaged concept would generally be unsuitable for the intended purpose.

hPin1<sub>WW</sub> is a well characterized WW domain<sup>21-23</sup> and has shown to be remarkably stable to engineering.<sup>20,24</sup> Although the loops, more precisely loop1, play an important role in WW-domain folding,<sup>25,26</sup> they are less conserved compared to the  $\beta$ -sheet regions (Fig. S1†). Hence, we split hPin1<sub>WW</sub> at the GR site (G20, R21) of loop 1 (CC-hPin1<sub>WW</sub>-L1) and the HI site (H27, I28) of loop 2 (CC-hPin1<sub>WW</sub>-L2), respectively (Fig. 1). The fragments were attached to the strands of an antiparallel heterodimeric CC, more precisely, to *de novo* designed Acid-Kg Base-Eg,<sup>27</sup> which, in the following, are referred to as A<sub>CC</sub>B<sub>CC</sub> (Fig. 1). The conjugates of coil peptides and WW-domain fragments are referred to as B<sub>CC</sub>-hPin1<sub>WW</sub>-L1 or B<sub>CC</sub>-hPin1<sub>WW</sub>-L2 for the N-terminal WW-domain fragments linked to coil strand B<sub>CC</sub>,

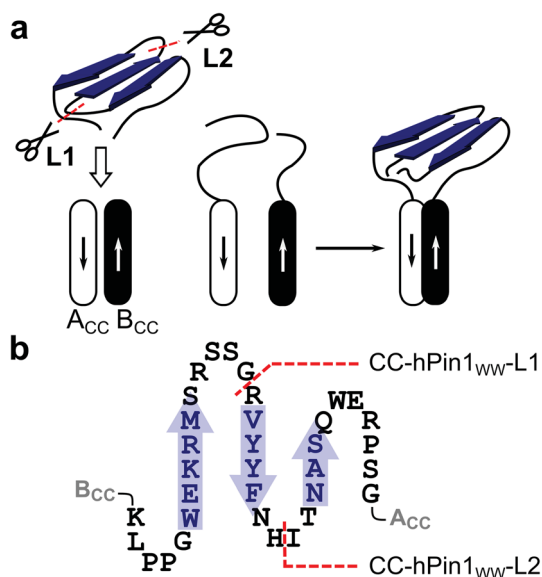


Fig. 1 Reconstitution of a fragmented WW domain on an antiparallel heterodimeric CC (A<sub>CC</sub>B<sub>CC</sub>). (a) Schematic representation of the structural concept applied in this study (A<sub>CC</sub> – acidic coil strand; B<sub>CC</sub> – basic coil strand). (b) Two-dimensional representation of the folding motif of the WW domain of hPin1 (hPin1<sub>WW</sub>) and the two possible fragmentation sites defining CC-hPin1<sub>WW</sub>-L1 (fragmentation at the GR site in loop 1) and CC-hPin1<sub>WW</sub>-L2 (fragmentation at the HI site in loop 2) in a WW-domain-CC construct. The  $\beta$ -sheets comprising hPin1<sub>WW</sub> are highlighted in blue and represented as arrows.



Table 1 Primary sequences for hPin1<sub>WW</sub> and hPin1<sub>WW</sub>-CC conjugates

Peptide	Sequence
hPin1 <sub>WW</sub>	H-KLPPGWKEKMRSSGRVYFNHITNAS QWERPSG-OH
B <sub>CC</sub> -hPin1 <sub>WW</sub> -L1	B <sub>CC</sub> -KLPPGWKEKMRSSG-OH
hPin1 <sub>WW</sub> -L1-A <sub>CC</sub>	H-RVYFNHITNASQWERPSG-A <sub>CC</sub>
B <sub>CC</sub> -hPin1 <sub>WW</sub> -L2	B <sub>CC</sub> -KLPPGWKEKMRSSGRVYFNH-OH
hPin1 <sub>WW</sub> -L2-A <sub>CC</sub>	H-ITNASQWERPSG-A <sub>CC</sub>

and hPin1<sub>WW</sub>-L1-A<sub>CC</sub> and hPin1<sub>WW</sub>-L2-A<sub>CC</sub> for the C-terminal WW-domain fragments linked to coil strand A<sub>CC</sub> (Table 1; note: The suffix L1/L2 identifies the fragmentation site – loop 1 or loop 2 – of the respective split-WW domains as well as their individual fragments). One-to-one mixtures of the related N- and C-terminal fragments then form CC-hPin1<sub>WW</sub>-L1 or CC-hPin1<sub>WW</sub>-L2.

All peptides were synthesized using microwave-assisted Fmoc/*t*Bu-based solid-phase peptide synthesis (SPPS). Interestingly, we observed an overall improved purity of the crude CC-hPin1<sub>WW</sub> peptides compared to crude hPin1<sub>WW</sub> (Fig. S2<sup>†</sup>), which we attribute to the known aggregation propensity of WW domains during the SPPS.<sup>28</sup> As only fragments of hPin1<sub>WW</sub> were attached to the CC strands, this effect was less pronounced.

### CD spectroscopic analysis of structural properties and thermodynamic stability of CC-hPin1<sub>WW</sub> variants

To study reconstitution of the fragmented CC-hPin1<sub>WW</sub> upon CC association, CD spectra and CD thermal denaturation profiles were recorded of both CC-hPin1<sub>WW</sub> constructs (CC-hPin1<sub>WW</sub>-L1 and CC-hPin1<sub>WW</sub>-L2), and hPin1<sub>WW</sub> and the antiparallel CC peptide A<sub>CC</sub>B<sub>CC</sub> were measured as references (Fig. 2). The CD spectrum of hPin1<sub>WW</sub> reveals the typical shape reported for WW domains showing a minimum at 197 nm, the characteristic shoulder or minimum at 206 nm and a maximum at

227 nm, which results from the disordered termini, the triple-stranded antiparallel  $\beta$ -sheet structure and the exciton coupling of the aromatic residues of the hydrophobic core, respectively.<sup>23</sup> The antiparallel CC peptide was  $\alpha$ -helical displayed by the strong minima at 208 and 222 nm. A CD spectrum of A<sub>CC</sub>B<sub>CC</sub> and hPin1<sub>WW</sub> in an one-to-one mixture was recorded as a reference spectrum (Fig. 2A, black dashed line). The CD spectra of CC-hPin1<sub>WW</sub>-L1 and CC-hPin1<sub>WW</sub>-L2 should have a similar profile in the case of a fully folded WW-domain region. Indeed, the CD spectrum of CC-hPin1<sub>WW</sub>-L1 (red) showed a good resemblance to this reference spectrum with an overall significantly reduced  $\alpha$ -helical content compared to A<sub>CC</sub>B<sub>CC</sub>, whereas the CD spectrum of CC-hPin1<sub>WW</sub>-L2 (green) revealed a higher  $\alpha$ -helical content (Fig. 2A). This observation alone led us to assume that the fragmented hPin1<sub>WW</sub> was more likely to be reconstituted in CC-hPin1<sub>WW</sub>-L1 than in CC-hPin1<sub>WW</sub>-L2. Deconvolution of the CD spectra using the CDSSTR algorithm on DichroWeb provided a more quantitative picture of the degree of folding of the fragmented WW-domain regions in CC-hPin1<sub>WW</sub>-L1 and CC-hPin1<sub>WW</sub>-L2.<sup>29</sup> The fractions of secondary structural elements are listed in Table 2. As expected, A<sub>CC</sub>B<sub>CC</sub> revealed an  $\alpha$ -helical content of almost 70% and hPin1<sub>WW</sub> a high  $\beta$ -strand content of 44% as well as 18% turn or 32% disordered regions. A folded WW-domain region in the CC-hPin1<sub>WW</sub> constructs should correspond to the fraction of secondary structure elements of the CD spectrum of the one-to-one mixture of hPin1<sub>WW</sub> and A<sub>CC</sub>B<sub>CC</sub>. The fraction of secondary structure in CC-hPin1<sub>WW</sub>-L1 very closely matches these values, which led us to conclude that WW-domain folding is indeed induced by CC association. In contrast, deconvolution of the CD spectrum of CC-hPin1<sub>WW</sub>-L2 revealed almost 60%  $\alpha$ -helical content and only 15%  $\beta$ -sheet content indicating, if at all, only partial folding of the fragmented WW domain.

Thermal denaturation profiles of hPin1<sub>WW</sub>, A<sub>CC</sub>B<sub>CC</sub> and the two CC-hPin1<sub>WW</sub> variants were recorded as the thermal stability gives additional information about the degree of folding. Thermodynamic analysis of the obtained profiles revealed cooperative unfolding and melting temperatures,  $T_m$ , of  $T_m = 58.5$  °C and  $T_m = 57.5$  °C for A<sub>CC</sub>B<sub>CC</sub> and hPin1<sub>WW</sub>, respectively.

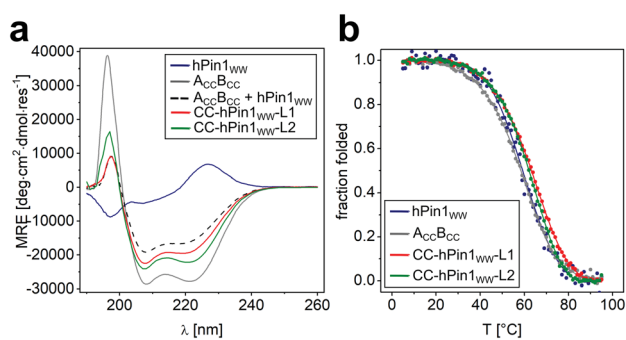


Fig. 2 CD spectroscopic data of CC-hPin1<sub>WW</sub> variants, hPin1<sub>WW</sub> and A<sub>CC</sub>B<sub>CC</sub>. (a) CD spectra of hPin1<sub>WW</sub> variants at 20 °C; (b) thermal denaturation profiles of hPin1<sub>WW</sub>, A<sub>CC</sub>B<sub>CC</sub>, CC-hPin1<sub>WW</sub>-L1, CC-hPin1<sub>WW</sub>-L2, and an one-to-one mixture of A<sub>CC</sub>B<sub>CC</sub> and hPin1<sub>WW</sub> to obtain a reference spectrum. Thermal denaturation profiles are depicted as fraction folded to enable comparison of the peptide species. Thermal denaturation profiles were fitted to a two-state folding model (see ESI<sup>†</sup>). Experimental conditions: PBS buffer, pH 7.4, 40  $\mu$ M peptide concentration.

Table 2 Fraction of secondary structural elements and melting temperatures,  $T_m$ , obtained for CC-hPin1<sub>WW</sub> variants and reference peptides based on CD spectroscopy

Peptide	Fraction secondary structure <sup>a</sup>				$T_m$ <sup>b</sup> [°C]
	$\alpha$ -Helix	$\beta$ -Strand	Turn	Disordered	
A <sub>CC</sub> B <sub>CC</sub> (CC)	0.69	0.11	0.10	0.12	58.5 $\pm$ 0.5
hPin1 <sub>WW</sub>	0.04	0.44	0.18	0.32	57.5 $\pm$ 1.0
CC/hPin1 <sub>WW</sub> (1 : 1)	0.48	0.23	0.14	0.19	—
CC-hPin1 <sub>WW</sub> -L1	0.47	0.24	0.11	0.18	65.0 $\pm$ 0.5
CC-hPin1 <sub>WW</sub> -L2	0.58	0.15	0.07	0.20	64.0 $\pm$ 0.5

<sup>a</sup> Fraction of secondary structure was obtained from deconvolution of CD spectra on DichroWeb using the CDSSTR algorithm (data set 4).<sup>29</sup>

<sup>b</sup>  $T_m$  values were calculated by least-square fitting of the thermal denaturation profiles assuming a two-state folding model for a monomer (hPin1<sub>WW</sub>) or a dimer (A<sub>CC</sub>B<sub>CC</sub>, CC-hPin1<sub>WW</sub>-L1, CC-hPin1<sub>WW</sub>-L2).

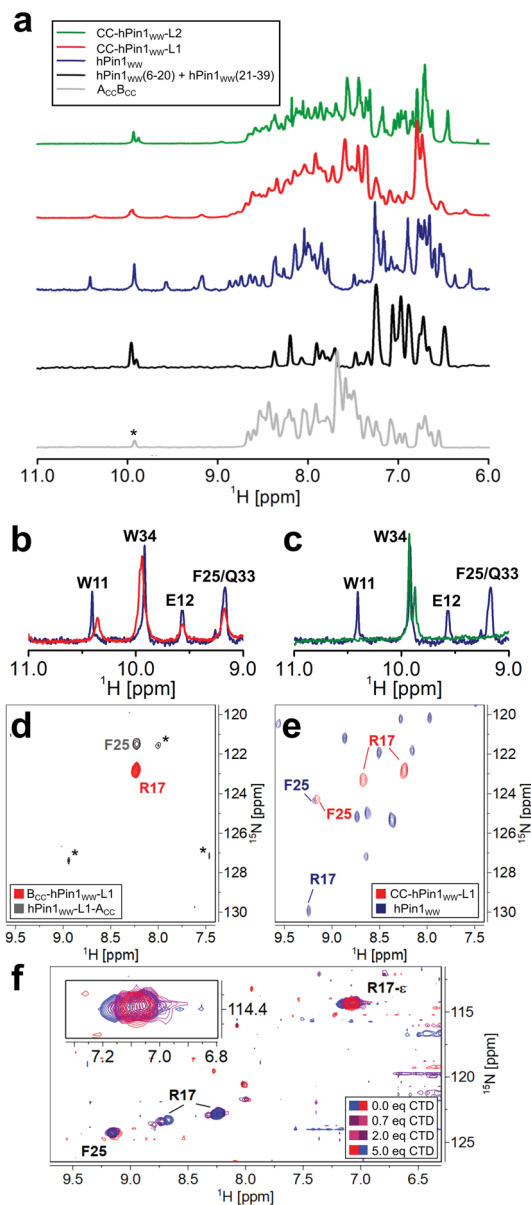


The  $T_m$  of hPin<sub>1WW</sub> is in accordance to the literature,<sup>20,23</sup> whereas the  $T_m$  of A<sub>CC</sub>B<sub>CC</sub> was increased by 18 °C from the reported literature value.<sup>27</sup> This discrepancy can best be explained by the lack of N- and C-terminal glycine residues in the studied sequences of the original publication (see Table S1†). The  $T_m$  values of the CC-hPin<sub>1WW</sub> variants were increased by 7 and 6 °C compared to the reference peptides, respectively. This led us to assume that the interaction of the CC is stabilized against thermal denaturation. However, CC-hPin<sub>1WW</sub>-L2 was slightly less stable, which indicates a less productive interaction of the WW-domain fragments.

The fact that the WW domain in CC-hPin<sub>1WW</sub>-L2 was not properly folded led to the hypothesis that the apparent reconstitution of the WW-domain structure in CC-hPin<sub>1WW</sub>-L1 was not solely due to CC association. Therefore, we synthesized the WW-domain fragments hPin<sub>1WW</sub>(6-20) and hPin<sub>1WW</sub>(21-39) lacking the CC tag and recorded CD spectra of an one-to-one mixture subsequently at different temperatures (Fig. S3†). When mixed at room temperature, the WW-domain structure was not reconstituted; however, at 0 °C, the CD spectrum showed the typical signals of an at least partially folded WW domain. This indicates that the structural information of split hPin<sub>1WW</sub> was not completely lost, which may positively contribute to the successful reconstitution into CC-hPin<sub>1WW</sub>-L1. However, the recorded thermal denaturation profile indicates a melting temperature below 0 °C, confirming that additional structural stabilization, e.g. by a CC dimerization domain, is required to achieve a stable assembly.

### NMR spectroscopic analysis of reconstitution of structure and function in CC-hPin<sub>1WW</sub> variants

NMR spectroscopic studies were performed to substantiate the conclusions on reconstitution of fragmented hPin<sub>1WW</sub> upon CC association in the CC-hPin<sub>1WW</sub> variants as obtained by CD spectroscopy so far. First, one-dimensional (1D) proton NMR spectra of hPin<sub>1WW</sub>, CC-hPin<sub>1WW</sub>-L1, CC-hPin<sub>1WW</sub>-L2, and, as control, of A<sub>CC</sub>B<sub>CC</sub> as well as an 1 : 1-mixture of hPin<sub>1WW</sub>(6-20) and hPin<sub>1WW</sub>(21-39) were recorded. All NMR spectra except the one-to-one mixture of hPin<sub>1WW</sub>(6-20) and hPin<sub>1WW</sub>(21-39) reveal predominantly folded peptide species. The spectrum of hPin<sub>1WW</sub> gives characteristic isolated proton resonance signals in the low field range between 9 and 11 ppm, which correspond to the indole NH protons of W11 and W34, as well as to the NH backbone protons of E12, F25 and Q33 (Fig. 3A–C, see Fig. S4† for assignment of NH backbone protons comprising hPin<sub>1WW</sub>). In this spectral range, no signals from neither folded A<sub>CC</sub>B<sub>CC</sub> nor the 1 : 1 mixture comprising hPin<sub>1WW</sub>(6-20) and hPin<sub>1WW</sub>(21-39), which show a pattern typical for non-structured peptide motifs, are resolved (Fig. 3A, grey and black traces, Fig. S5†), making these signals a strong indicator for successful structural reconstitution of the WW domain in the CC-hPin<sub>1WW</sub> variants. Indeed, the characteristic signals were clearly visible in the 1D proton spectrum of CC-hPin<sub>1WW</sub>-L1 with slightly altered chemical shifts of indole signals (Fig. 3B). CC-hPin<sub>1WW</sub>-L2, however, does not give this characteristic pattern of signals, which supports our findings obtained



**Fig. 3** NMR spectroscopic analysis of CC-hPin<sub>1WW</sub> variants assessing reconstitution of structure and function of fragmented hPin<sub>1WW</sub>. (a) 1D proton NMR spectra of A<sub>CC</sub>B<sub>CC</sub>, a 1 : 1 mixture of hPin<sub>1WW</sub>(6-20) and hPin<sub>1WW</sub>(21-39), hPin<sub>1WW</sub>, CC-hPin<sub>1WW</sub>-L1, and CC-hPin<sub>1WW</sub>-L2. The isolated proton NMR signals between 9 and 11 ppm are characteristic for natively folded hPin<sub>1WW</sub>. (\*: A<sub>CC</sub>B<sub>CC</sub> contains a tryptophan residue in A<sub>CC</sub> that is not present in the A<sub>CC</sub>-segment of hPin<sub>1WW</sub>-A<sub>CC</sub>). (b) Overlay of the low field proton NMR spectra of hPin<sub>1WW</sub> (blue) and CC-hPin<sub>1WW</sub>-L1 (red). The characteristic hPin<sub>1WW</sub> signals are also resolved in CC-hPin<sub>1WW</sub>-L1. (c) Overlay of the low field proton NMR spectra of hPin<sub>1WW</sub> (blue) and CC-hPin<sub>1WW</sub>-L2 (green). The characteristic hPin<sub>1WW</sub> signals are not resolved in CC-hPin<sub>1WW</sub>-L2. (d and e) Overlay of 2D <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra of B<sub>CC</sub>-hPin<sub>1WW</sub>-L1-<sup>15</sup>N<sub>4</sub><sup>13</sup>C<sub>6</sub>-R17 and hPin<sub>1WW</sub>-L1-A<sub>CC</sub>-<sup>15</sup>N-F25 (d), and CC-hPin<sub>1WW</sub>-L1-<sup>15</sup>N<sub>4</sub><sup>13</sup>C<sub>6</sub>-R17, <sup>15</sup>N-F25 and hPin<sub>1WW</sub> (e). Folding of CC-hPin<sub>1WW</sub>-L1 in (e) is apparent from the significant low field shift of F25 compared to (d) and the appearance of two cross-peaks for R17 indicating two backbone conformations. (\*: refers to noise signals; low signal intensity of cross-peaks of F25 in apo state due to solvent exchange) (f) chemical shift perturbation of R17 and R17-ε in <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra upon titration of CTD to CC-hPin<sub>1WW</sub>-L1 at T = 283 K. The signal of R17-ε is folded into the spectral range shown.



Table 3 Hydrodynamic radii of hPin<sub>1WW</sub>, CC-hPin<sub>1WW</sub>-L1 and CC-hPin<sub>1WW</sub> monomers

	Measured $r_h$ [nm]	Calculated $r_h^a$ [nm]	Meas./calc.
hPin <sub>1WW</sub>	1.45	1.33 (35 aa)	1.09
CC-hPin <sub>1WW</sub> -L1	2.20	1.78 (94 aa)	1.24
B <sub>CC</sub> -hPin <sub>1WW</sub> -L1	2.15	1.43 (45 aa)	1.50
hPin <sub>1WW</sub> -L1-A <sub>CC</sub>	1.78	1.48 (49 aa)	1.20
Calc. av. CC-hPin <sub>1WW</sub> monomers	1.98	1.46	1.35

<sup>a</sup>  $r_h$  is calculated for a globularly folded protein comprising the same number of residues as the particular variant of hPin<sub>1WW</sub> which has been measured.<sup>30</sup>

from the CD measurements that hPin<sub>1WW</sub> is not reconstituted in CC-hPin<sub>1WW</sub>-L2 upon CC association. The analysis of the NMR structure of hPin<sub>1WW</sub> published by Luh *et al.* revealed a conformationally defined loop 2 with a stabilizing hydrogen bond between the side chain at N26 and amide protons comprising the loop backbone (Fig. S6†).<sup>22</sup> Therefore, we assume that loop 2 is crucial for the structural integrity of hPin<sub>1WW</sub> and, for this reason, not suitable to act as a fragmentation site. Consequently, CC-hPin<sub>1WW</sub>-L2 was not considered in further experimental work in the present study.

Additionally, we have performed NMR diffusion experiments to determine the hydrodynamic radii ( $r_h$ ) of CC-hPin<sub>1WW</sub>-L1, hPin<sub>1WW</sub> and the monomers B<sub>CC</sub>-hPin<sub>1WW</sub>-L1 and hPin<sub>1WW</sub>-L1-A<sub>CC</sub> (Table 3, Fig. S7†). The hydrodynamic radius reports on a quantitative measure of the state of folding of the peptides investigated.<sup>30</sup> The value determined for hPin<sub>1WW</sub> corresponds well with the theoretical value for a globularly folded protein structure comprising the same number of residues, while the experimental value for CC-hPin<sub>1WW</sub>-L1 slightly deviates from the theoretical value. We attribute this to the rod-shaped CC extension of the peptide. However, compared to CC-hPin<sub>1WW</sub>-L1 monomers,  $r_h$  determined for CC-hPin<sub>1WW</sub>-L1 indicates a more globularly folded peptide, which becomes evident from the calculated average  $r_h$  values of B<sub>CC</sub>-hPin<sub>1WW</sub>-L1 and hPin<sub>1WW</sub>-L1-A<sub>CC</sub>.

To obtain insights into the structural and functional characteristics of CC-hPin<sub>1WW</sub>-L1 at atomic resolution, we incorporated <sup>15</sup>N-isotopically labelled amino acids, more specifically, <sup>15</sup>N<sub>4</sub><sup>13</sup>C<sub>6</sub>-R17 and <sup>15</sup>N-F25, since these are located (a) close to the fragmentation site and (b) to the ligand binding site. First, we studied the overall conformation of the fragmented WW domain by recording two-dimensional (2D) <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra of the isotopically labelled monomers (Fig. 3D) and CC-hPin<sub>1WW</sub>-L1, respectively. The chemical shift values of the backbone amide protons of R17 (8.23 ppm) and F25 (8.23 ppm) found for the individual monomers are close to the reported random coil chemical shifts of 8.27 ppm and 8.23 ppm,<sup>31</sup> thus indicating a disordered structure of the WW-domain fragments.

In CC-hPin<sub>1WW</sub>-L1 F25 is strongly low field shifted, indicating a  $\beta$ -sheet character, and also almost matches the chemical shift values found for F25 in hPin<sub>1WW</sub> (Fig. 3e). However, two conformations of the R17 backbone are evident, and the values of the <sup>15</sup>N chemical shifts differ significantly from those of the reference peptide hPin<sub>1WW</sub>. Both phenomena are likely due to the fact that R17 is part of the unstructured and

fragmented loop 1 and therefore can adopt different stable conformations compared to the non-fragmented parent peptide.

Since structural reconstitution of fragmented hPin<sub>1WW</sub> in CC-hPin<sub>1WW</sub>-L1 was proven independently by CD and NMR spectroscopy, we were particularly interested in whether this CC-induced reconstitution of structure also leads to reconstitution of the function. hPin1 is a peptidyl-prolyl-*cis/trans* isomerase that specifically isomerizes protein sequences with a phosphorylated S/TP motif recognized by hPin<sub>1WW</sub>.<sup>12,32</sup> One of the target proteins is the RNA polymerase II, and the typical recognition motif is the heptapeptide YSPTpSPS (pS – phosphorylated serine residue), which occurs repeatedly in the C-terminal domain and is therefore called CTD peptide.<sup>33</sup> We synthesized this peptide, titrated it to <sup>15</sup>N-isotopically labelled CC-hPin<sub>1WW</sub>-L1 and monitored the changes in the chemical shift values of cross-peaks belonging to the backbone and side chain of R17 in 2D <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra (Fig. 3f). By plotting the changes in chemical shift values against the stoichiometric ratio using the ligand and CC-hPin<sub>1WW</sub>-L1 concentrations, we were able to qualitatively estimate the binding dissociation constant ( $K_d$ ) that was in the range of 50  $\mu$ M (Fig. S8†). This is in good agreement with the  $K_d$  values previously reported for hPin<sub>1WW</sub>.<sup>34,35</sup> Since R17 is directly involved in ligand binding in native hPin<sub>1WW</sub>, these results suggest successful reconstitution of binding capacity to the CTD peptide in CC-hPin<sub>1WW</sub>-L1.

### hPin<sub>1WW</sub> and CC-hPin<sub>1WW</sub>-L1 show similar binding to CTD-peptide

Although the results obtained from the NMR titration experiment were promising, we aimed for an additional experiment to confirm the binding capacity between CTD and CC-hPin<sub>1WW</sub>-L1 and the value of the  $K_d$ . Furthermore, a direct comparison of the binding capacities to CTD peptide of hPin<sub>1WW</sub> and CC-hPin<sub>1WW</sub>-L1 improves the assessment of the extent of reconstitution regarding function of the fragmented WW-domain region in CC-hPin<sub>1WW</sub>-L1 significantly. Therefore, we measured the change in intrinsic tryptophan fluorescence of hPin<sub>1WW</sub> and CC-hPin<sub>1WW</sub>-L1 upon CTD peptide titration. It should be noted that direct titration of the ligand into the solution of the respective hPin<sub>1WW</sub> variant leads to false results, since tryptophan fluorescence in WW domains is highly sensitive to photobleaching.<sup>36</sup> Therefore, we decided to perform this experiment in a microplate format,



although signal scattering is sometimes an issue. An one-to-one dilution series of the ligand was prepared starting from 500  $\mu\text{M}$  CTD concentration, each dilution equilibrated with 2  $\mu\text{M}$  protein for 15 h at 4  $^{\circ}\text{C}$ , and measured in a microplate reader at room temperature to obtain the intrinsic tryptophan fluorescence at 350 nm. Both WW-domain peptides showed a very similar binding behaviour (Fig. 4), although the observed fluorescence enhancement was more pronounced for CC-hPin1<sub>WW</sub>-L1 than for hPin1<sub>WW</sub>. Surprisingly, the saturation binding curves indicated stronger binding for both peptides than initially indicated by NMR titration, and the non-linear least square fits (ESI Eqn. 14†) yielded  $K_d$  values of  $2.2 \pm 0.6 \mu\text{M}$  and  $2.0 \pm 0.7 \mu\text{M}$  for hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1, respectively. These values are by the order of one magnitude lower than those reported in the literature<sup>35</sup> and estimated from NMR titration experiments (Fig. S8†). One possible explanation for this discrepancy could be a reduced effect of undesired ligand depletion, since the measurements were performed at peptide concentrations 30 times lower than the reported  $K_d$ .<sup>37</sup> Additionally, we suspected conformational changes of the WW-domain backbone during ligand binding, which could result in cooperative changes of the fluorescence of both tryptophans, W11 in the hydrophobic core and W34 at the binding site, and thus to an apparent  $K_d$  value influenced by structural changes. At the extreme, such behaviour was observed in the Ess1<sub>WW</sub> domain during binding of similar CTD ligands. Ess1<sub>WW</sub> showed ligand-induced folding and provided a saturation-binding curve that indicated strong ligand binding.<sup>38</sup>

Therefore, we also performed CD measurements of hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1 in the presence of equimolar amounts of ligand CTD to determine possible structural changes. Indeed, CD spectra showed increased signal intensities for both, hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1 (Fig. 4b and d). Furthermore, CD thermal denaturation revealed cooperative unfolding with significantly steeper transition profiles and slightly lower melting temperatures (Fig. 4d, e and Table 4). The thermodynamic analysis of the thermal denaturation curves focusing on  $\Delta H_m$ ,  $\Delta C_{p,m}$  and  $\Delta G_{Di}$  at 20  $^{\circ}\text{C}$  using a two-state unfolding-folding model gave thermodynamic parameters that indicate a moderate structural stabilization of both hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1 upon ligand binding, respectively. Taken together, ligand titration monitored by NMR and fluorescence spectroscopy and CD measurements of hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1 gave similar results for both peptides, which suggest that both the structure and function of the fragmented WW-domain region, are reconstituted successfully in CC-hPin1<sub>WW</sub>-L1 upon CC association.

### A library of split-WW domains

As mentioned in the introduction, the concept of CC mediated reconstitution of split-WW domains was developed here to be used as a combinatorial approach to design WW domains possessing new functions. Advantageously, sets of N-terminal and C-terminal fragments can be combined, significantly expanding the size of the library while minimizing synthesis effort. To illustrate this, a simple example is given: Four N-terminal fragments combined with four C-terminal fragments

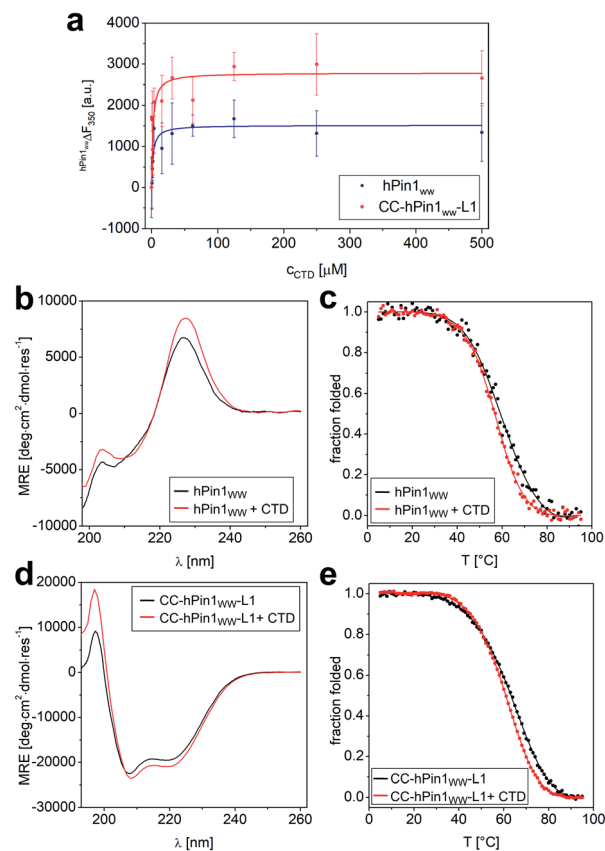


Fig. 4 Fluorescence and CD spectroscopic analysis of CTD binding to hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1. (a) Intrinsic tryptophan fluorescence titrations to determine dissociation constants ( $K_d$ ) of hPin1<sub>WW</sub> and CC-hPin1<sub>WW</sub>-L1. Saturation binding curves of hPin1<sub>WW</sub> ( $K_d = 2.2 \pm 0.6 \mu\text{M}$ ) and CC-hPin1<sub>WW</sub>-L1 ( $K_d = 2.0 \pm 0.7 \mu\text{M}$ ) obtained from tryptophan fluorescence enhancement upon CTD-peptide titration. Conditions:  $c$  (hPin1<sub>WW</sub>/CC-hPin1<sub>WW</sub>-L1) = 2  $\mu\text{M}$ , PBS buffer, pH 7.4, room temperature; titration steps: 0–500  $\mu\text{M}$  CTD peptide. All binding dissociation constants were obtained from non-linear least square fitting to the saturation bindings curves (see ESI Eqn. 14†). (b–e) Changes in CD spectra at 20  $^{\circ}\text{C}$  (b and d) and CD thermal denaturation profiles (c and e) of hPin1<sub>WW</sub> (b and c) and CC-hPin1<sub>WW</sub>-L1 (d and e) in the presence of equimolar amounts of CTD ligand. Thermodynamic data was obtained from non-linear least square fitting to thermal denaturation curves (Table 4). Experimental conditions: PBS buffer, pH 7.4, 40  $\mu\text{M}$  peptide concentration. Thermal denaturation profiles were fitted using a two-state folding-unfolding model (see ESI†).

yield a relatively small library of sixteen split-WW domains at half the synthetic effort, since only eight peptides need to be synthesized rather than sixteen; ten N-terminal fragments combined with ten C-terminal fragments, on the other hand, would expand the library to 100 members, requiring the synthesis of only twenty peptides, a comparatively small synthetic effort.

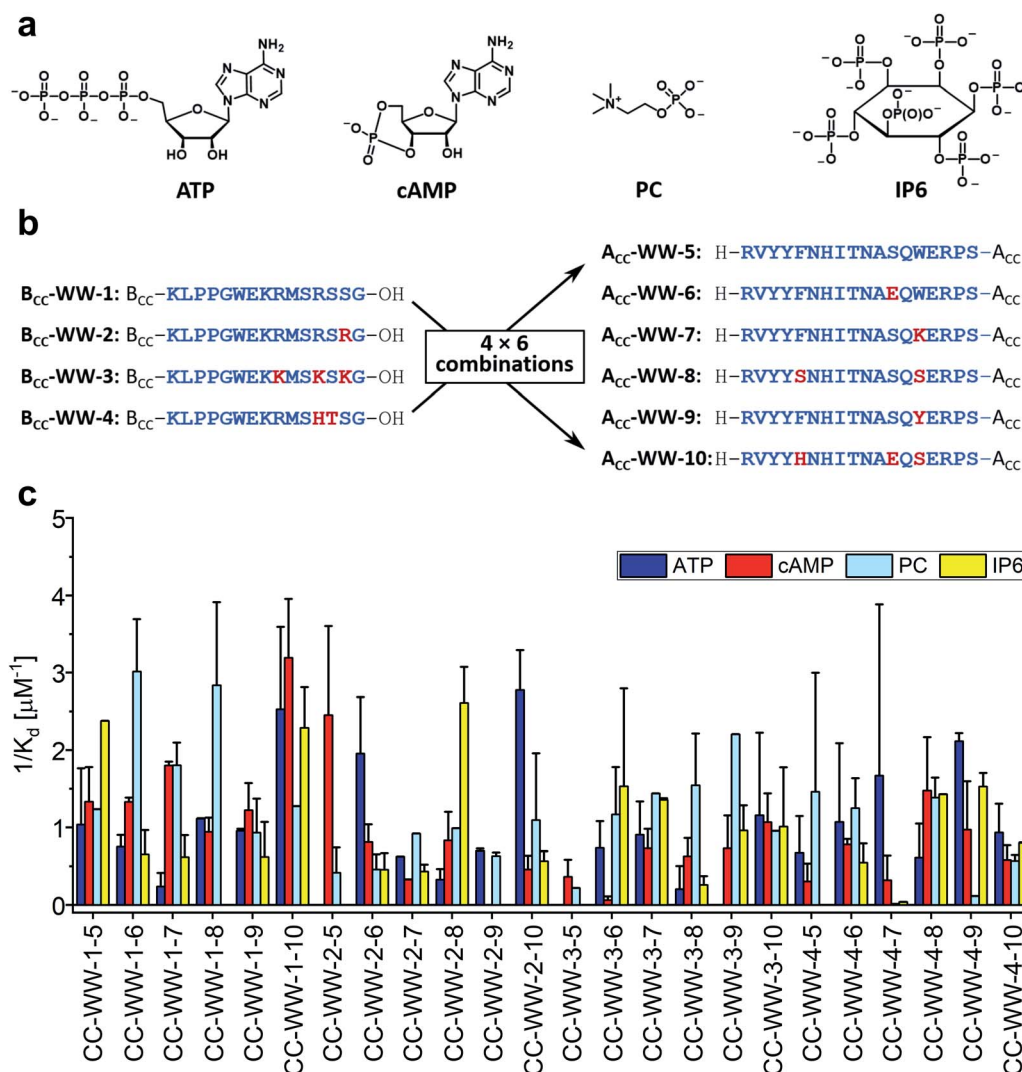
As a proof-of-concept, we chose to synthesize a small split-WW domain library derived from hPin1<sub>WW</sub> to be screened specifically for adenosine triphosphate (ATP) binding. Weak ATP binding with a binding dissociation constant of 0.6 mM was previously demonstrated with hPin1<sub>WW</sub>-M15C modified with a fluorescent stilbazole to which a Zn (II) dipicolylamine



**Table 4** Thermodynamic analysis of CD thermal denaturation profiles of hPin<sub>1WW</sub> and CC-hPin<sub>1WW</sub>-L1 in the absence and presence of ligand CTD

	$T_m$ ( $T_m^K$ ) [ $^{\circ}\text{C}$ ]	$\Delta H_m$ [ $\text{kJ mol}^{-1}$ ]	$\Delta C_{p,m}$ [ $\text{J mol}^{-1} \text{K}^{-1}$ ]	$\Delta G_{\text{Di}}^{T=293\text{K}}$ [ $\text{kJ mol}^{-1}$ ]
hPin <sub>1WW</sub> <sup>a</sup>	57.5 ± 0.9	-120 ± 12	-3300 ± 500	-6.0 ± 1.4
hPin <sub>1WW</sub> + CTD <sup>a</sup>	56.0 ± 0.6	-135 ± 6	-2600 ± 1100	-9.6 ± 0.8
CC-hPin <sub>1WW</sub> -L1 <sup>b</sup>	65.0 ± 0.4 (124.5 ± 0.1)	-168 ± 2	-190 ± 40	-41.3 ± 0.2
CC-hPin <sub>1WW</sub> -L1 + CTD <sup>b</sup>	60.5 ± 0.1 (111.5 ± 0.7)	-187 ± 5	-200 ± 120	-42.3 ± 0.2

<sup>a</sup> Thermodynamic parameters were obtained from non-linear least square fits using a two-state folding/unfolding model for a monomer.<sup>39</sup>  $T_m$  is the midpoint of thermal denaturation at fraction folded  $\alpha = 0.5$ ;  $\Delta H_m$  is the change in enthalpy at  $T_m$ ;  $\Delta C_{p,m}$  is the change in heat capacity at  $T_m$  and  $\Delta G_{\text{Di}}^{T=293\text{K}}$  is the change in free enthalpy at 20  $^{\circ}\text{C}$ . <sup>b</sup> Thermodynamic parameters were obtained from non-linear least square fits using a two-state folding/unfolding model for a dimer.<sup>39</sup>  $T_m^K$  is the concentration independent melting temperature at unfolding constant  $K = 1$ .<sup>40</sup>



**Fig. 5** Split-WW domain library design and screening for binding activity to organophosphates. (a) Organophosphate ligands for library screening: ATP (adenosine triphosphate), cAMP (cyclic adenosine monophosphate), PC (phosphorylcholine), IP6 (inositol hexaphosphate). (b) Sequences of the B<sub>CC</sub>-hPin<sub>1WW</sub>-L1 (B<sub>CC</sub>-WW-1 – B<sub>CC</sub>-WW-4) and hPin<sub>1WW</sub>-L1-A<sub>CC</sub> (A<sub>CC</sub>-WW-5 – A<sub>CC</sub>-WW-10) used in split-WW-domain library design. B<sub>CC</sub>-WW-1 corresponds to parent B<sub>CC</sub>-hPin<sub>1WW</sub>-L1 and A<sub>CC</sub>-WW-5 to hPin<sub>1WW</sub>-L1-A<sub>CC</sub>. Combination of the N- and C-terminal fragments results in a library of 24 split-WW domains. The amino acids highlighted in red mark the mutation sites. (c) Screening readout based on saturation binding monitored by intrinsic tryptophan fluorescence changes in split-WW domains upon ligand binding and quantified by calculation of the association constants ( $1/K_d$ ). Note: For practical reasons, only elementary titrations were performed. Therefore, the association constants obtained in this screening serve only as a relative comparison of the binding of the four ligands to a respective split-WW domain variant (for experimental details see ESI†).



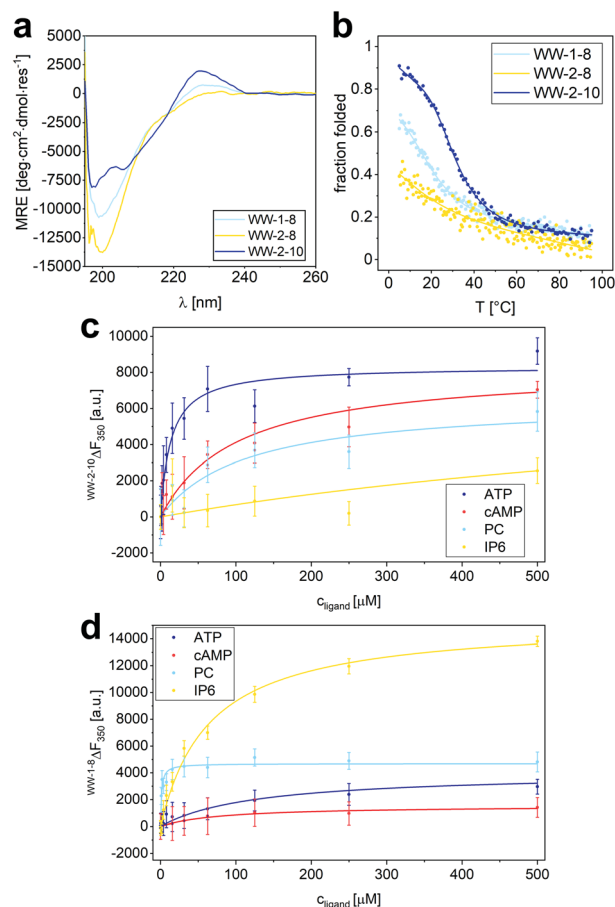
was attached as an additional phosphate binding motif.<sup>34</sup> However, our goal was to find an ATP-binding WW domain that consists exclusively of natural amino acids and, in addition, binds other biologically relevant organophosphates such as cyclic adenosine monophosphate (cAMP), phosphorylcholine (PC), or inositol hexaphosphate (IP6, Fig. 5a) less well, *i.e.*, that is ATP-specific. In addition to the original CC-hPin1<sub>WW</sub>-L1 fragments, in the context of this library now referred to as B<sub>CC</sub>-WW-1 and A<sub>CC</sub>-WW-5, eight other peptides – three N-terminal fragments and five C-terminal fragments that were derived from B<sub>CC</sub>-WW-1 and A<sub>CC</sub>-WW-5, respectively – were synthesized. This resulted in a total of four N-terminal fragments (B<sub>CC</sub>-WW-1 to B<sub>CC</sub>-WW-4) and six C-terminal fragments (B<sub>CC</sub>-WW-5 to B<sub>CC</sub>-WW-10), which after combination yielded a library of 24 split-WW domains (Fig. 5b). The most relevant interactions of ATP-binding proteins with the ligand are hydrogen bonds with the ribose moiety, aromatic interactions with the nucleobase, or electrostatic interactions of the negatively charged triphosphate with positively charged amino acid residues. Therefore, the variants of B<sub>CC</sub>-WW-1 and A<sub>CC</sub>-WW-5 contained mutations with positively charged lysine or arginine residues (B<sub>CC</sub>-WW-2, B<sub>CC</sub>-WW-3, A<sub>CC</sub>-WW-7), polar serine or glutamate residues (A<sub>CC</sub>-WW-6, A<sub>CC</sub>-WW-8, A<sub>CC</sub>-WW-10), and polar aromatic residues such as tyrosine and histidine (B<sub>CC</sub>-WW-4, A<sub>CC</sub>-WW-9, A<sub>CC</sub>-WW-10, Fig. 5b).

Library screening was performed by measuring the changes in intrinsic tryptophan fluorescence upon ligand binding. Since the change in tryptophan fluorescence is highly dependent on the peptide, elementary ligand titrations were performed to provide the binding association constants (Fig. S9–S12†). In a screening format, this methodology represents an estimate, but provides a relative comparison of the binding of the different ligands to one peptide species in a qualitative manner. The results of the library screening are summarized in Fig. 5c. Many of the peptide variants did not exhibit preferential binding to any of the organophosphates tested or showed ambiguous results. Only two combinations – CC-WW-2-6 and CC-WW-2-10 – showed specific binding to ATP over the other organophosphates. Of these two, CC-WW-2-10 was investigated further, because ATP binding appeared to be more selective for this particular split-WW domain. In addition to CC-WW-2-10, we also decided to study CC-WW-1-8 and CC-WW-2-8, as these variants showed selective binding to PC and IP6.

### WW-domain-derived peptides as mini-receptors for organophosphates

Based on the results of the library screening, WW domains WW-2-10, WW-1-8, and WW-2-8 were synthesized to further investigate their selective binding properties to ATP, PC, or IP6. Since these WW domains, unlike the CC-fusion variants, were not stabilized by CC interaction, CD spectroscopy was employed to characterize their structure and thermodynamic stability (Fig. 6a and b). The CD spectrum of WW-2-10 exhibits prominent features of a WW domain by showing the typical maximum at 227 nm resulting from the exciton coupling of the aromatic residues. However, compared to hPin1<sub>WW</sub>, the signal

intensity is lower, mainly due to the exchange of W34 for a serine residue. In the CD spectrum of WW-1-8, the exciton coupling is even less pronounced, and WW-2-8 appears to be mainly a random coil in solution. To investigate the overall thermodynamic stability of these WW domains, thermal denaturation profiles were recorded by monitoring the change in CD signal at a wavelength of 227 nm, which gave a  $T_m$  of 36.5 °C for WW-2-10 and of 14 °C for WW-1-8 (Fig. 6a and b). WW-2-8 did not show a cooperative folding to unfolding transition, allowing us to conclude that it represents an unfolded peptide species at ambient temperature. However, WW-1-8 also exhibits lower



**Fig. 6** Characterization of structure and binding properties of WW-1-8, WW-2-8 and WW-2-10. (a) CD spectrum of WW-1-8, WW-2-8 and WW-2-10 at  $T = 20$  °C; (b) thermal denaturation profile of WW-1-8, WW-2-8 and WW-2-10 depicted as fraction folded. The thermal denaturation profile was fitted to a two-state folding model giving a  $T_m$  value of  $36.5 \pm 0.5$  °C (see ESI†). Experimental conditions: PBS buffer, pH 7.4, 40  $\mu\text{M}$  peptide concentration. (c) Intrinsic tryptophan fluorescence titrations of ATP, cAMP, phosphorylcholine, and IP6 to WW-2-10 lead to the following  $K_d$  values:  $K_d$  (ATP) =  $14 \pm 3.4$   $\mu\text{M}$ ,  $K_d$  (cAMP) =  $102 \pm 22$   $\mu\text{M}$ ,  $K_d$  (PC) =  $112 \pm 51$   $\mu\text{M}$ ,  $K_d$  (IP6) =  $1.1 \pm 0.3$  mM. (d) Intrinsic tryptophan fluorescence titrations of ATP, cAMP, phosphorylcholine, and IP6 to WW-1-8 lead to the following  $K_d$  values:  $K_d$  (ATP) =  $130 \pm 40$   $\mu\text{M}$ ,  $K_d$  (cAMP) =  $81 \pm 28$   $\mu\text{M}$ ,  $K_d$  (PC) =  $1.1 \pm 0.3$   $\mu\text{M}$ ,  $K_d$  (IP6) =  $64 \pm 6$   $\mu\text{M}$ . Experimental conditions: c (peptide) = 2  $\mu\text{M}$ , PBS buffer, pH 7.4, room temperature; titration steps: 0–500  $\mu\text{M}$  phosphate ligand. All  $K_d$  values were obtained from non-linear least square fitting using the saturation bindings curves (see ESI Eqn. 14†).



thermodynamic stability, suggesting that mutations in the C-terminal fragment A<sub>CC</sub>-WW-8 significantly affect the structural integrity of the WW domains. WW-2-10, on the other hand, is a stably folded WW domain possessing a cooperative denaturation profile, although it shows lower thermodynamic stability compared with wild type hPin1<sub>WW</sub>. Taken together, these data clearly demonstrate that the split-WW-domain approach indeed allows screening for function and not primarily for thermodynamic stability.

After investigating the structure and thermodynamic stability, we analyzed the binding of organophosphates to WW-2-10 and WW-1-8 by measuring the changes in intrinsic tryptophan fluorescence. An one-to-one dilution series of each ligand was prepared starting from a concentration of 500 μM, each dilution was equilibrated with a concentration of 2 μM of the respective WW-domain variant at a temperature of 4 °C for a period of 15 h and measured in a microplate reader at room temperature to obtain the intrinsic tryptophan fluorescence at a wavelength of 350 nm.

When analyzing the binding properties of WW-2-10, tryptophan fluorescence quenching was observed upon ATP binding, while binding of the other ligands resulted in weak tryptophan fluorescence enhancement. By fitting the resulting titration curves to an one-site binding model, a  $K_d$  of  $14 \pm 4$  μM was determined for ATP binding. In comparison, the  $K_d$  values of cAMP and PC were increased by an order of magnitude, and IP6 showed only weak binding possessing a  $K_d$  value in the mM range (Fig. 6c). This confirmed the results from the library screening, which indicated a specific binding activity of WW-2-10 to ATP. To our knowledge, only one small peptide folding motif has been reported so far that exhibits specific ATP-binding properties, namely a β-hairpin of 11 amino acids.<sup>43</sup> ATP binding was achieved by electrostatic and aromatic interactions, but is one order of magnitude weaker than found for WW-2-10, which with its 34 amino acids can still be considered a peptide and, in this case, a mini-ATP receptor that binds comparatively strongly to ATP. By comparison, natural ATP-binding proteins bind their ligand in the high nM to mM range.<sup>44</sup>

To further investigate the relationship between sequence and activity, we also relied on computational methods. A structural model of WW-2-10 was constructed with the Rosetta Relax application<sup>41</sup> based on an available NMR structure of hPin1<sub>WW</sub> (pdb code 116c), in which the corresponding amino acid residues were mutated. Subsequently, AutoDock Vina<sup>42</sup> was used to perform molecular docking with the organophosphates considered in this study to predict their binding mode (Fig. S13†). Interestingly, WW-2-10 showed two binding pockets on the surface, a polar pocket (P1) and a positively charged pocket (P2). The key of this structural feature lies in the two mutations F25H and S32E. The two amino acid residues H and E form a hydrogen bond through which H25 moves out of the outer hydrophobic core, constituting a cavity (Fig. S13A and B†). While the long negatively charged phosphate tail of ATP binds to P2, where it is stabilized by electrostatic interactions and hydrogen bonds with the three arginine residues, its nucleoside moiety appears to fit perfectly into P1 and is stabilized mainly by

hydrogen bonds. The other organophosphates that were investigated either do not fit into the cavities or are too small to occupy both pockets, which is for instance the case for phosphorylcholine (Fig. S13†).

Although WW-1-8 exhibits lower thermodynamic stability, we investigated its binding behaviour to the four different organophosphates (Fig. 6d). In all titration experiments, we observed an enhancement of tryptophan fluorescence, and  $K_d$  values were determined by fitting the resulting titration curves to a one-site-binding model. For titration of phosphorylcholine to WW-1-8, we determined a  $K_d$  value of  $1.3 \pm 0.3$  μM, whereas titrations of ATP and cAMP yielded  $K_d$  values that were increased by approximately two orders of magnitude. Interestingly, titration of IP6 to WW-1-8 resulted in a strong tryptophan fluorescence enhancement. Although fitting the titration curve yielded a  $K_d$  value of only 65 μM, a factor of 1.5 increased compared with the  $K_d$  value of PC, we wondered whether this phenomenon was due to the lower thermodynamic stability of WW-1-8. Therefore, we also performed quantitative titrations of the phosphate ligands with the corresponding split variant CC-WW-1-8, which is thermodynamically stable due to the fused CC dimerization domain (Fig. S14†). Interestingly, CC-WW-1-8 showed a slightly different binding behaviour than the non-fusion WW domain WW-1-8. PC was similarly bound with a  $K_d$  value of  $1.7 \pm 0.4$  μM, however, the  $K_d$  values for ATP and cAMP were reduced by one order of magnitude, and IP6 was only weakly bound without causing a strong enhancement of tryptophan fluorescence that was observed upon binding of IP6 to WW-1-8. Thus, we can conclude that the thermodynamic stability of the modified WW domain most likely has a significant effect on ligand binding, although an influence of the fragmentation in loop 1 in CC-WW-1-8 cannot be ruled out completely. However, we state that both WW-1-8 and CC-WW-1-8 show the highest binding affinity for PC, which is consistent with the library screening.

## Conclusions

In this report, we have successfully introduced the concept of CC-mediated proximity-induced reconstitution of fragmented WW domains, which has high potential as a combinatorial approach for identifying WW domains with novel functions as demonstrated here. To develop this approach, we chose wild type hPin1<sub>WW</sub> as model peptide, which inherent structural and functional properties can be well investigated by applying a synergistic combination of CD, fluorescence and high-resolution NMR spectroscopy. Two fragmentation sites have been investigated: fragmentation in loop 1 (CC-hPin1<sub>WW</sub>-L1) at the GR site and fragmentation in loop 2 (CC-hPin1<sub>WW</sub>-L2) at the HI site. By using CD spectroscopy and 1D proton NMR experiments we could show that in CC-hPin1<sub>WW</sub>-L1 the fragmented WW domain is structurally reconstituted, although loop 1 is reported to be the folding nucleation site of hPin1<sub>WW</sub>.<sup>25,26</sup> These results are in accordance to the findings of Kier *et al.* who studied a circularly permuted and hyperstable variant of hPin1<sub>WW</sub>.<sup>20</sup> In contrast, the WW-domain region in CC-hPin1<sub>WW</sub>-L2 is not properly folded, which we attribute to the fact that loop



2 is conformationally well-defined and probably key to the overall stability of the WW-domain structure. A  $\Phi$ -value analysis of hPin1<sub>WW</sub> by Gruebele *et al.* indicated that loop 1 and loop 2 play a similar role in the folding process of hPin1<sub>WW</sub>, anticipating a two-state folding mechanism and not, as previously reported,<sup>25</sup> a sequential folding mechanism with loop 1 as initial nucleation site.<sup>45</sup>

Reconstitution of function was investigated by titration of the CTD peptide of RNA polymerase II, a natural hPin1<sub>WW</sub> ligand, to solutions of CC-hPin1<sub>WW</sub>-L1 and hPin1<sub>WW</sub> and monitored by recording both two-dimensional heteronuclear <sup>1</sup>H-<sup>15</sup>N HSQC NMR spectra of isotopically labelled CC-hPin1<sub>WW</sub>-L1 and intrinsic tryptophan fluorescence spectroscopy, respectively. Furthermore, structural changes upon ligand binding were investigated by CD spectroscopy and the acquisition of thermal denaturation profiles. These experiments show that not only the structure but also the binding properties of hPin1<sub>WW</sub> are fully reconstituted in CC-hPin1<sub>WW</sub>-L1, although loop 1 contains the phosphate binding site and is therefore directly involved in ligand binding. To put it in a nutshell, the split hPin1<sub>WW</sub> described here is the smallest protein/peptide to date that has been shown to reconstitute completely by CC association. This highlights the general robustness of WW-domain structure.

The WW domain has great potential as a scaffold for the development of biomimetic peptides and mini-proteins. The folding motif is extremely robust to modifications and mutations. However, unlike CCs, the sequence-structure relationships are less straightforward and therefore not as well studied. The split-WW domain approach presented here allows us to explore the sequence space of the WW domain in a combinatorial manner, facilitating the identification of novel functional peptides that exhibit WW-domain folding. We demonstrated the potential of this engineering strategy with a 24-member split-WW-domain library built from only ten peptides, from which we identified one specific ATP binder, namely WW-2-10, possessing a remarkable binding dissociation constant in the low micromolar range. This peptide competes very well with other engineered systems<sup>43</sup> and natural ATP-binding proteins, the latter having  $K_d$  values from the high nanomolar to the millimolar regime.<sup>44</sup> Sequence optimization of WW-2-10 could improve ATP binding even further. Only a single artificial ATP-binding protein of 80 amino acids, identified by *in vitro* selection of a messenger RNA displayed protein library with  $6 \times 10^{12}$  members, showed a  $K_d$  value of 0.1  $\mu$ M.<sup>46</sup> This is obviously very impressive, but the advantage of our approach is that the WW domain is comparatively small and therefore the peptides can be chemically synthesized in an automated manner, allowing library members with artificial amino acid residues. In addition to the ATP binder, we have also identified a phosphorylcholine-selective WW domain, WW-1-8, which, although not thermodynamically very stable, binds to PC with a  $K_d$  value of about 1  $\mu$ M and could be developed into a mini-receptor for PC or PC derivatives such as lysophosphatidylcholine, which is found, for example, as a so-called “Find-Me signal” on apoptotic cells.<sup>47</sup> We are convinced that the results reported here are promising for future applications, such as the development of other mini-

phosphate receptors or mini-phosphatases. Thus, we intend to establish this split-WW-domain approach as a platform for the development and identification of a variety of functional WW domains in future.

## Data availability

All experimental data associated with this work are available in the ESI.†

## Author contributions

FT and MK designed the project. HN and NBP synthesised the CC-hPin1<sub>WW</sub> constructs and investigated the reconstitution of split-hPin1<sub>WW</sub> by coiled-coil association using CD spectroscopy. FH synthesised WW-1-8, WW-2-8 and WW-2-10 and performed CD spectroscopic analysis. CL synthesised the split-WW-domain library and performed the library screening. FH and RG performed the computational modelling and docking of WW-2-10. MK performed all NMR experiments and analysed the NMR data. FT investigated the binding of CC-hPin1<sub>WW</sub>-L1, hPin1<sub>WW</sub>, WW-1-8 and WW-2-10 to their respective ligands using fluorescence spectroscopy and CD spectroscopy. She also designed the split-WW-domain library and screening experiment. The data were analysed and discussed by all authors. FT and MK wrote the manuscript.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The project was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 414261058 and supported by the DFG under Germany's Excellence Strategy 2082/1 390761711. MK acknowledges financial support from the Young Scholar Fund and the permanent financial support provided by the University of Konstanz enabling to operate a state-of-the-art NMR core facility. FH thanks the Carl Zeiss Foundation for the financial support. We would also like to thank the Dr Sophie-Berntsen-Stiftung for financial support in building our computational modeling infrastructure. We are grateful to Prof. Dr Michael Mastalerz (Heidelberg University) for providing access to a CD spectrometer. We thank the Centre for Advanced Materials (Heidelberg University) and Ulf Diederichsen (University of Göttingen) for general access to technical equipment.

## Notes and references

- (a) K. J. Grayson and J. L. R. Anderson, *J. R. Soc., Interface*, 2018, **15**, 20180472; (b) L. R. Marshall, O. Zozulia, Z. Lengyel-Zhand and I. V. Korendovych, *ACS Catal.*, 2019, **9**, 9265–9275.
- W. M. Dawson, G. G. Rhys and D. N. Woolfson, *Curr. Opin. Chem. Biol.*, 2019, **52**, 102–111.



- 3 D. N. Woolfson, *Fibrous Proteins*, 2017, **82**, 35–61.
- 4 (a) I. Nakase, S. Okumura, G. Tanaka, K. Osaki, M. Imanishi and S. Futaki, *Angew. Chem., Int. Ed.*, 2012, **51**, 7464–7467; (b) F. Thomas, W. M. Dawson, E. J. M. Lang, A. J. Burton, G. J. Bartlett, G. G. Rhys, A. J. Mulholland and D. N. Woolfson, *ACS Synth. Biol.*, 2018, **7**, 1808–1816.
- 5 W. M. Rink and F. Thomas, *Chem. – Eur. J.*, 2019, **25**, 1665–1677.
- 6 (a) K. E. Thompson, C. J. Bashor, W. A. Lim and A. E. Keating, *ACS Synth. Biol.*, 2012, **1**, 118–129; (b) A. J. Smith, F. Thomas, D. Shoemark, D. N. Woolfson and N. J. Savery, *ACS Synth. Biol.*, 2019, **8**, 1284–1293; (c) C. L. Edgell, A. J. Smith, J. L. Beesley, N. J. Savery and D. N. Woolfson, *ACS Synth. Biol.*, 2020, **9**, 427–436; (d) T. Lebar, D. Lainšček, E. Merljak, J. Aupič and R. Jerala, *Nat. Chem. Biol.*, 2020, **16**, 513–519.
- 7 (a) H. Robson Marsden, N. A. Elbers, P. H. H. Bomans, N. A. J. M. Sommerdijk and A. Kros, *Angew. Chem. Int. Ed.*, 2009, **48**, 2330–2333; (b) K. Meyenberg, A. S. Lygina, G. van den Bogaart, R. Jahn and U. Diederichsen, *Chem. Commun.*, 2011, **47**, 9405.
- 8 (a) B. Ma and R. Nussinov, *Curr. Opin. Chem. Biol.*, 2010, **14**, 652–659; (b) C. Micheletti, *Phys. Life Rev.*, 2013, **10**, 1–26.
- 9 (a) K. M. Wilcoxon, L. J. Lemman, D. A. Weinberger, Z.-Z. Huang and M. R. Ghadiri, *J. Am. Chem. Soc.*, 2007, **129**, 748–749; (b) M. C. Groth, W. M. Rink, N. F. Meyer and F. Thomas, *Chem. Sci.*, 2018, **9**, 4308–4316.
- 10 (a) Z. Salah, A. Alian and R. I. Aqeilan, *Front. Biosci.*, 2012, **17**, 331–348; (b) J. L. Ilsley, M. Sudol and S. J. Winder, *Cell. Signalling*, 2002, **14**, 183–189.
- 11 (a) M. J. Macias, M. Hyvönen, E. Baraldi, J. Schultz, M. Sudol, M. Saraste and H. Oschkinat, *Nature*, 1996, **382**, 646–649; (b) J. R. Pires, F. Taha-Nejad, F. Toepert, T. Ast, U. Hoffmüller, J. Schneider-Mergener, R. Kühne, M. J. Macias and H. Oschkinat, *J. Mol. Biol.*, 2001, **314**, 1147–1156; (c) M. Meiyappan, G. Birrane and J. A. A. Ladas, *J. Mol. Biol.*, 2007, **372**, 970–980.
- 12 R. Ranganathan, K. P. Lu, T. Hunter and J. P. Noel, *Cell*, 1997, **89**, 875–886.
- 13 P. Tian and R. B. Best, *Biophys. J.*, 2017, **113**, 1719–1730.
- 14 (a) M. J. Macias, V. Gervais, C. Civera and H. Oschkinat, *Nat. Struct. Biol.*, 2000, **7**, 375–379; (b) M. Socolich, S. W. Lockless, W. P. Russ, H. Lee, K. H. Gardner and R. Ranganathan, *Nature*, 2005, **437**, 512–518; (c) W. P. Russ, D. M. Lowery, P. Mishra, M. B. Yaffe and R. Ranganathan, *Nature*, 2005, **437**, 579–583; (d) J. M. Anderson and N. H. Andersen, *Angew. Chem., Int. Ed.*, 2017, **56**, 7074–7077.
- 15 A. L. Stewart, J. H. Park and M. L. Waters, *Biochemistry*, 2011, **50**, 2575–2584.
- 16 (a) T. L. Pham, M. Kovermann and F. Thomas, *ACS Synth. Biol.*, 2022, **11**, 254–264; (b) M. Werner, J. Pampel, T. L. Pham and F. Thomas, *Chem.–Eur. J.*, 2022, e202201339.
- 17 (a) S. S. Shekhawat and I. Ghosh, *Curr. Opin. Chem. Biol.*, 2011, **15**, 789–797; (b) E. Michel, A. Plückthun and O. Zerbe, *Angew. Chem., Int. Ed.*, 2018, **57**, 4576–4579.
- 18 (a) I. Ghosh, A. D. Hamilton and L. Regan, *J. Am. Chem. Soc.*, 2000, **122**, 5658–5659; (b) S. S. Shekhawat, J. R. Porter, A. Sriprasad and I. Ghosh, *J. Am. Chem. Soc.*, 2009, **131**, 15284–15290; (c) T. Azad, A. Tashakor and S. Hosseinkhani, *Anal. Bioanal. Chem.*, 2014, **406**, 5541–5560.
- 19 T. Fink, J. Lonžarić, A. Praznik, T. Plaper, E. Merljak, K. Leben, N. Jerala, T. Lebar, Ž. Strmšek, F. Lapenta, et al., *Nat. Chem. Biol.*, 2019, **15**, 115–122.
- 20 B. L. Kier, J. M. Anderson and N. H. Andersen, *J. Am. Chem. Soc.*, 2014, **136**, 741–749.
- 21 (a) R. Wintjens, J. M. Wieruszkeski, H. Drobecq, P. Rousselot-Pailley, L. Buée, G. Lippens and I. Landrieu, *J. Biol. Chem.*, 2001, **276**, 25150–25156; (b) J. A. Kowalski, K. Liu and J. W. Kelly, *Biopolymers*, 2002, **63**, 111–121; (c) C. Schelhorn, P. Martín-Malpartida, D. Sunöl and M. J. Macias, *Sci. Rep.*, 2015, **5**, 1–12; (d) D. E. Mortenson, D. F. Kreitler, H. G. Yun, S. H. Gellman and K. T. Forest, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2013, **69**, 2506–2512.
- 22 L. M. Luh, R. Hänsel, F. Löhr, D. K. Kirchner, K. Krauskopf, S. Pitzius, B. Schäfer, P. Tufar, I. Corbeski, P. Güntert, et al., *J. Am. Chem. Soc.*, 2013, **135**, 13796–13803.
- 23 M. Jäger, M. Dendle and J. W. Kelly, *Protein Sci.*, 2009, **18**, 1806–1813.
- 24 (a) W. Chen, S. Enck, J. L. Price, D. L. Powers, E. T. Powers, C. H. Wong, H. J. Dyson and J. W. Kelly, *J. Am. Chem. Soc.*, 2013, **135**, 9877–9884; (b) D. E. Mortenson, D. F. Kreitler, N. C. Thomas, I. A. Guzei, S. H. Gellman and K. T. Forest, *ChemBioChem*, 2018, **19**, 604–612; (c) A. A. Fuller, D. Du, F. Liu, J. E. Davoren, G. Bhabha, G. Kroon, D. A. Case, H. J. Dyson, E. T. Powers, P. Wipf, et al., *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 11067–11072.
- 25 D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan, et al., *Science*, 2010, **330**, 341–346.
- 26 (a) M. Jäger, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele and J. W. Kelly, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10648–10653; (b) F. Liu, D. Du, A. A. Fuller, J. E. Davoren, P. Wipf, J. W. Kelly and M. Gruebele, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 2369–2374.
- 27 D. L. McClain, H. L. Woods and M. G. Oakley, *J. Am. Chem. Soc.*, 2001, **123**, 3151–3152.
- 28 (a) I. Coin, *J. Pept. Sci.*, 2010, **16**, 223–230; (b) I. Coin, M. Beyermann and M. Bienert, *Nat. Protoc.*, 2007, **2**, 3247–3256.
- 29 (a) L. Whitmore and B. A. Wallace, *Nucleic Acids Res.*, 2004, **32**, 668–673; (b) L. Whitmore and B. A. Wallace, *Biopolymers*, 2008, **89**, 392–400.
- 30 D. K. Wilkins, S. B. Grimshaw, V. Receveur, C. M. Dobson, J. A. Jones and L. J. Smith, *Biochemistry*, 1999, **38**, 16424–16431.
- 31 K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, 1986.
- 32 K. P. Lu, G. Finn, T. H. Lee and L. K. Nicholson, *Nat. Chem. Biol.*, 2007, **3**, 619–629.
- 33 (a) Y. X. Xu, Y. Hirose, X. Z. Zhou, K. P. Lu and J. L. Manley, *Genes Dev.*, 2003, **17**, 2765–2776; (b) J. P. Hsin and J. L. Manley, *Genes Dev.*, 2012, **26**, 2119–2137.



- 34 T. Anai, E. Nakata, Y. Koshi, A. Ojida and I. Hamachi, *J. Am. Chem. Soc.*, 2007, **129**, 6233–6239.
- 35 M. A. Verdecia, M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel, *Nat. Struct. Biol.*, 2000, **7**, 639–643.
- 36 A. Yammine, J. Gao and A. H. Kwan, *Bio-Protoc.*, 2019, **9**, e3253.
- 37 C. M. S. Carter, J. R. Leighton-Davies and S. J. Charlton, *J. Biomol. Screening*, 2007, **12**, 255–266.
- 38 J. K. Myers, D. P. Morris, A. L. Greenleaf and T. G. Oas, *Biochemistry*, 2001, **40**, 8479–8486.
- 39 N. J. Greenfield, *Methods Enzymol.*, 2004, **383**, 282–317.
- 40 N. J. Greenfield, *Nat. Protoc.*, 2006, **1**, 2527–2535.
- 41 (a) F. Khatib, S. Cooper, M. D. Tyka, K. Xu, I. Makedon, Z. Popović, D. Baker and F. Players, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 18949–18953; (b) J. B. Maguire, H. K. Haddox, D. Strickland, S. F. Halabiya, B. Coventry, J. R. Griffin, S. V. S. R. K. Pulavarti, M. Cummins, D. F. Thieker, E. Klavins, T. Szyperski, F. DiMaio, D. Baker and B. Kuhlman, *Proteins: Struct., Funct., Bioinf.*, 2021, **89**, 436–449.
- 42 S. Forli, R. Huey, M. E. Pique, M. F. Sanner, D. S. Goodsell and A. J. Olson, *Nat. Protoc.*, 2016, **11**, 905–919.
- 43 S. M. Butterfield and M. L. Waters, *J. Am. Chem. Soc.*, 2003, **125**, 9580–9581.
- 44 D. J. Woodbury, E. C. Whitt and R. E. Coffman, *Biophys. Rep.*, 2021, **1**, 100012.
- 45 K. Dave, M. Jäger, H. Nguyen, J. W. Kelly and M. Gruebele, *J. Mol. Biol.*, 2016, **428**, 1617–1636.
- 46 A. D. Keefe and J. W. Szostak, *Nature*, 2001, **410**, 715–718.
- 47 T. O. J. Cockram, J. M. Dundee, A. S. Popescu and G. C. Brown, *Front. Immunol.*, 2021, **12**, 1–33.

