

Cite this: *Chem. Sci.*, 2020, 11, 10390

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

## Proteomimetic surface fragments distinguish targets by function†

Attila Tököli,<sup>a</sup> Beáta Mag,<sup>a</sup> Éva Bartus,<sup>ab</sup> Edit Wéber,<sup>a</sup> Gerda Szakonyi,<sup>c</sup> Márton A. Simon,<sup>d</sup> Ágnes Czibula,<sup>e</sup> Éva Monostori,<sup>e</sup> László Nyitray,<sup>id</sup>\*<sup>d</sup> and Tamás A. Martinek,<sup>id</sup>\*<sup>ab</sup>

The fragment-centric design promises a means to develop complex xenobiotic protein surface mimetics, but it is challenging to find locally biomimetic structures. To address this issue, foldameric local surface mimetic (LSM) libraries were constructed. Protein affinity patterns, ligand promiscuity and protein druggability were evaluated using pull-down data for targets with various interaction tendencies and levels of homology. LSM probes based on H14 helices exhibited sufficient binding affinities for the detection of both orthosteric and non-orthosteric spots, and overall binding tendencies correlated with the magnitude of the target interactome. Binding was driven by two proteinogenic side chains and LSM probes could distinguish structurally similar proteins with different functions, indicating limited promiscuity. Binding patterns displayed similar side chain enrichment values to those for native protein–protein interfaces implying locally biomimetic behavior. These analyses suggest that in a fragment-centric approach foldameric LSMs can serve as useful probes and building blocks for undruggable protein interfaces.

Received 25th June 2020

Accepted 9th September 2020

DOI: 10.1039/d0sc03525d

rsc.li/chemical-science

## Introduction

Protein surface mimetics are chemical tools that target protein–protein interactions (PPIs) that are undruggable with small molecules.<sup>1–3</sup> Although pioneering studies demonstrated the possibility of ribosome-assisted coupling of amino acids with non-natural backbones,<sup>4–7</sup> *in vitro* directed evolutionary approaches<sup>8</sup> are not routinely available for searching and optimization of fundamentally xenobiotic surface mimetic structures.<sup>9–11</sup> Four major experimental chemical approaches and their combinations have been applied to address this problem: (i) screening of large surface mimetic libraries,<sup>10,12,13</sup> (ii) top-down mutational design based on known ligands,<sup>14–21</sup> (iii) bottom-up design and assembly starting from structural hypotheses,<sup>3,22–31</sup> and (iv) the fragment-centric system chemistry approach leading to self-assembling ligands.<sup>32</sup> Recent

theoretical and experimental results strongly support that fragment-centric design built on recognition elements of reduced structural complexity is highly promising for the construction of surface mimetic ligands.<sup>33–35</sup> It is also in accord with the fragment-centric approach that protein interfaces are highly degenerate and strongly interconnected, and interface geometries are not dominated locally by specific secondary structure types.<sup>36,37</sup> To develop the fragment-centric concept further, the present study aimed to find minimal motifs that can serve as local surface mimetics (LSM) with sufficient affinity to probe otherwise undruggable protein surfaces. Dissociation constants below 100  $\mu\text{M}$ , the fragment-based drug design limit,<sup>38</sup> would facilitate observation of protein-specific binding patterns and their potential application as surface mimetic building blocks.

Our hypothesis was that the shortest foldamer sequences with folding tendency in water can be used as LSM probes (Fig. 1). The shortest water-stable foldameric structures are  $\beta$ -peptidic hexamer helices stabilized by a number of cyclic side chains.<sup>39–41</sup> These sequences display only two proteinogenic side chains on the same side that can be strongly shielded from the solvent by the bulky scaffold and the cyclic side chains. While longer foldameric sequences have been experimentally shown to recognize protein surfaces and inhibit protein–protein interactions (PPIs) for targets such as hDM2,<sup>42,43</sup> the Bcl-xL family,<sup>44</sup> gp-41,<sup>45,46</sup> the VEGF-receptor,<sup>47</sup> and  $\beta$ -amyloid oligomers,<sup>22</sup> the general druggability of protein interfaces with foldameric secondary structures has not been established. This

<sup>a</sup>Department of Medical Chemistry, University of Szeged, Dóm tér 8, H6720, Szeged, Hungary. E-mail: martinek.tamas@med.u-szeged.hu

<sup>b</sup>MTA-SZTE Biomimetic Systems Research Group, University of Szeged, Dóm tér 8, H6720, Szeged, Hungary

<sup>c</sup>Institute of Pharmaceutical Analysis, University of Szeged, Somogyi u. 4., H6720, Szeged, Hungary

<sup>d</sup>Department of Biochemistry, Eötvös Loránd University, Pázmány Péter sétány 1/C, H1077, Budapest, Hungary. E-mail: nyitray@elte.hu

<sup>e</sup>Lymphocyte Signal Transduction Laboratory, Institute of Genetics, Biological Research Centre, Temesvári krt. 62, H6726 Szeged, Hungary

† Electronic supplementary information (ESI) available: Experimental procedures, validation of foldamer–protein interactions, and peptide characterisation data. See DOI: 10.1039/d0sc03525d





Fig. 1 Design of local surface mimetic probe libraries using short foldamers. H14 scaffold (a), H12 scaffold (b), library composition showing sublibrary grouping (c), and structures of H14 and H12 helices (d).

might cast doubt on the general compatibility with protein surfaces that unnatural backbone segments are often not tolerated at protein–foldamer interfaces for ligands derived from natural templates in a top-down design scenario.<sup>48,49</sup> In contrast, compatibility was displayed toward some bottom-up designed foldameric sequences,<sup>50</sup> which led to high-affinity ligands.

Besides sufficient affinity (druggability), ligand promiscuity is an equally important problem for the design of specific surface mimetic ligands.<sup>51</sup> It is especially challenging to handle protein interface similarity and ligand promiscuity when exposed protein surfaces are targeted. The relevance of this point for foldamers is indicated by the fact that the chemical space of non-natural residues in bioactive sequences is focused on aliphatic and aromatic hydrophobic side chains that are primarily responsible for ligand promiscuity and non-specific binding.<sup>52,53</sup> Literature examples exemplify the existence of this phenomenon for peptidic<sup>54</sup> and aromatic oligoamide foldamers.<sup>25,55</sup> For acceptable LSM probes, druggability and promiscuity properties should be in balance, mimicking the behaviour of natural proteins. The number of known interactions, and therefore, current knowledge, is limited for druggability–promiscuity properties of foldamers. This situation calls for a systematic approach that generates data on protein–foldamer interactions and facilitates an understanding of molecular recognition patterns depending on the side chain composition, secondary structures of foldamer segments, and target protein surfaces.

In this work, the druggability–promiscuity problem for foldameric LSMs is addressed with a bottom-up fragment-centric pull-down methodology. To test these parameters, target proteins that display different levels of interface similarity and promiscuity in their own PPI network were selected. After mapping LSM binding fingerprints, we found that proposed probes displayed biomimetic features on the protein test set, suggesting the suitability of short foldameric sequences to probe protein surfaces and serve as surface mimetic building blocks.

## Results

### Foldameric LSM probe libraries

Peptide sequences constructed from β-amino acid residues exhibit a strong tendency to fold into secondary structures at short chain lengths, and the specific geometry is controlled by the backbone stereochemistry pattern and topology of structure-inducing side chains.<sup>41,56</sup> Here, helices H14 and H12 were utilized. The closely packed side chain distances in α-helices ( $C_i^\alpha-C_{i+3}^\alpha \approx 5 \text{ \AA}$  and  $C_i^\alpha-C_{i+4}^\alpha \approx 6 \text{ \AA}$ ) and in β-sheets ( $C_i^\alpha-C_{i+2}^\alpha \approx 6 \text{ \AA}$  and “sideways”  $C_i^\alpha-C_{i+n}^\alpha \approx 5 \text{ \AA}$ ) are comparable with the closest side chain distances in foldameric helices H14 ( $C_i^\beta-C_{i+3}^\beta \approx 5 \text{ \AA}$ ) and H12 ( $C_i^\beta-C_{i+2}^\beta \approx 6 \text{ \AA}$ ). Helix H14 exhibits a strong tendency to fold and its side chains are oriented in parallel. Helix H12 is conformationally less stable in aqueous medium, demonstrating an elongated geometry, and side chains are juxtaposed with a small tilt angle.<sup>41</sup>

These secondary structures can be stabilized by cyclic side chain residues so that they display folding at lengths of hexamers and octamers for H14 ((1*S*,2*S*)-2-aminocyclohexanecarboxylic acid, ACHC) and H12 ((1*S*,2*S*)-2-aminocyclopentanecarboxylic acid, ACPC), respectively. Such oligomers allow two proteingenic side chains that point to the identical face of the helix to be incorporated into the sequence (Fig. 1). Sixteen different β<sup>3</sup>-amino acids were substituted at positions R<sup>1</sup> and R<sup>2</sup> generating altogether 512 local surface mimetics. LSM probes were synthesized and screened as sublibraries (L1–L4) (Fig. 1c).

This library includes a considerable fraction of peptide foldamers that fold at such a short chain length. These peptides can be considered as minimally folded building blocks in the peptide foldamer space.<sup>32</sup> These structural units can display a surface area of *ca.* 500 Å<sup>2</sup> toward a protein surface, which can provide the binding free energy for affinities in the range of 1–500 μM. Such affinity is sufficient for probing the surface and is useful for subsequent fragment-based design.

### Test protein set

Five proteins (Table 1) were selected based on numbers of their interacting protein partners (Fig. 2), levels of interprotein structural homology, and types of interactions (protein–protein, protein–carbohydrate). The reference protein in this set was CaM,<sup>57</sup> a highly pleiotropic molecule that plays a vital role in every eukaryotic cell. It displays EF-hand motifs and mediates an extremely large number of signal pathways *via* control of enzyme and ion channel activities in a Ca<sup>2+</sup>-dependent manner. Two additional EF-hand proteins with high structural homology were selected from the S100 family: S100A4 and S100B,<sup>58</sup> both



Table 1 Characteristics of applied protein probes

	CaM	S100A4	S100B	Gal-1	RecQ-WH
Helix content (%)	62	55	58	0	50
$\beta$ -sheet content (%)	4	3	2	54	13
Number of PPIs <sup>a</sup>	645	67	41	5	1
Molecular weight (kDa)	16.8	$11.7 \times 2^b$	$10.7 \times 2^b$	$14.6 \times 2^b$	12.9
Isoelectric point <sup>c</sup>	4.09	5.85	4.52	5.30	10.18

<sup>a</sup> Average values from databases, BioGRID, Wiki-Pi, GPS-Prot, IntAct, and APID. For Gal-1,  $N_{PPI}$  is given based on the review of Camby *et al.*<sup>66</sup>

<sup>b</sup> Homodimers. <sup>c</sup> Calculated values based on amino acid composition.

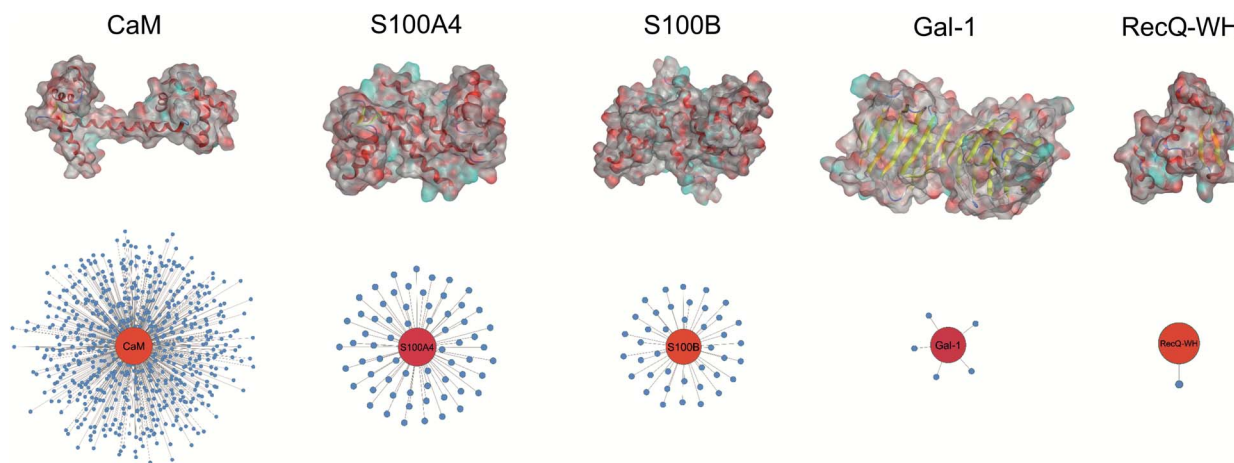


Fig. 2 Surface representation and PPI network of selected proteins. PPI data were obtained from BioGRID,<sup>61</sup> Wiki-Pi,<sup>62</sup> GPS-Prot,<sup>63</sup> IntAct<sup>64</sup> and APID.<sup>65</sup> For the lectin Gal-1,  $N_{PPI}$  is based on ref. 66.

important pharmacological targets for cancer treatment.<sup>59</sup> S100A4 is involved in tumour progression and metastasis and has many interacting partners. S100B is expressed in melanocyte-derived tumours and mature astrocytes,<sup>60</sup> where it plays an important role in neurite extension. Numbers of its known protein partners are significantly less than that for S100A4. Galectin-1 (Gal-1) and the winged helix domain of RecQ helicase (RecQ-WH) were used as target models with low tendency to form PPIs. Gal-1 has an immunosuppressive effect, promoting cancer progression and metastasis through recognition of  $\beta$ -galactoside motifs on cell surface glycoproteins.<sup>66</sup> Gal-1 has a  $\beta$ -sandwich structure with a jelly roll topology,<sup>67–69</sup> that may interact with proteins in a carbohydrate independent manner.<sup>66,70–72</sup> RecQ is a prokaryotic intracellular ATP-dependent DNA helicase, which plays important roles in DNA damage response, DNA recombination, replication, and repair.<sup>73,74</sup> RecQ interacts with the disordered C-terminal of the single-stranded DNA-binding protein (SSB) through its WH domain, which is the single PPI reported for RecQ-WH.<sup>75</sup> All selected targets exhibit solvent exposed shallow binding sites; however, CaM is able to wrap around  $\alpha$ -helices with special side chain motifs, thereby forming a hydrophobic pocket.<sup>76,77</sup> For competitive binding studies, peptide motifs of PPI partners were selected from the following proteins: transient receptor potential cation channel subfamily V member 1 (TRPV1), non-muscle myosin IIA (NMIIA), ribosomal S6 kinase 1 (RSK1), and SSB, for CaM, S100A4, S100B, and RecQ-WH, respectively.

### LSM affinity patterns indicate secondary structure dependent structural compatibility

Pull-down assays with immobilized test proteins and foldameric LSM libraries were carried out. Proteins were equilibrated with LSM probes, and equilibrium bound fractions ( $F_B$ ) were determined from the unbound fractions measured by HPLC-MS methods. Hits were corrected with background binding of the resin. The protocol allowed all binding events to equilibrate irrespective of affinity, specificity, and binding kinetics. This was essential to judge overall druggability and promiscuity. Importantly, the concentration of protein and the total concentration of foldameric LSMs were equimolar, which minimized competition between the probes. This experimental setup approximated independent binding of individual library members. Results were visualized as heat maps of dissociation constants (Fig. 3). Considering the experimental error of  $F_B$  (see ESI, Fig. S1†), the affinity limit was set at 150  $\mu$ M to filter successful hits (Fig. 3, indicated in red), which is 50  $\mu$ M above the 100  $\mu$ M guideline normally applied in fragment-based drug design.<sup>38</sup>

A number of hits were obtained for all proteins studied, which supports the hypothesis that foldameric LSMs can display sufficient contact area toward target proteins (Fig. 4). Helical geometry had profound effects on binding tendency. Helix H12 yielded fewer hits in general than did H14. Moreover, binders having the H12 skeleton in the affinity range 1–10  $\mu$ M





Fig. 3 Binding patterns of foldameric LSM probes represented as heat maps in the  $K_D$  dimension (in  $\mu\text{M}$ ). Apparent dissociation constants are given for H14- and H12-helical LSM libraries. One letter codes displayed at the headings of the rows and columns correspond to the proteinogenic side chains of the  $\beta^3$ -homo-amino acids used in the  $R^1$  and  $R^2$  positions. (Exact  $K_D$  values are given in the ESI, Tables S1–S5.†)

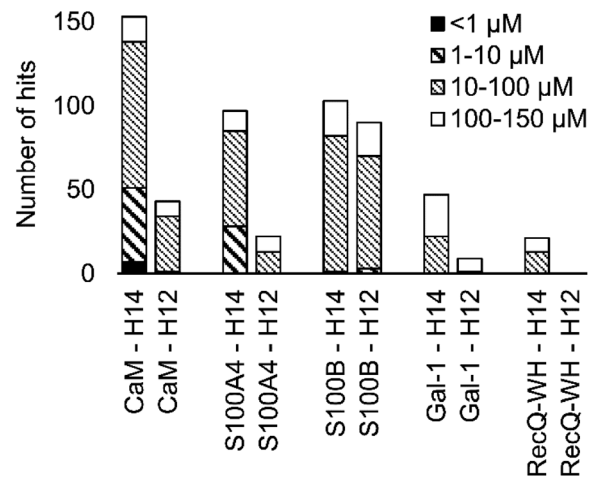


Fig. 4 Number of hits obtained in the H14 and H12 LSM pull-down experiments and their affinity distributions.

could only be observed for protein S100B, while H14 displayed more uniform affinity distribution.

Overall, these findings indicated that H14 helical probes had an overall ability to adapt to surface features of the protein set studied. The stable, short, and bulky structure of H14 with its closely packed side chains is free to reorient to attain the optimum affinity as a local probe. Our conclusion was that the H14 library exhibits superior properties for probing protein surfaces. To validate protein binding of this library, selected H14 LSM probes were synthesized separately and tested for binding in the solution phase (see ESI for details, Fig. S2–6†).

#### Foldameric LSMs detect orthosteric and non-orthosteric spots

Interactions between the H14 helical LSMs and the targets were abundant, but safe recognition of the native PPI interfaces on the targets is a criterion of druggability. To gain insight into the ability of the LSMs to detect orthosteric spots, we carried out pull-down assays in the presence of the native ligands (Table 2) of the targets (Fig. 5). For highlighting the effect of the competitor on the binding fingerprints,  $K_D$  ratios and the LSM replacement percentages were calculated and are represented as heat maps (see ESI, Fig. S7†). We found that native ligands radically changed the LSM affinity patterns for targets exhibiting many direct protein–protein contacts (CaM, S100A4, S100B, and RecQ-WH), and many of the LSM probes were displaced. Since these proteins exhibit geometrically a single surface region to form PPI interfaces, residual binders obviously interacted with the non-orthosteric spots. In contrast, only a few LSM probes displayed replacement for Gal-1. This could be explained by the glycan selective native recognition domain of the protein. We note that the replacement pattern for a single probe can, in theory, be complex. While orthosteric weak-binder probes could be completely replaced from interaction sites, high-affinity foldamer hits might bind to targets even in the presence of competitors, although significantly increased apparent  $K_D$  values clearly indicated orthosteric foldamer probes. Moreover, a single probe can bind to both orthosteric and non-orthosteric



Table 2 Sequences and dissociation constants of the native ligands applied as competitors

Protein	Competitor	$K_D$
CaM	TRPV1-Ct <sub>15</sub>	30.9 ± 2.1 nM (ref. 32)
S100A4	NMIIA (1893–1923)	7 ± 1 nM (ref. 83)
S100B	RSK1 (689–735)	1.8 μM (ref. 83)
Gal-1	Lactose	409 μM (ref. 67)
RecQ-WH	SSB-Ct <sub>8</sub>	16.6 ± 0.8 μM (ref. 75)

spots, which limits overall displacement levels. To avoid overlooking interesting orthosteric weak binders, direct replacement levels were assessed (see ESI, Fig. S7†).

### Affinity patterns of foldameric LSMs are characteristic of the target proteins

An important aspect of this study was to test if foldameric LSM probes are able to distinguish proteins in terms of their PPI forming behaviour. First, different levels of PPI promiscuities were compared among test proteins using the overall ability to bind LSM probes. PPI promiscuity was measured with the number of known PPI partners in databases, BioGRID,<sup>61</sup> Wiki-Pi,<sup>62</sup> GPS-Prot,<sup>63</sup> IntAct<sup>64</sup> and APID<sup>65</sup> (Fig. 2). False positive rates in proteomics range from 20–34% (ref. 78 and 79) and databases may not provide exact numbers of PPIs.<sup>80,81</sup> Nevertheless, these data can be used to indicate the magnitude of promiscuity and protein moonlighting.<sup>82</sup> The binding promiscuity in foldameric LSM space was estimated with the average bound fractions in the H14 library. A connection was found between the number of database PPI partners ( $N_{\text{PPI}}$ ) and the average bound fraction values obtained for the H14 helical LSM probe library (Fig. 6, and ESI, Fig. S8†). Logarithmic scaling is explained by exponential relationships between the fragment space of LSM probes and the chemical space of full proteins. This finding suggests that the overall affinity of the H14 LSM probes distinguishes among proteins with strongly different levels of promiscuity. With this behaviour, foldameric LSM probes displayed a biomimicking feature, an essential criterion for surface mimetic drug design.

The ability to distinguish between proteins can also be measured by similarity between the affinity patterns. Such similarity was calculated using pairwise covariance of  $F_B$  values scaled to the theoretical maximum value for the 256-membered library. Scaled covariance indicates maximum similarity with a value of 100%, while zero covariance leads to 0%.

Results of the calculations are displayed for total (Fig. 7a) and the orthosteric (Fig. 7b) binding fractions. Pairwise scaled covariances were uniformly lower for the orthosteric binders, showing that non-orthosteric binding obscured differences between proteins as detected by the affinity pattern of the LSM probes. This finding supports that the non-orthosteric binding for these proteins can be considered non-specific in nature and is a source of promiscuous behaviour for LSM probes. However, non-orthosteric spots can contribute to allosteric sites, which can be detected with the LSM probes. In this study, CaM, S100A4, and S100B were selected as proteins with high levels of structural and functional homology, albeit their specific roles



Fig. 5 Competition pull-down experiments with the H14 LSM library.  $K_D$  heat maps show side chain preferences using H14 helices alone and in the presence of native ligands as competitors (left and right, respectively). Foldamers with significant replacement percentages (>80% for RecQ-WH, >90% for the other proteins) are marked with an open circle, while foldamer hits with significant  $K_D$  increase are marked with a filled circle. Exact  $K_D$  values are given in Tables S1–S5.†



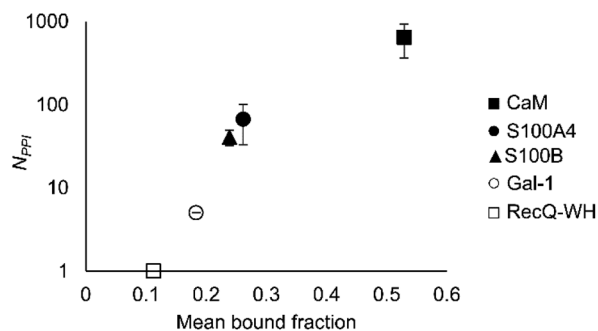


Fig. 6 Average bound fractions for the H14 helical LSM library compared with the mean number of PPIs found in databases BioGRID, Wiki-Pi, GPS-Prot, IntAct, and APID. For the lectin, Gal-1,  $N_{\text{PPI}}$  is based on ref. 66.

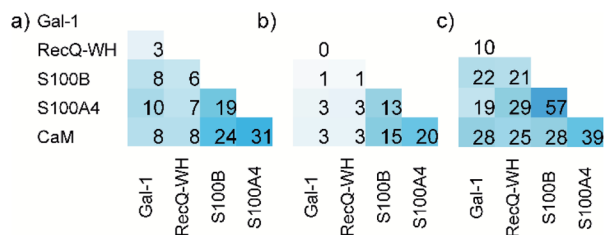


Fig. 7 Pairwise scaled covariances (% maximum similarity: 100%, zero covariance: 0%) of H14 LSM  $F_B$  patterns calculated for total (a) and the orthosteric (b)  $F_B$  values. The sequence homologies of proteins are given in panel (c).

and pathways in signal transduction are different and require different PPI interfaces. Scaled covariances for the orthosteric binders in the range of 13–20% were found among these proteins. This low level of interface similarity detected by LSM affinity patterns supports the conclusion that LSM probes are able to distinguish the proteins with different functions even with high levels of structural similarity (Fig. 7c).

### Side chain binding propensities are biomimetic

Features at the level of the side chains that are responsible for binding of LSM probes were also investigated. The analysis of the affinity patterns revealed that binding affinity did not depend on the dominant presence of highly hydrophobic cyclic  $\beta$ -amino acid residues (ACHC and ACPC). If the sheer hydrophobic surface of these residues drove interaction with the test proteins, higher uniform baseline binding would have been observed. Many sequences displayed no significant binding. Interestingly, certain levels of diagonal symmetry could be observed in the affinity maps (Fig. 3), suggesting that the direction of the order of the  $R^1$  and  $R^2$  proteinogenic side chains exhibits a limited effect on binding. These findings indicate that protein–LSM interactions are driven by proteinogenic side chains present on the compact scaffold.

Focusing on the native-like anchor points, normalized frequencies ( $w_j$ ) of the residues at protein–LSM contacts were analysed. The literature definition of  $w_j$ <sup>37</sup> is given in eqn (1):

$$w_j = \frac{f_j}{\sum_m f_m \times v_j} \quad (1)$$

where  $f_j$  is the number of interface residues of type  $j$ , and normalization factor  $v_j$  is the overall prevalence of the amino acid of type  $j$ . Index  $m$  denotes the residue type. To approximate residue frequencies based on bound fractions for the H14 LSM library, the following formula was applied:

$$w_j = \frac{\sum_i F_B^{ij}}{2 \times \sum_i F_B^i \times v_j} = \frac{\sum_i F_B^{ij}}{2 \times \sum_i F_B^i \times 0.0625} \quad (2)$$

Index  $i$  denotes the LSM sequence, and index  $j$  is the residue type.  $F_B^{ij}$  is the bound fraction of the LSM probe  $i$ , if it contains residue type  $j$ .  $F_B^{ij}$  is zero, if sequence  $i$  does not contain residue type  $j$ . This formula assumes that a binder LSM probe has both proteinogenic side chains in contact with the protein interface. The LSM library was uniform with respect to the 16 side chains, and therefore  $v_j = 1/16 = 0.0625$  was applied as a uniform normalization factor. Normalized frequencies were compared with literature data dissecting large datasets from the Protein Data Bank: (i) computational Ala scanning categorized by secondary structure types<sup>37</sup> and (ii) protein–protein interface analysis, defining interface residues with a maximum distance of 6 Å for the contacting residues<sup>84</sup> (Fig. 8, calculation details in the ESI†).

The LSM library displayed enrichment for residues W, F, Y, I, L, M, V, and R and overall depletion for residues A, D, E, N, Q, S, and T was observed. Statistically significant (Z-score,  $p < 0.05$ ) changes were observed for W, F, I, L, D, E, N, S and T (Table S6†). These results are qualitatively consistent with analyses of the PPI interface databases that demonstrate enrichment of aromatic<sup>37,84</sup> and aliphatic side chains at PPI interfaces,<sup>85</sup> along with the importance of ion pair interactions.<sup>84,86</sup> Interestingly, engineered proteins exhibit a more dominant enrichment in

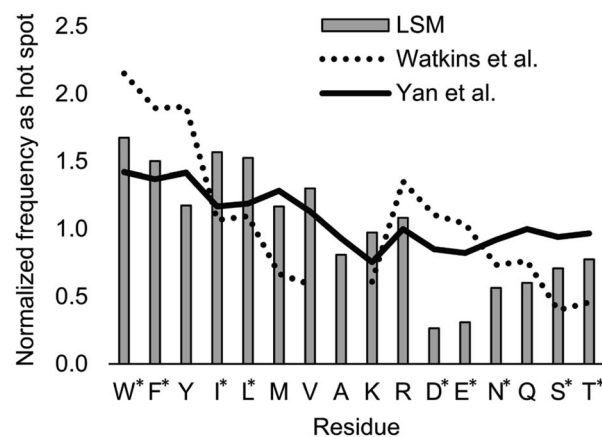


Fig. 8 Normalized frequencies obtained from literature data and calculated for the H14 LSM library. Values from Watkins *et al.*<sup>37</sup> and Yan *et al.*<sup>84</sup> are based on computational Ala scanning and on PPI interface analysis, respectively. LSM data were calculated from experimentally determined  $F_B$  data using eqn (2). The enrichment and depletion of residues marked with asterisks (\*) were found to be statistically significant (Z-score,  $p < 0.05$ ).



aromatic and ionic residues.<sup>87</sup> However, variations of the side chain frequencies have been observed across different structural databases.<sup>37,84,85,88,89</sup>

For the test protein set, side chain frequency levels were dependent on specific proteins (see ESI, Fig. S9†), and enrichment/depletion values can be associated with the selectivity of PPI interactions. Initially, foldameric LSM probes were hypothesized to respond to key side chain requirements of binding at orthosteric sites, especially sites involving electrostatic and H-bonding interactions.

For CaM, a high-affinity interaction is possible only if cationic residues are present in specific patterns on the mostly hydrophobic helical interface.<sup>90</sup> Inspection of binding data (Fig. 3) clearly revealed that residues Lys and Arg were enriched in the binder LSM probes. S100A4 interacts with sequences containing Ser, Thr, and Met as key residues.<sup>91</sup> The selection toward these side chains is evident in binding fingerprints of S100A4. Protein S100B displays different binding preferences compared with S100A4.<sup>92</sup> In protein complexes of S100B, besides Leu and Ile, cationic side chains are also prevalent as anchoring points of interacting partners.<sup>83,93–95</sup> NMR experiments reported that the binding surface of the S100B-p53 peptide complex is defined by Arg, Lys, Leu, and Ile residues of p53, forming salt bridges with Glu side chains and participating in hydrophobic contacts with a hydrophobic patch on S100B.<sup>83,93–95</sup> As opposed to S100A4, the specific Arg–Lys block was observed in LSM probe binding fingerprints of S100B.

Gal-1 and the RecQ-WH domain were models for proteins with a low tendency to form direct protein–protein complexes. Gal-1 does not exhibit preference for specific hydrophilic side chains, and accordingly, its LSM affinity pattern did not show a tendency to bind other than hydrophobic side chains. The binding cleft of RecQ-WH includes a hydrophobic graft flanked by polar and positively charged residues<sup>96</sup> that recognizes the intrinsically disordered and highly anionic C-terminal of SSB<sup>97</sup> (DFDDDIPF). This segment fits into a narrow and slightly bent site on RecQ-WH. As expected, surface mapping of RecQ-WH with bulky LSM probes led to a low number of hits. Even the propensity of the protein to bind anionic sequences did not drive anionic LSM probes to the binding site, possibly because of steric incompatibility.

These findings suggest that the protein interfaces studied select preferred side chain chemistries from the LSM probe library in a manner similar to selection of natural partners. On this ground, we concluded that LSM probes exhibit biomimetic features in presenting proteinogenic side chains for their protein partners.

## Conclusions

Foldameric LSM probe libraries were constructed with H14 and H12 scaffolds and were tested with five proteins. Affinity patterns of 512 LSM probes were recorded, and the resulting 2560 protein–foldamer interactions were analysed using a fragment-centric approach. LSM affinities were secondary structure dependent, and a sufficient number of hits in the affinity range of 1–150  $\mu$ M was obtained for the hexameric H14 library. Binding was driven

by two proteinogenic side chains displayed. The hydrophobic foldamer scaffold served as a template and solvent shield. The better performance of H14 LSM probes may be explained by their ability to geometrically adapt their stable and bulky skeleton to the local environment at protein hot spots. Foldameric LSMs sensitively probed protein surfaces, and thus, detected both orthosteric and non-orthosteric hot spots, based on competition experiments. Orthosteric hits identified in this way along with their geometrically confined binding to the PPI interface offer the possibility of linking them to a high-affinity ligand. This concept was demonstrated earlier in a dynamic covalent system with CaM.<sup>32</sup> Despite the limited structural complexity of LSM probes, affinity fingerprints were characteristic of target proteins. The overall tendency to bind LSM probes correlates with the number of PPI partners identified for the protein. Pairwise correlations between orthosteric binding patterns remained low even for proteins with high structural homology. LSM probes are thus sensitive to the function encoded by PPI interfaces. LSM probes achieve such biomimetic behaviour by presenting proteinogenic side chains in a manner similar to natural PPIs. This behaviour is reflected in protein-like enrichment and depletion values obtained for the side chain chemistries. Moreover, test proteins were able to select key residue types needed for specific PPI recognition patterns from the libraries.

In the early days of chemical biology, Linus Pauling proposed target-induced self-organization of biopolymers as a possible mechanism to generate high-affinity ligands against proteins.<sup>98</sup> His chemistry-inspired concept did not explain the operation of the adaptive immune system, but Pauling's ingenious idea is still there mostly untapped to offer a route to synthetic biopolymers that mimic molecular recognition properties of natural antibodies. This work takes a step toward Pauling's concept, using self-organizing protein mimetic LSM probes as promising candidates for recognition segments. An additional understanding of dynamic covalent chemistry may eventually solve the problem of *in situ* coupling.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

Prof. Andrew Wilson (University of Leeds) is acknowledged for discussion about Pauling's idea. This research was funded by the National Research, Development and Innovation Office of Hungary, grant number GINOP-2.2.1-15-2016-00007 (T. A. M.), K134754 (T. A. M.) and K119359 (L. N.). T. A. M. acknowledges support from the Hungarian Academy of Sciences LEND-ULET-Foldamer and L. N. from HunProtEx 2018-1.2.1-NKP-2018-00005. Ministry of Human Capacities, Hungary grant 20391-3/2018/FEKUSTRAT is acknowledged.

## Notes and references

- 1 V. Azzarito, K. Long, N. S. Murphy and A. J. Wilson, *Nat. Chem.*, 2013, 5, 161–173.



- 2 S. H. Hewitt, M. H. Filby, E. Hayes, L. T. Kuhn, A. P. Kalverda, M. E. Webb and A. J. Wilson, *ChemBioChem*, 2017, **18**, 223–231.
- 3 S. H. Hewitt and A. J. Wilson, *Chem. Commun.*, 2016, **52**, 9745–9756.
- 4 C. Tsiamantas, S. Kwon, C. Douat, I. Huc and H. Suga, *Chem. Commun.*, 2019, **55**, 7366–7369.
- 5 J. M. Rogers, S. Kwon, S. J. Dawson, P. K. Mandal, H. Suga and I. Huc, *Nat. Chem.*, 2018, **10**, 405–412.
- 6 C. Melo Czekster, W. E. Robertson, A. S. Walker, D. Soll and A. Schepartz, *J. Am. Chem. Soc.*, 2016, **138**, 5194–5197.
- 7 J. A. Iannuzzelli and R. Fasan, *Chem. Sci.*, 2020, **11**, 6202–6208.
- 8 A. M. Davis, A. T. Plowright and E. Valeur, *Nat. Rev. Drug Discovery*, 2017, **16**, 681–698.
- 9 W. S. Horne and T. N. Grossmann, *Nat. Chem.*, 2020, **12**, 331–337.
- 10 Z. P. Gates, A. A. Vinogradov, A. J. Quartararo, A. Bandyopadhyay, Z. N. Choo, E. D. Evans, K. H. Halloran, A. J. Mijalis, S. K. Mong, M. D. Simon, E. A. Standley, E. D. Styduhar, S. Z. Tasker, F. Touti, J. M. Weber, J. L. Wilson, T. F. Jamison and B. L. Pentelute, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E5298–E5306.
- 11 R. Boehringer, B. Kieffer and V. Torbeev, *Chem. Sci.*, 2018, **9**, 5594–5599.
- 12 J. A. Kritzer, N. W. Luedtke, E. A. Harker and A. Schepartz, *J. Am. Chem. Soc.*, 2005, **127**, 14584–14585.
- 13 L. R. Whitby and D. L. Boger, *Acc. Chem. Res.*, 2012, **45**, 1698–1709.
- 14 Z. E. Reinert and W. S. Horne, *Chem. Sci.*, 2014, **5**, 3325–3330.
- 15 R. W. Cheloha, B. Chen, N. N. Kumar, T. Watanabe, R. G. Thorne, L. Li, T. J. Gardella and S. H. Gellman, *J. Med. Chem.*, 2017, **60**, 8816–8833.
- 16 S. Liu, R. W. Cheloha, T. Watanabe, T. J. Gardella and S. H. Gellman, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 12383–12388.
- 17 R. W. Cheloha, A. W. Woodham, D. Bousbaine, T. Wang, S. Liu, J. Sidney, A. Sette, S. H. Gellman and H. L. Ploegh, *J. Immunol.*, 2019, **203**, 1619–1628.
- 18 G. A. Eddinger and S. H. Gellman, *Angew. Chem., Int. Ed.*, 2018, **57**, 13829–13832.
- 19 K. J. Peterson-Kaufman, H. S. Haase, M. D. Boersma, E. F. Lee, W. D. Fairlie and S. H. Gellman, *ACS Chem. Biol.*, 2015, **10**, 1667–1675.
- 20 J. D. Sadowsky, M. A. Schmitt, H. S. Lee, N. Umezawa, S. Wang, Y. Tomita and S. H. Gellman, *J. Am. Chem. Soc.*, 2005, **127**, 11966–11968.
- 21 K. L. George and W. S. Horne, *J. Am. Chem. Soc.*, 2017, **139**, 7931–7938.
- 22 L. Fulop, I. M. Mandity, G. Juhasz, V. Szegedi, A. Hetenyi, E. Weber, Z. Bozso, D. Simon, M. Benko, Z. Kiraly and T. A. Martinek, *PLoS One*, 2012, **7**, e39485.
- 23 E. Bartus, G. Olajos, I. Schuster, Z. Bozso, M. A. Deli, S. Veszelka, F. R. Walter, Z. Datki, Z. Szakonyi, T. A. Martinek and L. Fulop, *Molecules*, 2018, **23**, 2523.
- 24 S. H. Hewitt and A. J. Wilson, *Eur. J. Org. Chem.*, 2018, **2018**, 1872–1879.
- 25 G. M. Burslem, H. F. Kyle, A. L. Breeze, T. A. Edwards, A. Nelson, S. L. Warriner and A. J. Wilson, *Chem. Commun.*, 2016, **52**, 5421–5424.
- 26 G. Olajos, E. Bartus, I. Schuster, G. Lautner, R. E. Gyurcsanyi, T. Szogi, L. Fulop and T. A. Martinek, *Anal. Chim. Acta*, 2017, **960**, 131–137.
- 27 C. M. Lin, M. Arancillo, J. Whisenant and K. Burgess, *Angew. Chem., Int. Ed.*, 2020, **59**, 9398–9402.
- 28 S.-Q. Zhang, H. Huang, J. Yang, H. T. Kratochvil, M. Lolicato, Y. Liu, X. Shu, L. Liu and W. F. DeGrado, *Nat. Chem. Biol.*, 2018, **14**, 870–875.
- 29 A. Lombardi, F. Pirro, O. Maglio, M. Chino and W. F. DeGrado, *Acc. Chem. Res.*, 2019, **52**, 1148–1159.
- 30 Z. Hegedus, C. M. Grison, J. A. Miles, S. Rodriguez-Marin, S. L. Warriner, M. E. Webb and A. J. Wilson, *Chem. Sci.*, 2019, **10**, 3956–3962.
- 31 V. Azzarito, J. A. Miles, J. Fisher, T. A. Edwards, S. L. Warriner and A. J. Wilson, *Chem. Sci.*, 2015, **6**, 2434–2443.
- 32 E. Bartus, Z. Hegedus, E. Weber, B. Csipak, G. Szakonyi and T. A. Martinek, *ChemistryOpen*, 2017, **6**, 236–241.
- 33 D. Rooklin, C. Wang, J. Katigbak, P. S. Arora and Y. Zhang, *J. Chem. Inf. Model.*, 2015, **55**, 1585–1599.
- 34 M. G. Wuo, S. H. Hong, A. Singh and P. S. Arora, *J. Am. Chem. Soc.*, 2018, **140**, 16284–16290.
- 35 A. M. Beekman, M. M. D. Cominetti, S. J. Walpole, S. Prabhu, M. A. O'Connell, J. Angulo and M. Searcey, *Chem. Sci.*, 2019, **10**, 4502–4508.
- 36 M. Gao and J. Skolnick, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 22517–22522.
- 37 A. M. Watkins, R. Bonneau and P. S. Arora, *J. Am. Chem. Soc.*, 2016, **138**, 10386–10389.
- 38 P. J. Hajduk and J. Greer, *Nat. Rev. Drug Discovery*, 2007, **6**, 211–219.
- 39 D. Seebach and J. L. Matthews, *Chem. Commun.*, 1997, **1997**, 2015–2022.
- 40 A. Hetenyi, I. M. Mandity, T. A. Martinek, G. K. Toth and F. Fulop, *J. Am. Chem. Soc.*, 2005, **127**, 547–553.
- 41 A. Hetenyi, Z. Szakonyi, I. M. Mandity, E. Szolnoki, G. K. Toth, T. A. Martinek and F. Fulop, *Chem. Commun.*, 2009, **2009**, 177–179.
- 42 J. A. Kritzer, J. D. Lear, M. E. Hodsdon and A. Schepartz, *J. Am. Chem. Soc.*, 2004, **126**, 9468–9469.
- 43 E. A. Harker, D. S. Daniels, D. A. Guarracino and A. Schepartz, *Bioorg. Med. Chem.*, 2009, **17**, 2038–2046.
- 44 W. S. Horne, M. D. Boersma, M. A. Windsor and S. H. Gellman, *Angew. Chem., Int. Ed.*, 2008, **47**, 2853–2856.
- 45 O. M. Stephens, S. Kim, B. D. Welch, M. E. Hodsdon, M. S. Kay and A. Schepartz, *J. Am. Chem. Soc.*, 2005, **127**, 13126–13127.
- 46 E. P. English, R. S. Chumanov, S. H. Gellman and T. Compton, *J. Biol. Chem.*, 2006, **281**, 2661–2667.
- 47 H. S. Haase, K. J. Peterson-Kaufman, S. K. Lan Levensgood, J. W. Checco, W. L. Murphy and S. H. Gellman, *J. Am. Chem. Soc.*, 2012, **134**, 7652–7655.
- 48 S. H. Gellman, *Acc. Chem. Res.*, 1998, **31**, 173–180.
- 49 W. S. Horne and S. H. Gellman, *Acc. Chem. Res.*, 2008, **41**, 1399–1408.





- 50 R. Gopalakrishnan, A. I. Frolov, L. Knerr, W. J. Drury 3rd and E. Valeur, *J. Med. Chem.*, 2016, **59**, 9599–9621.
- 51 V. Kasproicz, S. M. Ward, A. Turner, A. Grammatikos, B. E. Nolan, L. Lewis-Ximenez, C. Sharp, J. Woodruff, V. M. Fleming, S. Sims, B. D. Walker, A. K. Sewell, G. M. Lauer and P. Klenerman, *J. Clin. Invest.*, 2008, **118**, 1143–1153.
- 52 M. M. Hann, *MedChemComm*, 2011, **2**, 349–355.
- 53 M. M. Hann and G. M. Keseru, *Nat. Rev. Drug Discovery*, 2012, **11**, 355–365.
- 54 B. J. Smith, E. F. Lee, J. W. Checco, M. Evangelista, S. H. Gellman and W. D. Fairlie, *ChemBioChem*, 2013, **14**, 1564–1572.
- 55 V. Azzarito, J. A. Miles, J. Fisher, T. A. Edwards, S. L. Warriner and A. J. Wilson, *Chem. Sci.*, 2015, **6**, 2434–2443.
- 56 I. M. Mandity, E. Weber, T. A. Martinek, G. Olajos, G. K. Toth, E. Vass and F. Fulop, *Angew. Chem., Int. Ed.*, 2009, **48**, 2171–2175.
- 57 M. Zhang and T. Yuan, *Biochem. Cell Biol.*, 1998, **76**, 313–323.
- 58 A. R. Bresnick, D. J. Weber and D. B. Zimmer, *Nat. Rev. Cancer*, 2015, **15**, 96–109.
- 59 I. Salama, P. S. Malone, F. Mihaimed and J. L. Jones, *Eur. J. Surg. Oncol.*, 2008, **34**, 357–364.
- 60 D. D. Wang and A. Bordey, *Prog. Neurobiol.*, 2008, **86**, 342–367.
- 61 C. Stark, B. J. Breitzkreutz, T. Reguly, L. Boucher, A. Breitzkreutz and M. Tyers, *Nucleic Acids Res.*, 2006, **34**, D535–D539.
- 62 N. Orii and M. K. Ganapathiraju, *PLoS One*, 2012, **7**, e49029.
- 63 M. E. Fahey, M. J. Bennett, C. Mahon, S. Jäger, L. Pache, D. Kumar, A. Shapiro, K. Rao, S. K. Chanda, C. S. Craik, A. D. Frankel and N. J. Krogan, *BMC Bioinf.*, 2011, **12**, 298.
- 64 S. Orchard, M. Ammari, B. Aranda, L. Breuza, L. Briganti, F. Broackes-Carter, N. H. Campbell, G. Chavali, C. Chen, N. del-Toro, M. Duesbury, M. Dumousseau, E. Galeota, U. Hinz, M. Iannuccelli, S. Jagannathan, R. Jimenez, J. Khadake, A. Lagreid, L. Licata, R. C. Lovering, B. Meldal, A. N. Melidoni, M. Milagros, D. Peluso, L. Peretto, P. Porras, A. Raghunath, S. Ricard-Blum, B. Roechert, A. Stutz, M. Tognolli, K. van Roey, G. Cesareni and H. Hermjakob, *Nucleic Acids Res.*, 2014, **42**, D358–D363.
- 65 D. Alonso-Lopez, F. J. Campos-Laborie, M. A. Gutierrez, L. Lambourne, M. A. Calderwood, M. Vidal and J. De Las Rivas, *Database*, 2019, **2019**, baz005.
- 66 I. Camby, M. Le Mercier, F. Lefranc and R. Kiss, *Glycobiology*, 2006, **16**, 137R–157R.
- 67 M. F. Lopez-Lucendo, D. Solis, S. Andre, J. Hirabayashi, K. Kasai, H. Kaltner, H. J. Gabius and A. Romero, *J. Am. Chem. Soc.*, 2004, **343**, 957–970.
- 68 D. Gupta, M. Cho, R. D. Cummings and C. F. Brewer, *Biochemistry*, 1996, **35**, 15236–15243.
- 69 I. V. Nesmelova, E. Ermakova, V. A. Daragan, M. Pang, M. Menendez, L. Lagartera, D. Solis, L. G. Baum and K. H. Mayo, *J. Mol. Biol.*, 2010, **397**, 1209–1230.
- 70 H. J. Gabius, *Int. J. Biochem.*, 1994, **26**, 469–477.
- 71 O. Blaževič, Y. G. Mideksa, M. Šolman, A. Ligabue, N. Ariotti, H. Nakhaeizadeh, E. K. Fansa, A. C. Papageorgiou, A. Wittinghofer, M. Ahmadian and D. Abankwa, *Sci. Rep.*, 2016, **6**, 24165.
- 72 Z. Bojic-Trbojevic, M. Jovanovic Krivokuca, I. Stefanoska, N. Kolundzic, A. Vilotic, T. Kadoya and L. Vicovac, *J. Biochem.*, 2018, **163**, 39–50.
- 73 K. Umezū, K. Nakayama and H. Nakayama, *Proc. Natl. Acad. Sci. U. S. A.*, 1990, **87**, 5363–5367.
- 74 D. A. Bernstein, M. C. Zittel and J. L. Keck, *EMBO J.*, 2003, **22**, 4910–4921.
- 75 R. D. Shereda, D. A. Bernstein and J. L. Keck, *J. Biol. Chem.*, 2007, **282**, 19247–19258.
- 76 K. P. Hoefflich and M. Ikura, *Cell*, 2002, **108**, 739–742.
- 77 H. Tidow and P. Nissen, *FEBS J.*, 2013, **280**, 5551–5565.
- 78 S. Lievens, S. Eyckerman, I. Lemmens and J. Tavernier, *Expert Rev. Proteomics*, 2010, **7**, 679–690.
- 79 A. Patil, K. Nakai and H. Nakamura, *Nucleic Acids Res.*, 2011, **39**, D744–D749.
- 80 J. Snider, M. Kotlyar, P. Saraon, Z. Yao, I. Jurisica and I. Stagljär, *Mol. Syst. Biol.*, 2015, **11**, 848.
- 81 B. Lehne and T. Schlitt, *Hum. Genomics*, 2009, **3**, 291.
- 82 A. Gomez, S. Hernandez, I. Amela, J. Pinol, J. Cedano and E. Querol, *Mol. Biosyst.*, 2011, **7**, 2379–2382.
- 83 G. Gogl, A. Alexa, B. Kiss, G. Katona, M. Kovacs, A. Bodor, A. Remenyi and L. Nyitray, *J. Biol. Chem.*, 2016, **291**, 11–27.
- 84 C. Yan, F. Wu, R. L. Jernigan, D. Dobbs and V. Honavar, *Protein J.*, 2008, **27**, 59–70.
- 85 A. A. Bogan and K. S. Thorn, *J. Mol. Biol.*, 1998, **280**, 1–9.
- 86 I. S. Moreira, P. A. Fernandes and M. J. Ramos, *Proteins*, 2007, **68**, 803–812.
- 87 M. G. Wuo and P. S. Arora, *Curr. Opin. Chem. Biol.*, 2018, **44**, 16–22.
- 88 B. N. Bullock, A. L. Jochim and P. S. Arora, *J. Am. Chem. Soc.*, 2011, **133**, 14220–14223.
- 89 S. Lise, C. Archambeau, M. Pontil and D. T. Jones, *BMC Bioinf.*, 2009, **10**, 365.
- 90 K. Mruk, B. M. Farley, A. W. Ritacco and W. R. Kobertz, *J. Gen. Physiol.*, 2014, **144**, 105–114.
- 91 B. Kiss, A. Duelli, L. Radnai, K. A. Kékesi, G. Katona and L. Nyitray, *Proc. Natl. Acad. Sci. U. S. A.*, 2012.
- 92 M. C. Cavalier, A. D. Pierce, P. T. Wilder, M. J. Alasady, K. G. Hartman, D. B. Neau, T. L. Foley, A. Jadhav, D. J. Maloney, A. Simeonov, E. A. Toth and D. J. Weber, *Biochemistry*, 2014, **53**, 6628–6640.
- 93 M. Ikura and K. L. Yap, *Nat. Struct. Mol. Biol.*, 2000, **7**, 525.
- 94 R. R. Rust, D. M. Baldisseri and D. J. Weber, *Nat. Struct. Mol. Biol.*, 2000, **7**, 570.
- 95 D. J. Weber, R. R. Rustandi, F. Carrier and D. B. Zimmer, in *Calcium: The molecular basis of calcium action in biology and medicine*, Springer, 2000, pp. 521–539.
- 96 R. D. Shereda, N. J. Reiter, S. E. Butcher and J. L. Keck, *J. Mol. Biol.*, 2009, **386**, 612–625.
- 97 S. Raghunathan, A. G. Kozlov, T. M. Lohman and G. Waksman, *Nat. Struct. Biol.*, 2000, **7**, 648–652.
- 98 L. Pauling, *J. Am. Chem. Soc.*, 1940, **62**, 2643–2657.

