



Cite this: *RSC Chem. Biol.*, 2021, 2, 1474

Received 26th June 2021.
Accepted 30th July 2021

DOI: 10.1039/d1cb00137j

rsc.li/rsc-chembio

A simple-to-implement and experimentally validated computational workflow for sequence modification of peptide inhibitors of protein–protein interactions (PPIs) is described.

Understanding and modulating PPIs is important both for delineating molecular mechanisms of healthy cells and disease states, and for directing drug discovery.^{1–3} However, PPIs are challenging targets for molecular design.⁴ A significant proportion of PPIs rely on short peptide motifs (SPMs)⁵ for affinity. SPMs are often found in intrinsically disordered regions of proteins⁶ and undergo disorder-to-order transitions to adopt defined structures, *e.g.* α helices⁷ and β -strands,⁸ on interaction with a target domain. The sequences of SPMs serve as powerful templates for inhibitor design, and have motivated efforts to develop peptide and peptidomimetic ligands,^{9,10} including stapled¹¹ and macrocyclic peptides,¹² as potential therapeutics. Strategies for peptide-based ligand development should explore sequence space to maximise binding affinity whilst maintaining good pharmacokinetic properties, such as solubility, cell permeability, and resistance to proteolysis. Concerning affinity, truncation and the identification of hot residues (*i.e.* side chains that contribute significantly to the affinity of a PPI)^{13,14} through alanine scanning¹⁵ represent practical approaches for obtaining key information on the minimal determinants of binding. Enhancing affinity and/or selectivity through sequence variation is also desirable as exemplified by alternative systematic experimental

Towards optimizing peptide-based inhibitors of protein–protein interactions: predictive saturation variation scanning (PreSaVS)†

Kristina Hetherington,^{‡,ab} Som Dutt,^{‡,ab} Amaury A. Ibarra,^{‡,c} Emma E. Cawood,^{‡,ab} Fruzsina Hobor,^{ad} Derek N. Woolfson,^{‡,cef} Thomas A. Edwards,^{‡,ad} Adam Nelson,^{‡,ab} Richard B. Sessions^{‡,cef} and Andrew J. Wilson^{‡,ab}

sequence variation strategies, *e.g.* hydrophile scanning.¹⁶ However, generally, sequence space is too great to explore using synthetic chemistry. Biological selection methods are powerful and do allow more space to be searched,^{17–19} however, these can be experimentally demanding and expensive. For all these reasons, development and improvement of computational approaches to examine and design PPIs are important.

Computational alanine scanning (CAS) is an important tool within this armoury to speed up and direct experiments.^{20,21} Recent studies show the power of computational design;^{22,23} *e.g.*, the use of Rosetta to identify helix bundles that selectively modulate BCL-2 family interactions.²³ Affinity mapped SORTCERY uses biological selection and deep sequencing to obtain binding information and then develop computational models of sequence-binding relationships to inform peptide design.²⁴ Statistical data on tertiary structural motifs (TERMs) in the RCSB Protein Data Bank (PDB) and the resulting TERM energies (dTERMen) have been used in design, predicting peptide binding energies as accurately as structure-based tools, leading to high-affinity peptide binders.²⁵ However, these methods rely on large data sets and/or multiple experimental designs being pursued. Rosetta Backrub identifies tolerated sequences using flexible backbone protein design. It uses a simulated annealing and genetic algorithm optimization method to create a single estimate for the tolerated sequence space.^{26–28} Finally, AlphaSpace can be used to identify unoccupied spaces in PPIs, which can be filled by natural or unnatural side chains in designs targeting interfaces.²⁹

Here, we describe a simple-to-implement and experimentally validated computational workflow for sequence modification of peptide-based PPI inhibitors. We call this *in silico* Predictive Saturation Variation Scanning (PreSaVS). Rather than knocking-out affinity (as in CAS), modifying a sequence whilst retaining or even improving its potency is challenging: it is unclear if further optimization of hot residues, or making new interactions using non-hot residues³⁰ is the best approach and to what extent affinity can be increased. PreSaVS computationally substitutes each residue in a peptide sequence of

^a Astbury Centre for Structural Molecular Biology, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK. E-mail: a.j.wilson@leeds.ac.uk

^b School of Chemistry, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK

^c School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK. E-mail: r.sessions@bristol.ac.uk

^d School of Molecular and Cellular Biology, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK

^e School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK

^f BrisSynBio, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK

† Electronic supplementary information (ESI) available. See DOI: [10.1039/d1cb00137j](https://doi.org/10.1039/d1cb00137j)

‡ Authors contributed equally to this work.



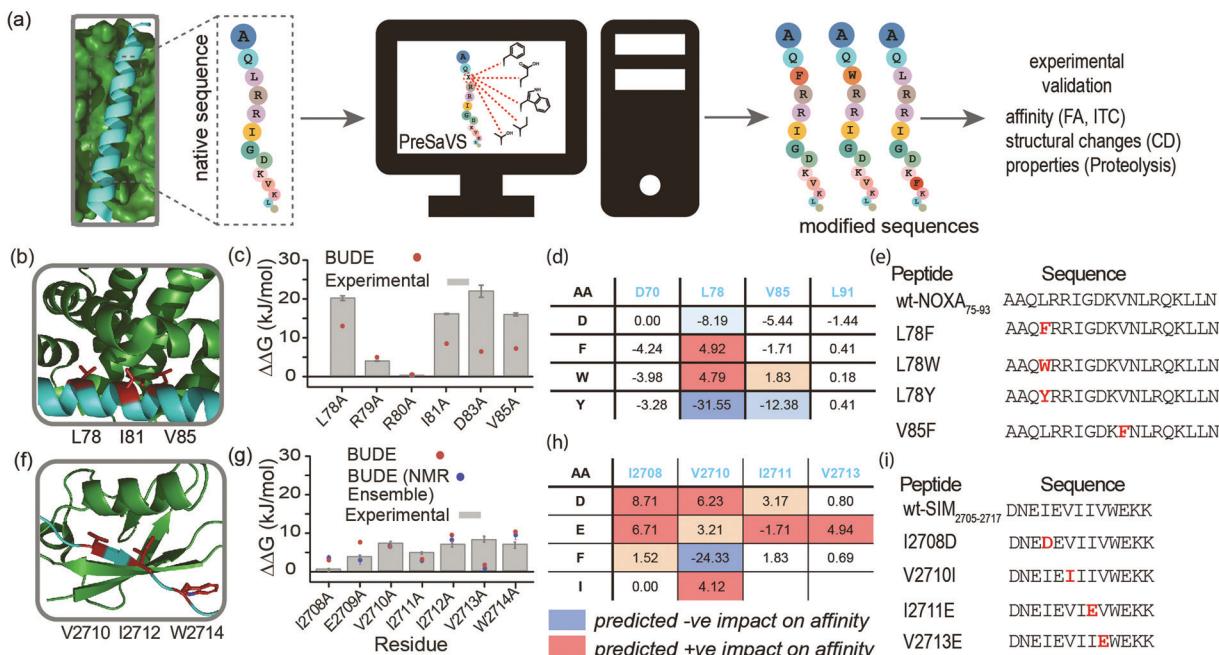


Fig. 1 PreSaVS applied to NOXA₇₅₋₉₃/MCL-1 and SIM₂₇₀₅₋₂₇₁₇/SUMO. (a) Schematic depicting workflow; (b) lowest energy NMR-derived structure (PDB ID: 2J6M) of *m*NOXA₆₈₋₉₃ (cyan)/*m*MCL-1 (green, hot residues in red); (c) BudeAlaScan and experimental data for NOXA₇₅₋₉₃/MCL-1; (d) key PreSaVS results for NOXA₇₅₋₉₃; (e) NOXA₇₅₋₉₃ sequences selected for experimental analyses; (f) lowest energy NMR derived structure (PDB ID: 2LAS) of SIM₂₇₀₅₋₂₇₁₇ (cyan)/SUMO (green, hot residues in red); (g) BudeAlaScan and experimental data for SIM₂₇₀₅₋₂₇₁₇/SUMO; (h) key PreSaVS results for SIM₂₇₀₅₋₂₇₁₇; (i) SIM₂₇₀₅₋₂₇₁₇ sequences selected for experimental analyses.

interest to 16 of the proteinogenic amino acids (*i.e.*, standard residues except Ala, Gly, Pro and Cys) and calculates the difference in binding free energy ($\Delta\Delta G$) relative to the native sequence (Fig. 1a). To do this, we adapted BUDE alanine scanning^{20,31} to allow variation to different amino acids (see ESI†). The soft nature of the BUDE forcefield avoids penalising small geometric overlaps in the modelled structures and this is key to the speed of these methods, unlike the Rosetta forcefield, where extensive sampling is required.²⁰ To establish the PreSaVS workflow, it was applied to both α helix- and β strand-mediated PPIs (Fig. 1b–i). A modified *m*NOXA-B/*h*MCL-1^{32,33} interaction (referred to as NOXA₇₅₋₉₃/MCL-1) was selected as the model α -helix-mediated PPI (Fig. 1b). NOXA and MCL-1 are proteins of the B-cell Lymphoma 2 (BCL-2) family of apoptosis regulators,³⁴ and have been the focus of oncology drug-discovery efforts.³⁵ The interaction between the SIM peptide found in the M-IR2 region of RanBP2, and *h*SUMO-1 (referred to as SIM₂₇₀₅₋₂₇₁₇/SUMO)³⁶ was chosen as the model β strand-mediated PPI (Fig. 1f). Small ubiquitin-like modifiers (SUMO) regulate many cellular processes through their interaction with SUMO-interacting motifs (SIMs) within other proteins, and are the subject of ongoing investigation.³⁷

By analogy to hot residues ($\Delta\Delta G \geq -4.5 \text{ kJ mol}^{-1}$), we set $\Delta\Delta G \geq 4.5 \text{ kJ mol}^{-1}$ ^{13,14} as a threshold for experimental analyses. The two PPIs were subjected to PreSaVS and the outcomes visually inspected (Fig. 1d and h; full data in ESI†, Tables S1 and S2). Based on prediction of a favourable increase in affinity, we selected NOXA₇₅₋₉₃ peptides with substitutions at L78 and V85 (both hot residues, Fig. 1b and e) for experimental

analyses, specifically focussing on Trp and Phe variants. For L78, we also selected the Tyr variant as a negative control. Predictions for SIM₂₇₀₅₋₂₇₁₇ identified hot and non-hot residues, focussing predominantly on substitution of hydrophobic for charged amino acids (Fig. 1f–i): I2708D, V2710I (hot residue), I2711E and V2713E (hot residue) variations were selected for experimental validation.

Peptide variants were prepared as N-terminal acetamides and C-terminal amides (see ESI†). These were tested in fluorescence anisotropy (FA) competition assays using a fluorescein-labelled BIM₇₅₋₈₅ peptide (FITC-Ahx-BIM₇₅₋₈₅, $K_d = 204 \pm 16 \text{ nM}$) or SIM₂₇₀₅₋₂₇₁₇ (FITC-peg-SIM₂₇₀₅₋₂₇₁₂, $K_d = 1.5 \pm 0.2 \mu\text{M}$) peptides described previously (see ESI† for sequences and direct titration data, Fig. S1).²⁰ Competition FA revealed that NOXA₇₅₋₉₃L78F and NOXA₇₅₋₉₃L78W variations were tolerated, whilst the NOXA₇₅₋₉₃V85F variant displayed weaker inhibitory activity (Fig. 2a). As predicted by PreSaVS, the NOXA₇₅₋₉₃L78Y variant was a poor inhibitor providing confidence in the predictions. The NOXA₇₅₋₉₃ peptide binds MCL-1 selectively over other BCL-2 family members,³⁸ and peptide variants retained this selectivity in a FA competition assay against a further BCL-2 family member protein, BCL-x_L (Fig. S2, ESI†). Isothermal calorimetry (ITC) experiments showed comparable K_d values for the variant peptides relative to NOXA₇₅₋₉₃, confirming these results; variant peptides exhibited a lower entropy and compensatory decrease in enthalpy of binding compared to NOXA₇₅₋₉₃ (Fig. 2b, Table 1 and Fig. S3, ESI†). Lastly, structural effects of these variations were investigated using circular dichroism (CD) spectroscopy (Fig. S4, ESI†). The peptides



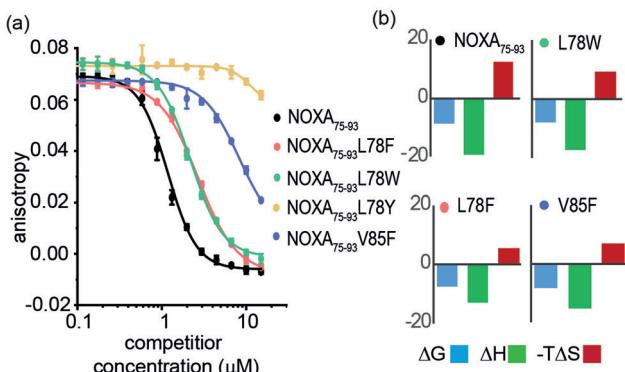


Fig. 2 FA competition and ITC data for NOXA₇₅₋₉₃ and variant sequences. (a) Competition FA (50 mM Tris, 150 mM NaCl, pH 7.5, using 25 nM tracer and 150 nM MCL-1). (b) ITC data (50 mM Tris, 150 mM NaCl, pH 7.5, 150 μM peptide).

exhibited CD spectra consistent with predominance of random coil in buffer (peptides exhibited increased helicity in 30% trifluoroethanol, Fig S5, ESI[†]).⁴⁵ The sequence-driven variation in helicity is greater than the variation in affinity (see Fig. S6, ESI[†]) although both are small supporting the hypothesis that side chain interactions dominate the observed effects of variation as opposed to a significant change in conformational preference.

For SIM₂₇₀₅₋₂₇₁₇ peptide variants (Fig. 3a and Table 2), similar inhibitory potency was observed for SIM₂₇₀₅₋₂₇₁₇ I2708D ($IC_{50} = 14.8 \pm 0.7 \mu\text{M}$) in comparison to SIM₂₇₀₅₋₂₇₁₇ ($IC_{50} = 21.9 \pm 0.3 \mu\text{M}$). In the bound state, the hydrophobic I2708 SIM side chain lies proximal to K45 and K46 of SUMO; swapping this position for Asp may introduce hydrogen-bonding interactions (Fig. 3b). Less surprising was the more conservative SIM₂₇₀₅₋₂₇₁₇ V2710I variant peptide, which maintained SUMO binding with an $IC_{50} = 16.2 \pm 1.6 \mu\text{M}$. (Fig. 3a). Finally, the SIM₂₇₀₅₋₂₇₁₇ I2711E and SIM₂₇₀₅₋₂₇₁₇ V2713E variants lost inhibitory potency. In the NMR ensemble, I2711 rests against the surface of the SUMO protein between the hydrophobic Y21 and the charged K37 residues of (Fig. 3b), implying any benefit from introduction of a salt bridge may be countered by repulsion from the electron-rich aromatic ring. Lastly, the V2713 side chain is solvent exposed (Fig. 3b), so the weaker inhibitory potency of the V2713E variant cannot be reconciled by considering potential interactions it may make with the SUMO interface. Alanine scanning, however, previously showed loss in binding for

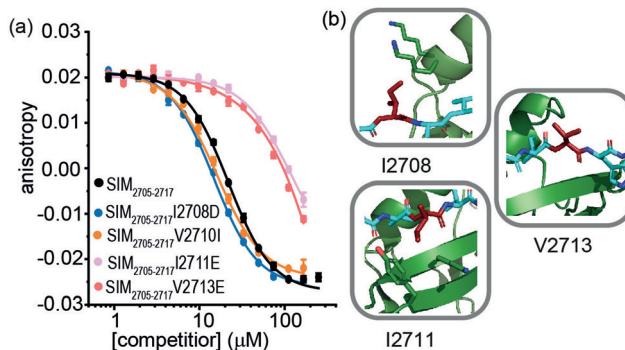


Fig. 3 FA competition data and structural analysis for SIM₂₇₀₅₋₂₇₁₇ and variant sequences; (a) competition FA (50 mM Tris, 150 mM NaCl, 25 nM tracer, 100 nM SUMO); (b) SIM₂₇₀₅₋₂₇₁₇/SUMO (PDB ID: 2LAS) variant residues highlighted in red showing I2708 proximity to SUMO K45 and K46 residues, I2711 proximity to SUMO residues Y21 and K37 and V2713 pointing towards solvent.

Table 2 SUMO binding parameters for SIM₂₇₀₅₋₂₇₁₇ variant peptides^a

	SIM ₂₇₀₅₋₂₇₁₇	I2708D	V2710I	I2711E	V2713E
IC_{50} (μM)	21.9 ± 0.3	14.8 ± 0.7	16.1 ± 1.6	>200	>200

^a Determined using conditions as noted in Fig. 3.

V2713A.²⁰ Valine favours β structure,³⁹ and we attribute the loss of affinity observed for V2713E to a critical structure imposing role for V2713 in adopting a compliant SUMO-binding conformation. CD spectra on SIM peptides were consistent with a random coil conformation as expected (not shown).

An advantage of PreSaVS lies in the ability to identify tolerated variants – in terms of target binding affinity – with more desirable physicochemical properties (Table S3 for calculated physicochemical properties, ESI[†]). Notable changes were observed in proteolysis studies. Both NOXA₇₅₋₉₃ variants tested (L78F and L78Y) exhibited ≈ 10 -fold increased rate of cleavage by α -chymotrypsin relative to NOXA₇₅₋₉₃ (Fig. S7a, ESI[†]). In contrast, SIM variants showed improved Proteinase K stability over SIM₂₇₀₅₋₂₇₁₇ (Fig S7b, ESI[†]), with the greatest protection observed for the V2713E variant (≈ 9 -fold decreased proteolysis rate). For NOXA₇₅₋₉₃, while the peptide bond between residues L78 (P1) and R79 (P1') is a substrate for α -chymotrypsin, cleavage susceptibility for peptides containing Leu, Phe, or Tyr at the P1 position is similar.⁴⁰ Although small, the increased helical propensity of NOXA₇₅₋₉₃ (Fig. S4 and S5, ESI[†]) may contribute to its higher protease resistance compared to the variant sequences. The broad-range specificity of Proteinase K,⁴⁰ renders most peptide bonds in the SIM₂₇₀₅₋₂₇₁₇ sequence targets for this protease. The relative Proteinase K resistance of these peptides (SIM₂₇₀₅₋₂₇₁₇ \sim I2708D \sim V2710I $<$ I2711E $<$ V2713E) is similar to their relative affinity for SUMO. This may arise from subtle changes in protease recognition specificity, although given proteases recognise substrates in an extended, β -like conformation as is the case for the SIM₂₇₀₅₋₂₇₁₇/SUMO interaction, secondary structure propensity may also play a role.

Table 1 MCL-1 binding parameters and helicities for NOXA₇₅₋₉₃ peptides^a

	NOXA ₇₅₋₉₃	L78F	L78W	L78Y	V85F
K_d (μM)	0.4 ± 0.1	1.7 ± 0.1	2.4 ± 0.8	—	0.8 ± 0.1
IC_{50} (μM)	1.2 ± 0.1	2.7 ± 0.1	2.2 ± 0.1	>100	>20
ΔG^b	-8.71	-7.88	-7.66	—	-8.35
ΔH^b	-19.2 ± 1.5	-17.3 ± 3.2	-13.1 ± 3.9	—	-15.3 ± 3.4
$-\Delta S^b$	12.2	9.44	5.48	—	6.90
% helicity ^c	15	8	12	—	9

^a Determined using conditions as noted in Fig. 2. ^b (kJ mol⁻¹).

^c In buffer.



Conclusions

We have developed *in silico* Predictive Saturation Variation Scanning (PreSaVS) and tested it on α helix- and β strand-mediated PPIs. Variants of NOXA_{75–93} and SIM_{2705–2717} peptides generated by PreSaVS retained inhibitory potency/affinity for MCL-1 and SUMO, respectively. Tolerated modifications could be made at hot and non-hot residues. Further experiments revealed changes in proteolysis rates of the peptides as a consequence of the sequence variation. Whereas suppressed proteolysis is often desired, accelerated proteolysis can be advantageous for fast-acting peptides.⁴¹ This validates PreSaVS as a fast predictive tool for sequence variation and further extends the capabilities of the Bristol University Docking Engine (BUDE).⁴² We expect the approach to be useful for other topologies *e.g.* loops, and, whilst NOXA and SIM are intrinsically disordered their PPIs are well defined; thus utility of the approach for “fuzzy” interactions remains to be explored. Ongoing studies are focussed on exploring the scope to identify affinity enhancing and/or selectivity modifying variations across a broader array of PPI targets.⁴³

Author contributions

D. N. W., T. A. E., A. N., R. B. S., and A. J. W., conceived and designed the research program, K. H. and S. D. prepared and tested peptides, A. A. I. developed the computational workflow, E. E. C carried out proteolysis, F. H. produced protein and carried out ITC. The manuscript was written by K. H. and A. J. W. with contributions from all authors.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by EPSRC (EP/N013573/1 and EP/KO39292/1) and the BBSRC/EPSRC-funded Synthetic Biology Research Centre, BrisSynBio (BB/L01386X/1). AN holds an EPSRC Established Career Fellowship (EP/N025652/1).

Notes and references

- 1 J. D. L. Rivas and C. Fontanillo, *PLoS Comput. Biol.*, 2010, **6**, e1000807.
- 2 D. E. Scott, A. R. Bayly, C. Abell and J. Skidmore, *Nat. Rev. Drug Discovery*, 2016, **15**, 533–550.
- 3 L.-G. Milroy, T. N. Grossmann, S. Hennig, L. Brunsved and C. Ottmann, *Chem. Rev.*, 2014, **114**, 4695–4748.
- 4 M. R. Arkin, Y. Tang and J. A. Wells, *Chem. Biol.*, 2014, **21**, 1102–1114.
- 5 P. Tompa, Norman E. Davey, Toby J. Gibson and M. M. Babu, *Mol. Cell*, 2014, **55**, 161–169.
- 6 P. E. Wright and H. J. Dyson, *Nat. Rev. Mol. Cell Biol.*, 2015, **16**, 18–29.

- 7 A. L. Jochim and P. S. Arora, *Mol. BioSyst.*, 2009, **5**, 924–926.
- 8 A. M. Watkins and P. S. Arora, *ACS Chem. Biol.*, 2014, **9**, 1747–1754.
- 9 M. Pelay-Gimeno, A. Glas, O. Koch and T. N. Grossmann, *Angew. Chem., Int. Ed.*, 2015, **54**, 8896–8927.
- 10 V. Azzarito, K. Long, N. S. Murphy and A. J. Wilson, *Nat. Chem.*, 2013, **5**, 161–173.
- 11 H. Wang, R. S. Dawber, P. Zhang, M. Walko, A. J. Wilson and X. Wang, *Chem. Sci.*, 2021, **12**, 5977–5993.
- 12 A. A. Vinogradov, Y. Yin and H. Suga, *J. Am. Chem. Soc.*, 2019, **141**, 4167–4181.
- 13 T. Clackson and J. Wells, *Science*, 1995, **267**, 383–386.
- 14 N. London, B. Raveh and O. Schueler-Furman, *Curr. Opin. Chem. Biol.*, 2013, **17**, 952–959.
- 15 B. Cunningham and J. Wells, *Science*, 1989, **244**, 1081–1085.
- 16 M. D. Boersma, J. D. Sadowsky, Y. A. Tomita and S. H. Gellman, *Protein Sci.*, 2008, **17**, 1232–1240.
- 17 A. I. Ekanayake, L. Sobze, P. Kelich, J. Youk, N. J. Bennett, R. Mukherjee, A. Bhardwaj, F. Wuest, L. Vukovic and R. Derda, *J. Am. Chem. Soc.*, 2021, **143**, 5497–5507.
- 18 S. Chen, S. Lovell, S. Lee, M. Fellner, P. D. Mace and M. Bogyo, *Nat. Biotechnol.*, 2021, **39**, 490–498.
- 19 J. Miles, F. Hobor, C. Trinh, J. Taylor, C. Tiede, P. Rowell, B. Jackson, F. Nadat, P. Ramsahye, H. Kyle, B. Wicky, J. Clarke, D. Tomlinson, A. Wilson and T. Edwards, *ChemBioChem*, 2021, **22**, 232–240.
- 20 A. A. Ibarra, G. J. Bartlett, Z. Hegedüs, S. Dutt, F. Hobor, K. A. Horner, K. Hetherington, K. Spence, A. Nelson, T. A. Edwards, D. N. Woolfson, R. B. Sessions and A. J. Wilson, *ACS Chem. Biol.*, 2019, **14**, 2252–2263.
- 21 I. Massova and P. A. Kollman, *J. Am. Chem. Soc.*, 1999, **121**, 8133–8143.
- 22 M. Yu, L. Ghamsari, J. A. Rotolo, B. J. Kappel and J. M. Mason, *RSC. Chem. Biol.*, 2021, **2**, 656–668.
- 23 S. Berger, E. Prock, D. Margineantu, E. F. Lee, B. W. Shen, A. Zelter, D.-A. Silva, K. Chawla, M. J. Herold, J.-M. Garnier, R. Johnson, M. J. MacCoss, G. Lessene, T. N. Davis, P. S. Stayton, B. L. Stoddard, W. D. Fairlie, D. M. Hockenberry and D. Baker, *eLife*, 2016, **5**, e20352.
- 24 J. M. Jenson, V. Xue, L. Stretz, T. Mandal, L. L. Reich and A. E. Keating, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, E10342–E10351.
- 25 V. Frappier, J. M. Jenson, J. Zhou, G. Grigoryan and A. E. Keating, *Structure*, 2019, **27**, 606–617.e605.
- 26 C. A. Smith and T. Kortemme, *PLoS One*, 2011, **6**, e20451.
- 27 C. A. Smith and T. Kortemme, *J. Mol. Biol.*, 2010, **402**, 460–474.
- 28 F. Lauck, C. A. Smith, G. F. Friedland, E. L. Humphris and T. Kortemme, *Nucleic Acids Res.*, 2010, **38**, W569–W575.
- 29 D. Rooklin, A. E. Modell, H. Li, V. Berdan, P. S. Arora and Y. Zhang, *J. Am. Chem. Soc.*, 2017, **139**, 15560–15563.
- 30 C. N. Reddy, N. Manzar, B. Ateeq and R. Sankararamakrishnan, *Biochemistry*, 2020, **59**, 4379–4394.



31 C. W. Wood, A. A. Ibarra, G. J. Bartlett, A. J. Wilson, D. N. Woolfson and R. B. Sessions, *Bioinformatics*, 2020, **36**, 2917–2919.

32 A. M. Beekman, M. A. O'Connell and L. A. Howell, *ChemMedChem*, 2016, **11**, 840–844.

33 P. E. Czabotar, E. F. Lee, M. F. V. Delft, C. L. Day, B. J. Smith, D. C. S. Huang, W. D. Fairlie, M. G. Hinds and P. M. Colman, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 6217–6222.

34 R. Singh, A. Letai and K. Sarosiek, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 175–193.

35 A. W. Hird and A. E. Tron, *Pharmacol. Ther.*, 2019, **198**, 59–67.

36 A. T. Namanja, Y.-J. Li, Y. Su, S. Wong, J. Lu, L. T. Colson, C. Wu, S. S. C. Li and Y. Chen, *J. Biol. Chem.*, 2012, **287**, 3231–3240.

37 J. R. Gareau and C. D. Lima, *Nat. Rev. Mol. Cell Biol.*, 2010, **11**, 861.

38 M. Certo, V. D. G. Moore, M. Nishino, G. Wei, S. Korsmeyer, S. A. Armstrong and A. Letai, *Cancer Cell*, 2006, **9**, 351–365.

39 V. Muñoz and L. Serrano, *Proteins: Struct., Funct., Bioinf.*, 1994, **20**, 301–311.

40 B. Keil, *Specificity of Proteolysis*, Springer-Verlag, Berlin-Heidelberg, New York, 1992.

41 M. Muttenthaler, G. F. King, D. J. Adams and P. F. Alewood, *Nat. Rev. Drug Discovery*, 2021, **20**, 309–325.

42 S. McIntosh-Smith, J. Price, R. B. Sessions and A. A. Ibarra, *Int. J. High Perform. Comput. Appl.*, 2015, **29**, 119–134.

43 H. T. H. Chan, M. A. Moesser, R. K. Walters, T. R. Malla, R. M. Twidale, T. John, H. M. Deeks, T. Johnston-Wood, V. Mikhailov, R. B. Sessions, W. Dawson, E. Salah, P. Lukacik, C. Strain-Damerell, C. D. Owen, T. Nakajima, K. Świderek, A. Lodziola, V. Moliner, D. R. Glowacki, M. A. Walsh, C. J. Schofield, L. Genovese, D. K. Shoemark, A. J. Mulholland, F. Duarte and G. M. Morris, *bioRxiv*, 2021, DOI: 10.1101/2021.06.18.446355.

