

D-Protein and D-Peptide Technology in Applied Research



Submitted to Cardiff University for the degree of Doctor of
Philosophy by

Alexander John Lander

2022

Supervisors: Dr. Louis Y.P. Luk and Prof. Rudolf K. Allemann

Abstract

Chemical protein synthesis provides access to entire D-protein enantiomers, with unique utility in applied research. Firstly, proteins may be crystallized from a racemic mixture of L- and D-enantiomers, known as racemic protein crystallography, providing facile access to high resolution X-ray structures. In this thesis, to understand how membrane-active bacteriocins have acquired their antibacterial properties, an analysis based on racemic protein crystallography (0.9–1.2 Å) of two key representatives, aureocin a53 (AucA) and lacticin Q (LnqQ) was conducted. Through structural analysis and systematic residue substitutions, conserved surface tryptophans appeared to play important roles in coordination of lipid phosphate and were critical for antibacterial activity. Additional tryptophans in AucA were also involved in forming an oligomeric assembly for stability enhancement. Collectively, racemic protein crystallography shed light on the molecular interactions of tryptophans, demonstrating how a bacteriocin delivers its antibacterial properties.

Secondly, D-enantiomers of protein drug targets can be used in mirror-image phage display (MIPD), allowing discovery of non-proteolytic D-peptide ligands as lead candidates. Development of a D-peptide capable of targeting tumor necrosis factor receptor I (TNFR1) activation could be a beneficial solution in controlling inflammation caused by cytokine storm. However, containing 144 residues and 12 disulfide bonds, synthesis of the D-enantiomer for MIPD would be a significant challenge. Here, isolation of the mid-chain, cytokine binding domain (TNRC2) is reported, and its enantiomer enabled MIPD to discover a cyclic peptide binder with low micromolar affinity. Additionally, the use of computational methods to develop D-protein ligands based only on the target structure was explored.

Table of Contents

Abstract	i
Table of figures	vi
Table of tables	xix
Table of abbreviations	xx
Acknowledgements	xxv
Chapter 1 : Introduction	1
1.1 Research surrounding D-protein enantiomers	2
1.1.1 Towards ‘mirror-image life’	3
1.1.2 Racemic protein crystallography	6
1.1.3 Identification of drug candidates through mirror-image phage display and related screening technologies	16
1.2 The current state of the art in D-polypeptide preparation	22
1.2.1 Solid phase peptide synthesis (SPPS)	22
1.2.2 Chemical ligation	22
1.2.3 Transformation from D-polypeptide to mirror-image protein	27
1.3 Conclusion	42
1.4 Thesis aims	43
Chapter 2 : Investigations of the role of tryptophan in membrane-active bacteriocins guided by racemic protein crystallography	44
2.1. Preface	45
2.2. Introduction	45
2.3. Results and discussion	46
2.3.1 Chemical synthesis and antibacterial activity of aureocin A53, lacticin Q and their enantiomers	46
2.3.2 Bacteriocin racemic protein crystallography	49
2.3.3 Effect of tryptophan substitutions on antibacterial activity	50
2.3.4 Investigation of bacteriocin oligomerization and its effect on stability	53
2.4. Conclusion	54
Chapter 3 : Isolated domain of TNFR-1 for the discovery of D-peptide inhibitors via mirror image discovery technologies	56
3.1. Preface	57
3.2. Introduction	57
3.3. Results and discussion	59
3.3.1. Chemical synthesis of TNFR1 cysteine rich domain 2 (TNRC2)	59
3.3.2. Optimization of TNRC2 <i>in vitro</i> protein folding	60
3.3.3. Validation of TNRC2 conformation by racemic protein crystallography	61

3.3.4.	Mirror-image phage display to identify D-peptide ligands for TNRC2D	62
3.3.5.	Identification of the cyclic peptide active isomer	63
3.3.6.	Mirror-image <i>de novo</i> protein design – a case study using TNRC2D	66
3.4.	Conclusion	69
Chapter 4 : Thesis summary and future directions		70
4.1.	Summary	71
4.2.	Further investigation into the mechanisms of bacteriocin activity	73
4.2.1.	Probing the role of tryptophan-lipid interaction	73
4.2.2.	Investigation of active bacteriocin conformations by racemic protein crystallography	75
4.3.	Development of an expedient route to D-peptide binder discovery	75
4.2.1.	Towards a D-peptide inhibitor of TNFR-1	75
4.2.2.	Optimization of mirror image <i>de novo</i> protein design	76
Chapter 5 : Materials and methods		78
5.1.	Materials and instruments	79
5.2.	HPLC, UPLC and ESI mass spec procedures	80
5.3.1.	Peptide LCMS analysis	80
5.3.2.	Peptide UPLC/HDMS analysis	80
5.3.3.	Peptide semi-preparative HPLC	80
5.3.4.	Peptide preparative HPLC	81
5.3.	Synthetic procedures	81
5.3.1.	Manual SPPS	81
5.3.2.	Automated SPPS with low amino acid excess	81
5.3.3.	Automated SPPS with high amino acid excess	82
5.3.4.	Peptide hydrazide preparation	83
5.3.5.	C-terminal biotinylated peptide linker for TNRC2D	83
5.3.6.	Peptide cleavage	84
5.3.7.	One-pot native chemical ligation and desulfurization of bacteriocins	84
5.3.8.	Native chemical ligation of TNRC2D for refolding optimization	85
5.3.9.	Screening of TNRC2D refolding conditions	86
5.3.10.	One-pot native chemical ligation and refolding of TNRC2D	88
5.3.11.	TCPB-E solution-phase oxidation	89
5.3.12.	TCPB-E solid-phase oxidation	89
5.4.	Circular dichroism spectroscopy	90
5.5.	Racemic protein crystallography	90
5.5.1.	AucA crystallization	90
5.5.2.	LnqQ crystallization	91

5.5.3. TNRC2 crystallization	92
5.5.4. X-ray diffraction data collection	92
5.5.5. AucA structure solution	92
5.5.6. LncQ structure solution	93
5.5.7. TNRC2 structure solution	94
5.6. Antibiotic activity assay	94
5.7. Analytical size exclusion chromatography	95
5.8. AucA proteolytic stability assay	95
5.9. Bacteriophage bio-panning	96
5.10. LCMS disulfide bond mapping of TCPB-E's	96
5.11. Grating coupled interferometry	97
5.12. ROSETTA <i>de novo</i> protein binder design	98
Chapter 6 : References	101
Chapter 7 : Appendices	119
7.1. Supplementary figures – Chapter 2	120
7.2. Supplementary figures – Chapter 3	127
7.3. Peptide and protein LCMS data	137
7.3.1. L-aureocin A53 (Met ¹ -Ile ¹⁰) peptide hydrazide	137
7.3.2. L-aureocin A53 (Cys ¹¹ -Leu ⁵¹) N-cysteine peptide	138
7.3.3. D-aureocin A53 (Met ¹ -Ile ¹⁰) peptide hydrazide	139
7.3.4. D-aureocin A53 (Cys ¹¹ -Leu ⁵¹) N-cysteine peptide	140
7.3.5. Example LCMS ligation monitoring of lactacin Q	141
7.3.6. L-aureocin A53 (Met ¹ -Leu ⁵¹)	142
7.3.7. D-aureocin A53 (Met ¹ -Leu ⁵¹)	146
7.3.8. L-aureocin A53 W3L (Met ¹ -Leu ⁵¹)	149
7.3.9. L-aureocin A53 W22L (Met ¹ -Leu ⁵¹)	152
7.3.10. L-aureocin A53 W31L (Met ¹ -Leu ⁵¹)	153
7.3.11. L-aureocin A53 W40L (Met ¹ -Leu ⁵¹)	156
7.3.12. L-aureocin A53 W42L (Met ¹ -Leu ⁵¹)	159
7.3.13. L-aureocin A53 W3E (Met ¹ -Leu ⁵¹)	160
7.3.14. L-aureocin A53 W31E (Met ¹ -Leu ⁵¹)	161
7.3.15. L-aureocin A53 W40E (Met ¹ -Leu ⁵¹)	162
7.3.16. L-lactacin Q (Met ¹ -Trp ²³) peptide hydrazide	163
7.3.17. L-lactacin Q (Cys ²⁴ -Lys ⁵³) N-cysteine peptide	164
7.3.18. D-lactacin Q (Met ¹ -Trp ²³) peptide hydrazide	165
7.3.19. D-lactacin Q (Cys ²⁴ -Lys ⁵³) N-cysteine peptide	166

7.3.20.	Example LCMS ligation monitoring of lacticin Q	167
7.3.21.	L-lacticin Q (Met ¹ -Lys ⁵³)	168
7.3.22.	D-lacticin Q (Met ¹ -Lys ⁵³)	171
7.3.23.	L-lacticin Q W21L (Met ¹ -Lys ⁵³)	174
7.3.24.	L-lacticin Q W23L (Met ¹ -Lys ⁵³)	177
7.3.25.	L-lacticin Q W32L (Met ¹ -Lys ⁵³)	180
7.3.26.	L-lacticin Q W41L (Met ¹ -Lys ⁵³)	183
7.3.27.	L-TNRCD2 (Ser ⁷² -Ser ⁸⁷) peptide hydrazide	186
7.3.28.	D-TNRCD2 (Ser ⁷² -Lys ⁸⁷) peptide hydrazide	186
7.3.29.	L-TNRCD2 (Cys ⁸⁸ -Asn ¹¹⁶) N-cysteine peptide	187
7.3.30.	D-TNRCD2 (Cys ⁸⁸ -Asn ¹¹⁶) N-cysteine peptide	188
7.3.31.	D-TNRCD2 (Cys ⁸⁸ -Asn ¹¹⁶ -Gly ¹¹⁷ -Gly ¹²¹ -Lys(biotin) ¹²²) N-cysteine peptide.....	189
7.3.32.	L-TNRCD2 (Ser ⁷² -Asn ¹¹⁶) reduced (not isolated).....	190
7.3.33.	L-TNRCD2 (Ser ⁷² -Asn ¹¹⁶)	191
7.3.34.	D-TNRCD2 (Ser ⁷² -Asn ¹¹⁶)	192
7.3.35.	D-TNRCD2 (Ser ⁷² -Asn ¹¹⁶ -Gly ¹¹⁷ -Gly ¹²¹ -Lys(biotin) ¹²²).....	193
7.3.36.	L-TCPB-E-C1	194
7.3.37.	L-TCPB-E-C4	195
7.3.38.	L-TCPB-E-C11	196
7.3.39.	L-TCPB-E-G1	197
7.3.40.	L-TCPB-E-G11	198
7.3.41.	L-TMPB-T2 (Arg ¹ -Lys ⁴³).....	199
7.3.42.	L-TMPB-T3 (Arg ¹ -Lys ⁴³).....	200
7.3.43.	L-TMPB-T4 (Arg ¹ -Lys ⁴³).....	201
7.3.44.	L-TMPB-Sc (Arg ¹ -Leu ⁴³)	202
7.4.	Crystallography data refinement statistics	203
7.5.	ROSETTA protein <i>de novo</i> protein design scripts	204
7.5.1.	Example command line for running ROSSETA scripts in Linux	204
7.5.2.	Relax script	204
7.5.3.	Relax flag	205
7.5.4.	GlobalDock script.....	206
7.5.5.	GlobalDock flag.....	208
7.5.6.	FlexDesign script.....	209
7.5.7.	Design flag	215
7.5.8.	FullDock script.....	216
7.5.9.	FullDock flag	219

Table of figures

Figure 1.1 Proteins comprised entirely of D-amino acids fold into the mirror image of the corresponding L-protein.	2
Figure 1.2 Recent advances in realizing mirror-image synthetic biology using D-protein enantiomers. Representative structures for illustrative purposes as D-ASFV pol X (PDB: 1JQE), D-Dpo4 (PDB: 3PR4), D-Pfu (PDB: 2JGU), D-LigA (PDB:2Q2T), L5, L18 and L25 (PDB: 4YBB) constructed in PYMOL. ¹⁶	4
Figure 1.3 Illustrative comparison of (A) a homochiral crystal unit cell and (B) a racemic crystal unit cell containing an inversion center (e.g. <i>P</i> -1).	7
Figure 1.4 Quaternary structures by racemic protein crystallography; (A) Melittin tetramer (PDB: 6O4M); (B) BTD-2 extended fibril-like structure (PDB: 5INZ); (C) Magainin 2 phenylalanine zipper motif unaffected by β -amino acid substitutions. D-magainin 2 is shown in red and mutants L-1 (Ala) shown in blue (PDB: 4MPG), L-2 (APC) in green (PDB: 4CGN) and L-3 (ACPC)(PDB: 5CGO) in magenta. β -amino acids highlighted in orange; (D) M2-TM helix forms heterochiral coiled coils, with a hendecad repeat identified in lipidic cubic phase (LCP, PDB: 4RWB) but absent in racemic β -octylglucoside (DL-OG, PDB: 4RWC). Mutation of sterically disruptive isoleucine residues to alanine (DL-OG(I39A), PDB: 6MPL), (DL-OG(I42A), PDB: 6MPM) or glutamate (DL-OG(I42E), PDB: 6MPN) favored hendecad repeat motifs in chiral lipids; (E) Quaternary structure of VEGF-A dimer bound to two D-protein antagonist molecules (PDB:4GLN). All structures were modelled in CCP4MG ⁶⁶ with data obtained from the Protein Data Bank.	12
Figure 1.5 Quasi-racemic protein crystallography of homogenous, post-translationally modified proteins; (A) D-Ser-CCL facilitated crystallization of glycosylated L-Ser-CCL1 (PDB: 4OIK); (B) Monomeric D-ubiquitin (D-Ub) facilitated crystallization of L-K6-linked diUb (PDB: 5GOB); (C) Quasi-racemic protein crystallography of L- and D-K27-linked diUb (PDB: 5J8P). All structures modelled in CCP4MG ⁶⁶ with data obtained from the protein data bank.	14
Figure 1.6 Process of mirror-image phage display. Natural chirality L-proteins are represented in blue, and mirror-image synthetic D-proteins are represented in red. Purple beads represent protein immobilization.	17
Figure 1.7 Reaction scheme of native chemical ligation.	23
Figure 1.8 (A) Mechanism of metal-free radical desulfurization, (B) acetamidomethyl cysteine (Cys(Acm)) and (C) penicillamine buildings blocks for Fmoc-SPPS.	25

Figure 1.9 Second generation removable backbone modification. 4-methoxy-5-nitrosalicylaldehyde is installed onto backbone nitrogen during Fmoc-SPPS, followed by assembly of peptide main chain and desired tag sequence. Reversible acetylation of phenol group controls TFA lability for tag removal.	26
Figure 1.10: Examples of disulfide-rich D-proteins prepared by chemical synthesis, with structures resolved by racemic protein crystallography.	28
Figure 1.11. Methods for branched protein ligation used in preparation of poly-ubiquitins; (A) glycyI auxilliary mediated NCL for preparation of branched L-proteins and (B) cysteine mediated NCL for preparation of branched D-proteins.	30
Figure 2.1: A) Sequence of AucA and LnqQ, B) Ligation scheme for synthesis of AucA and LnqQ, C) UPLC traces and deconvoluted high-definition ESI+ MS of isolated L- and D- LnqQ and AucA. ⁿ denotes the residue number of the Ala→Cys substitution used to facilitate ligation.	47
Figure 2.2: Insights from racemic protein crystallography of AucA; a-c) coordination of solvent exposed W31, W3 and W40 with sulfate (PDB: 8AVR) and d) coordination of W31 with glycerol 3-phosphate (G3P) (PDB: 8AVT); and comparison of e) dimeric and f) monomeric AucA. (PDB: 8AVU and 8AVS).	50
Figure 2.3: Analytical HP-SEC showing upfield retention of AucA W3L with respect to wild-type, dimeric AucA, and similar retention to monomeric LnqQ. See Figure S6 for extended analysis.	54
Figure 3.1: Interaction of TNFR-1 with TNF- α is significantly governed by the cysteine rich domain 2 (TNRC2, magenta), serving as a good target for mirror-image phage display (MIPD).	58
Figure 3.2: The use of TNFR-1 cysteine rich domain 2 (TNRC2) as a target in mirror-image phage display (MIPD) to identify D-peptide inhibitors.	59
Figure 3.3: A) amino acid sequence of TNRC2, numbered based on full extracellular domain, B) synthetic scheme to prepare TNRC2, C) LCMS analysis of TNRC2 intermediate and enantiomer products, D) mirror image folding of L- and D- TNCRD2, E) σ A-weighted $2F_o-F_c$ omit electron density maps contoured at 1σ (0.46 electrons per \AA^3) showing correct disulfide bond formation, F) structural overlay of the L-TNRC2 structure with TNF- α /TNFR-1 complex (PDB: 1TNR).	60
Figure 3.4: Sequence consensus of the enriched peptide library following the second biopanning experiment, using a CX ₄ WLGX ₂ C library format.	62
Figure 3.5: Six possible conformers of TCPB-E, oxidized with either glutathione (TCPB-E-G) or cysteine (TCPB-E-C).	63

Figure 3.6: Solution-phase oxidation of reduced TCPB-E peptide (Cystine 0.5 mM) and isolation of each tautomer TCPB-E-Ci (top) and -Cii (bottom). Chymotrypsin digestion of the isolated oxidized peptide followed by LCMS analysis allowed assignment of the present conformer...	64
Figure 3.7: Multi-cycle kinetic analysis of L-TCPB-E-C1 binding to D-TNRCD2 (left), and D-TCPB-E-C1 binding to L-TNRCD2 (right) using grating-coupled interferometry (GCI). Equilibrium binding constant K_D is indicated in insets, using data fitted with a 1:1 Langmuir model (black lines).....	65
Figure 3.8: Proposed method of mirror-image <i>de novo</i> protein binder design	66
Figure 3.9: Mini protein scaffolds used for binder design in this work, denoting the PDB ID's and secondary structure compositions (H = α -helix, E = β -sheet).	67
Figure 3.10: The top 20 poses of 5UOI docked to D-TNRCD2, used in the subsequent scaffold binding interface design. The pose giving rise to the best candidate binders (TMPB-TX) is highlighted.	67
Figure 3.11: Candidate mini protein binders for the D-TNRCD2 target. A) Structure of designed TMPB-T4 bound to D-TNRCD2, B) interactions of the C-terminal tail of TMPB-T4 with D-TNRCD2, and C) sequence alignment of the three candidate binders (TMPB-T2, -T3, and -T4), a low scoring design TMPB-TL, and the undesigned scaffold protein (TMPB-Sc).....	68
Figure 4.1: Tryptophan derivatives for incorporation into AucA and LnqQ, allowing further investigations into the role of the indole hydrogen bond (available from Aralez Bio ²⁴¹).	74
Figure S7.1: Positions of AucA Ala11 and Ala21 in the folded state, shown as green van der Waals (VDW) spheres.	120
Figure S7.2: Circular dichroism spectra of AucA and variants.....	121
Figure S7.3: Circular dichroism spectra of LnqQ and variants.....	122
Figure S7.4: Reorientation of AucA W22 between dimeric and monomeric forms. Overlay of AucA (SO ₄ ²⁻) (PDB: 8AVR, Cyan) and AucA (dimer) (PDB: 8AVU, Magenta) chain A showing reorientation of Trp22 (Bold).	123
Figure S7.5: Analysis of AucA dimeric interface (PDB: 8AVU) using Protein Interfaces, Surfaces and Assemblies (PISA) software ²⁰⁷ in the CCP4 software suite.	123
Figure S7.6: Analytical HP-SEC of AucA and Trp→Leu variants. Gel filtration standard shown (top) was used to create calibration plot (top right), with approximate calculated MW shown along the X-axis. LnqQ is used as monomeric bacteriocin reference.	124
Figure S7.7: AucA proteolytic stability assay. LC traces at UV210nm of protease reaction quenched after 24 hours incubation with proteinase K (+ProK) at 37 °C, and samples prepared and incubated in the same manner without addition of protease (blank). HPLC conducted using	

a 0-99% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm).	125
Figure S7.8: Solution-phase oxidation of TCPB-E with cystine (left) or glutathione disulfide (right) for deduction of active bacteriophage conformation. See below for identification of peaks by disulfide bond mapping experiments. Peak corresponding to dimerized product with three disulfide bonds indicate by asterisk (*).	130
Figure S7.9: LCMS disulfide bond mapping of isolated TCPB-E peak ii, oxidized with cystine. Lyophilized peptide was digested with chymotypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 6 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peak is shown in bottom left. Identity of the major digested fragment and the deduced, undigested conformer are shown to the right. No digestion was observed between Phe2 and His3 after 6 hours.	130
Figure S7.10: LCMS disulfide bond mapping of isolated TCPB-E peak i, oxidized with cystine. Lyophilized peptide was digested with chymotypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 6 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peaks are shown in bottom left. Identity of the major digested fragments and the deduced, undigested conformer are shown to the right.	131
Figure S7.11: LCMS disulfide bond mapping of isolated TCPB-E peak ii, oxidized with glutathione disulfide. Lyophilized peptide was digested with chymotypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 6 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peak is shown in bottom left. Identity of the major digested fragment and the deduced, undigested conformer are shown to the right. No digestion was observed between Phe2 and His3 after 6 hours.	131
Figure S7.12: LCMS disulfide bond mapping of isolated TCPB-E peak i, oxidized with glutathione disulfide. Lyophilized peptide was digested with chymotypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 12 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peaks are shown in bottom left. Identity of the major digested fragments and the deduced, undigested conformer are shown to the right.	132
Figure S7.13: Grating-coupled interferometry (GCI) analysis of TCPB-E isomers binding to D-TNRCD2. Repeated analyte pulses of increasing duration (RAPID) were passed over the sensor surface, with fixed concentrations of each peptide conformer shown above. Sensograms shown	

are blank subtracted (target flow cell - flow cell with no target). The highest sensor response with respect to concentration was observed for TCPB-E-C1.	132
Figure S7.14: Illustrative representation of a solid-phase synthetic route to TCPB-E peptide conformers. An orthogonal protection scheme was implemented using cysteine acetamidomethyl (acm) or trityl (trt). Disulfide bond formation was controlled by the relative reactivity of iodine towards protected cysteine residues in non-polar (trt-trt) and polar (acm-acm) solvents, based on a modified procedure. ²²⁷	133
Figure S7.15: Comparison of HPLC retention times of TCPB-E-C1 and -C4 prepared on solid-phase with the solution-phase oxidation of TCPB-E with cystine, reinforcing the results of LCMS disulfide bond mapping. Small quantities of TCPB-E-C4 appear to have been formed in solution-phase but was insufficient for isolation, likely suggesting that this conformer is thermodynamically unfavourable.	133
Figure S7.16: Grating-coupled interferometry (GCI) analysis of TCPB-E-C1 and -C4 isomers binding to D-TNRCD2. A single cycle of association and dissociation were passed over the sensor surface, with 50 μ M of each peptide conformer. Sensograms shown are blank subtracted (target flow cell - flow cell with no target). The highest sensor response with was observed for TCPB-E-C1, with little binding observed for TCPB-E-C4.....	134
Figure S7.17: RosettaDock results of <i>de novo</i> mini protein docking to the D-TNRCD2, showing: A) undesigned scaffold (TMPB-Sc), B) low scoring protein design (TMPB-TL), and C-E) binder candidates (TMPB-T2, -T3 and -T4). Plots of the estimated binding energy of the docked pose (Ddg) versus the deviation (RMSD) of the docked structure to the initial design pose are shown on the left. The lowest Ddg docked structure overlaid with the initial design pose is shown on the right, as a visualization of the single lowest energy point on the plots (highlighted by purple circle). Proteins with low Ddg docked poses and high RMSD to the initial design pose are characterized as poor binders <i>in silico</i>	135
Figure S7.18: Circular dichroism (CD) spectra of designed proteins TMPB-T2, -T3 and -T4 compared to the undesigned scaffold protein TMPB-S, showed a lack of alpha-helical fold following <i>in silico</i> protein design.	136
Figure S7.19: LC trace at UV210nm (left) of isolated L-aureocin A53 (Met ¹ -Ile ¹⁰) peptide hydrazide using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 μ m). Right - (ESI-MS (m/z): calculated 665.3 [M+2H] ²⁺ , 443.9 [M+3H] ³⁺ , observed 665.0 [M+2H] ²⁺ , 443.8 [M+3H] ³⁺).	137
Figure S7.20: LC trace at UV210nm (left) of isolated L-aureocin A53 (Cys ¹¹ -Leu ⁵¹) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfinityLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 μ m). Right - (ESI-MS (m/z): calculated 1180.7 [M+4H] ⁴⁺ ,	

944.7 [M+5H] ⁵⁺ , 787.4 [M+6H] ⁶⁺ , 675.1 [M+7H] ⁷⁺ , 590.8 [M+8H] ⁸⁺ , 525.3 [M+9H] ⁹⁺ , observed 1180.7 [M+4H] ⁴⁺ , 944.6 [M+5H] ⁵⁺ , 787.4 [M+6H] ⁶⁺ , 675.0 [M+7H] ⁷⁺ , 590.8 [M+8H] ⁸⁺ , 525.3 [M+9H] ⁹⁺).	138
Figure S7.21: LC trace at UV210nm (left) of isolated D-aureocin A53 (Met ¹ -Ile ¹⁰) peptide hydrazide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfinityLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 665.3 [M+2H] ²⁺ , 443.9 [M+3H] ³⁺ , observed 665.0 [M+2H] ²⁺ , 443.8 [M+3H] ³⁺).	139
Figure S7.22: LC trace at UV210nm (left) of isolated D-aureocin A53 (Cys ¹¹ -Leu ⁵¹) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfinityLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 1572.9 [M+3H] ³⁺ , 1180.7 [M+4H] ⁴⁺ , 944.7 [M+5H] ⁵⁺ , 787.4 [M+6H] ⁶⁺ , 675.1 [M+7H] ⁷⁺ , observed 1574.2 [M+3H] ³⁺ , 1180.7 [M+4H] ⁴⁺ , 944.6 [M+5H] ⁵⁺ , 787.4 [M+6H] ⁶⁺ , 675.0 [M+7H] ⁷⁺).	140
Figure S7.23: Example LCMS ligation scheme for AucA. a) AucA amino acid sequence, b) one-pot ligation-desulurization scheme, c) HPLC reaction monitoring, and d) ESI+ MS for AucA product, reactants and intermediates.....	141
Figure S7.24: LC trace at UV210nm (left) of isolated L-aureocin A53 (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1197.6 [M+5H] ⁵⁺ , 998.2 [M+6H] ⁶⁺ , 855.7 [M+7H] ⁷⁺ , 748.9 [M+8H] ⁸⁺ , 665.8 [M+9H] ⁹⁺ , 599.3 [M+10H] ¹⁰⁺ , observed 1197.4 [M+5H] ⁵⁺ , 998.2 [M+6H] ⁶⁺ , 855.7 [M+7H] ⁷⁺ , 748.9 [M+8H] ⁸⁺ , 665.8 [M+9H] ⁹⁺ , 599.3 [M+10H] ¹⁰⁺).	142
Figure S7.25: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 (Met ¹ -Leu ⁵¹) [19.47 min].	143
Figure S7.26: High-definition ESI+ mass spec of isolated L-aureocin A53 (Met ¹ -Leu ⁵¹), (ESI-MS (m/z): calculated 1197.6 [M+5H] ⁵⁺ , 998.2 [M+6H] ⁶⁺ , 855.7 [M+7H] ⁷⁺ , 748.9 [M+8H] ⁸⁺ , 665.8 [M+9H] ⁹⁺ , observed 1197.5 [M+5H] ⁵⁺ , 998.0 [M+6H] ⁶⁺ , 855.6 [M+7H] ⁷⁺ , 748.9 [M+8H] ⁸⁺ , 665.7 [M+9H] ⁹⁺ , deconvoluted: calculated 5983.2, observed 5983.5).	144
Figure S7.27: High-definition ESI+ mass spec of isolated L-aureocin A53 (Met ¹ -Leu ⁵¹), (ESI-MS (m/z) deconvoluted: calculated 5983.2, observed 5983.5).	145
Figure S7.28: LC trace at UV210nm (left) of isolated D-aureocin A53 (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfintiyLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 1197.6 [M+5H] ⁵⁺ , 998.2 [M+6H] ⁶⁺ , 855.7 [M+7H] ⁷⁺ , 748.9 [M+8H] ⁸⁺ , 665.8 [M+9H] ⁹⁺ , observed 1197.5 [M+5H] ⁵⁺ , 998.0 [M+6H] ⁶⁺ , 855.7 [M+7H] ⁷⁺ , 748.8 [M+8H] ⁸⁺ , 665.8 [M+9H] ⁹⁺).	146
Figure S7.29: UPLC total ion chromatogram in ESI+ mode of isolated D-aureocin A53 (Met ¹ -Leu ⁵¹) [20.57 min].	147

Figure S7.30: High-definition ESI+ mass spec of isolated D-aureocin A53 (Met ¹ -Leu ⁵¹), (ESI-MS (m/z): calculated 1197.6 [M+5H] ⁵⁺ , 998.2 [M+6H] ⁶⁺ , 855.7 [M+7H] ⁷⁺ , 748.9 [M+8H] ⁸⁺ , 665.8 [M+9H] ⁹⁺ , observed 1197.5 [M+5H] ⁵⁺ , 998.1 [M+6H] ⁶⁺ , 855.6 [M+7H] ⁷⁺ , 748.8 [M+8H] ⁸⁺ , 665.7 [M+9H] ⁹⁺ , deconvoluted: calculated 5983.2, observed 5983.0).	148
Figure 7.31: LC trace at UV210nm (left) of isolated L-aureocin A53 W3L (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.8 [M+5H] ⁵⁺ , 985.8 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺).	149
Figure S7.32: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 W3L (Met ¹ -Leu ⁵¹) [18.78 min].	150
Figure S7.33: High-definition ESI+ mass spec of isolated L-aureocin A53 W3L (Met ¹ -Leu ⁵¹), (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.9 [M+5H] ⁵⁺ , 985.9 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.6 [M+9H] ⁹⁺ , 591.9 [M+10H] ¹⁰⁺ , deconvoluted: calculated 5910.2, observed 5909.5).	151
Figure S7.34: LC trace at UV210nm (left) of isolated L-aureocin A53 W22L (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.7 [M+5H] ⁵⁺ , 985.8 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺).	152
Figure S7.35: LC trace at UV210nm (left) of isolated L-aureocin A53 W31L (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 986.0 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.1 [M+10H] ¹⁰⁺).	153
Figure S7.36: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 W31L (Met ¹ -Leu ⁵¹) [18.77 min].	154
Figure S7.37: High-definition ESI+ mass spec of isolated L-aureocin A53 W31L (Met ¹ -Leu ⁵¹), (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.9 [M+5H] ⁵⁺ , 985.9 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.6 [M+9H] ⁹⁺ , 591.9 [M+10H] ¹⁰⁺ , deconvoluted: calculated 5910.2, observed 5909.5).	155
Figure S7.38: LC trace at UV210nm (left) of isolated L-aureocin A53 W40L (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ ,	

739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.7 [M+5H] ⁵⁺ , 985.8 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺).	156
Figure S7.39: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 W40L (Met ¹ -Leu ⁵¹) [18.50 min].	157
Figure S7.40: High-definition ESI+ mass spec of isolated L-aureocin A53 W40L (Met ¹ -Leu ⁵¹), (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.9 [M+5H] ⁵⁺ , 985.9 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.6 [M+9H] ⁹⁺ , 591.9 [M+10H] ¹⁰⁺⁺ , deconvoluted: calculated 5910.2, observed 5909.5).	158
Figure S7.41: LC trace at UV210nm (left) of isolated L-aureocin A53 W42L (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H] ⁵⁺ , 986.0 [M+6H] ⁶⁺ , 845.3 [M+7H] ⁷⁺ , 739.8 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺ , observed 1182.8 [M+5H] ⁵⁺ , 985.8 [M+6H] ⁶⁺ , 845.2 [M+7H] ⁷⁺ , 739.7 [M+8H] ⁸⁺ , 657.7 [M+9H] ⁹⁺ , 592.0 [M+10H] ¹⁰⁺).	159
Figure S7.42: LC trace at UV210nm (left) of isolated L-aureocin A53 W3E (Met ¹ -Leu ⁵¹) using a 20-90% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 988.7 [M+6H] ⁶⁺ , 847.6 [M+7H] ⁷⁺ , 741.8 [M+8H] ⁸⁺ , 659.5 [M+9H] ⁹⁺ , 593.6 [M+10H] ¹⁰⁺ , observed 988.6 [M+6H] ⁶⁺ , 847.4 [M+7H] ⁷⁺ , 741.7 [M+8H] ⁸⁺ , 659.5 [M+9H] ⁹⁺ , 593.6 [M+10H] ¹⁰⁺).	160
Figure S7.43: LC trace at UV210nm (left) of isolated L-aureocin A53 W31E (Met ¹ -Leu ⁵¹) using a 5-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 988.7 [M+6H] ⁶⁺ , 847.6 [M+7H] ⁷⁺ , 741.8 [M+8H] ⁸⁺ , 659.5 [M+9H] ⁹⁺ , 593.6 [M+10H] ¹⁰⁺ , observed 988.5 [M+6H] ⁶⁺ , 847.5 [M+7H] ⁷⁺ , 741.8 [M+8H] ⁸⁺ , 659.4 [M+9H] ⁹⁺ , 593.5 [M+10H] ¹⁰⁺).	161
Figure S7.44: LC trace at UV210nm (left) of isolated L-aureocin A53 W40E (Met ¹ -Leu ⁵¹) using a 20-90% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 988.7 [M+6H] ⁶⁺ , 847.6 [M+7H] ⁷⁺ , 741.8 [M+8H] ⁸⁺ , 659.5 [M+9H] ⁹⁺ , 593.6 [M+10H] ¹⁰⁺ , observed 988.5 [M+6H] ⁶⁺ , 847.5 [M+7H] ⁷⁺ , 741.7 [M+8H] ⁸⁺ , 659.5 [M+9H] ⁹⁺ , 593.5 [M+10H] ¹⁰⁺).	162
Figure S7.45: LC trace at UV210nm (left) of isolated L-lacticin Q (Met ¹ -Trp ²³) peptide hydrazide using a 10-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 870.7 [M+3H] ³⁺ , 653.3 [M+4H] ⁴⁺ , 522.8 [M+5H] ⁵⁺ , observed 870.6 [M+3H] ³⁺ , 653.2 [M+4H] ⁴⁺ , 522.8 [M+5H] ⁵⁺).	163
Figure S7.46: LC trace at UV210nm (left) of isolated L-lacticin Q (Cys ²⁴ -Lys ⁵³) N-cysteine peptide using a 10-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 839.0	

[M+4H] ⁴⁺ , 671.4 [M+5H] ⁵⁺ , 559.7 [M+6H] ⁶⁺ , observed 838.8 [M+4H] ⁴⁺ , 671.3 [M+5H] ⁵⁺ , 559.6 [M+6H] ⁶⁺).	164
Figure S7.47: LC trace at UV210nm (left) of isolated D-lactacin Q (Met ¹ -Trp ²³) peptide hydrazide using a 20-80% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 870.7 [M+3H] ³⁺ , 653.3 [M+4H] ⁴⁺ , 522.8 [M+5H] ⁵⁺ , observed 870.7 [M+3H] ³⁺ , 653.2 [M+4H] ⁴⁺ , 522.8 [M+5H] ⁵⁺).	165
Figure S7.48: LC trace at UV210nm (left) of isolated D-lactacin Q (Cys ²⁴ -Lys ⁵³) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfintiyLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 839.0 [M+4H] ⁴⁺ , 671.4 [M+5H] ⁵⁺ , 559.7 [M+6H] ⁶⁺ , observed 838.8 [M+4H] ⁴⁺ , 671.3 [M+5H] ⁵⁺ , 559.7 [M+6H] ⁶⁺).	166
Figure S7.49: Example LCMS ligation scheme for LnqQ. a) LnqQ amino acid sequence, b) one-pot ligation-desulfurization scheme, c) HPLC reaction monitoring, and d) ESI+ MS for LnqQ product, reactants and intermediates.....	167
Figure S7.50: LC trace at UV210nm (left) of isolated L-lactacin Q (Met ¹ -Lys ⁵³) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated *990.0 [M+5H+K] ⁶⁺ , 983.8 [M+6H] ⁶⁺ , *848.7 [M+6H+K] ⁷⁺ , 843.4 [M+7H] ⁷⁺ , *742.8 [M+7H+K] ⁸⁺ , 738.1 [M+8H] ⁸⁺ , 656.2 [M+9H] ⁹⁺ , observed *990.2 [M+5H+K] ⁶⁺ , 983.7 [M+6H] ⁶⁺ , *848.7 [M+6H+K] ⁷⁺ , 843.3 [M+7H] ⁷⁺ , *742.8 [M+5H+K] ⁶⁺ , 738.1 [M+8H] ⁸⁺ , 656.2 [M+9H] ⁹⁺).	168
Figure S7.51: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q (Met ¹ -Lys ⁵³) [22.30 min].....	169
Figure S7.52: High-definition ESI+ mass spec of isolated L-lactacin Q (Met ¹ -Lys ⁵³) (left), (ESI-MS (m/z): calculated 1180.4 [M+5H] ⁵⁺ , 983.8 [M+6H] ⁶⁺ , 843.4 [M+7H] ⁷⁺ , 738.1 [M+8H] ⁸⁺ , observed 1180.3 [M+5H] ⁵⁺ , 983.7 [M+6H] ⁶⁺ , 843.3 [M+7H] ⁷⁺ , 738.1 [M+8H] ⁸⁺). (Right – deconvoluted mass), calculated 5897.1, observed 5898.0.	170
Figure S7.53: LC trace at UV210nm (left) of isolated D-lactacin Q (Met ¹ -Lys ⁵³) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated *990.0 [M+5H+K] ⁶⁺ , 983.8 [M+6H] ⁶⁺ , *848.7 [M+6H+K] ⁷⁺ , 843.4 [M+7H] ⁷⁺ , *742.8 [M+7H+K] ⁸⁺ , 738.1 [M+8H] ⁸⁺ , 656.2 [M+9H] ⁹⁺ , observed *990.0 [M+5H+K] ⁶⁺ , 983.7 [M+6H] ⁶⁺ , *848.8 [M+6H+K] ⁷⁺ , 843.5 [M+7H] ⁷⁺ , *742.6 [M+5H+K] ⁶⁺ , 738.0 [M+8H] ⁸⁺ , 656.2 [M+9H] ⁹⁺).	171
Figure S7.54: UPLC total ion chromatogram in ESI+ mode of isolated D-lactacin Q (Met ¹ -Lys ⁵³) [22.40 min].....	172

Figure S7.55: High-definition ESI+ mass spec of isolated D-lactacin Q (Met ¹ -Lys ⁵³) (left), (ESI-MS (m/z): calculated 1180.4 [M+5H] ⁵⁺ , 983.8 [M+6H] ⁶⁺ , 843.4 [M+7H] ⁷⁺ , 738.1 [M+8H] ⁸⁺ , , 656.2 [M+9H] ⁹⁺ , observed 1180.3 [M+5H] ⁵⁺ , 983.7 [M+6H] ⁶⁺ , 843.3 [M+7H] ⁷⁺ , 738.2 [M+8H] ⁸⁺ , 656.2 [M+9H] ⁹⁺). (Right – deconvoluted mass), calculated 5897.1, observed 5896.5.	173
Figure S7.56: LC trace at UV210nm (left) of isolated L-lactacin Q W21L (Met ¹ -Lys ⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.3 [M+5H] ⁵⁺ , 971.3 [M+6H] ⁶⁺ , 832.8 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.0 [M+9H] ⁹⁺).	174
Figure S7.57: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W21L (Met ¹ -Lys ⁵³) [32.85 min].....	175
Figure S7.58: High-definition ESI+ mass spec of isolated L-lactacin Q W21L (Met ¹ -Lys ⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.7 [M+5H] ⁵⁺ , 971.6 [M+6H] ⁶⁺ , 832.9 [M+7H] ⁷⁺ , 728.9 [M+8H] ⁸⁺ , 648.0 [M+9H] ⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5824.0.	176
Figure S7.59: LC trace at UV210nm (left) of isolated L-lactacin Q W23L (Met ¹ -Lys ⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.5 [M+5H] ⁵⁺ , 971.5 [M+6H] ⁶⁺ , 832.9 [M+7H] ⁷⁺ , 728.9 [M+8H] ⁸⁺ , 648.0 [M+9H] ⁹⁺).	177
Figure S7.60: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W23L (Met ¹ -Lys ⁵³) [32.75 min].....	178
Figure S7.61: High-definition ESI+ mass spec of isolated L-lactacin Q W23L (Met ¹ -Lys ⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.7 [M+5H] ⁵⁺ , 971.6 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5825.0.	179
Figure S7.62: LC trace at UV210nm (left) of isolated L-lactacin Q W32L (Met ¹ -Lys ⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.6 [M+5H] ⁵⁺ , 971.5 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 728.9 [M+8H] ⁸⁺ , 648.2 [M+9H] ⁹⁺).	180
Figure S7.63: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W32L (Met ¹ -Lys ⁵³) [23.96 min] using a 5-95% gradient of A/B over 30 minutes.....	181
Figure S7.64: High-definition ESI+ mass spec of isolated L-lactacin Q W32L (Met ¹ -Lys ⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.7 [M+5H] ⁵⁺ , 971.6 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.1 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5825.0.	182

Figure S7.65: LC trace at UV210nm (left) of isolated L-lactacin Q W41L (Met ¹ -Lys ⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 832.8 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺).	183
Figure S7.66: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W41L (Met ¹ -Lys ⁵³) [32.53 min].....	184
Figure S7.67: High-definition ESI+ mass spec of isolated L-lactacin Q W41L (Met ¹ -Lys ⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H] ⁵⁺ , 971.7 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺ , observed 1165.7 [M+5H] ⁵⁺ , 971.6 [M+6H] ⁶⁺ , 833.0 [M+7H] ⁷⁺ , 729.0 [M+8H] ⁸⁺ , 648.1 [M+9H] ⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5825.0.	185
Figure S7.68: LC trace at UV210nm (left) of isolated L-TNRCD2 (Ser ⁷² -Ser ⁸⁷) peptide hydrazide using a 0-60% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 895.0 [M+2H] ²⁺ , 597.0 [M+3H] ³⁺ , 448.0 [M+4H] ⁴⁺ , 894.5 [M+2H] ²⁺ , 596.2 [M+3H] ³⁺ , 447.4 [M+4H] ⁴⁺).	186
Figure S7.69: LC trace at UV210nm (left) of isolated L-TNRCD2 (Ser ⁷² -Ser ⁸⁷) peptide hydrazide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 895.0 [M+2H] ²⁺ , 597.0 [M+3H] ³⁺ , 448.0 [M+4H] ⁴⁺ , 894.5 [M+2H] ²⁺ , 596.2 [M+3H] ³⁺ , 447.4 [M+4H] ⁴⁺).	186
Figure S7.70: LC trace at UV210nm (left) of isolated L-TNRCD2 (Cys ⁸⁸ -Asn ¹¹⁶) N-cysteine peptide using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1196.3 [M+3H] ³⁺ , 897.5 [M+4H] ⁴⁺ , 718.2 [M+5H] ⁵⁺ , 598.7 [M+6H] ⁶⁺ , observed 1195.7 [M+3H] ³⁺ , 897.3 [M+4H] ⁴⁺ , 718.0 [M+5H] ⁵⁺ , 598.6 [M+6H] ⁶⁺).	187
Figure S7.71: LC trace at UV210nm (left) of isolated D-TNRCD2 (Cys ⁸⁸ -Asn ¹¹⁶) N-cysteine peptide using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 897.5 [M+4H] ⁴⁺ , 718.2 [M+5H] ⁵⁺ , 598.7 [M+6H] ⁶⁺ , observed 897.3 [M+4H] ⁴⁺ , 718.0 [M+5H] ⁵⁺ , 598.5 [M+6H] ⁶⁺).	188
Figure S7.72: LC trace at UV210nm (left) of isolated D-TNRCD2 (Cys ⁸⁸ -Asn ¹¹⁶ -Gly ¹¹⁷ -Gly ¹²¹ -Lys(biotin) ¹²²) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1072.3 [M+4H] ⁴⁺ , 858.1 [M+5H] ⁵⁺ , 715.2 [M+6H] ⁶⁺ , observed 1071.9 [M+4H] ⁴⁺ , 858.0 [M+5H] ⁵⁺ , 715.3 [M+6H] ⁶⁺).	189
Figure S7.73: LC trace at UV210nm (left) of reduced L-TNRCD2 (Ser ⁷² -Asn ¹¹⁶) following native chemical ligation (without isolation). Right - (ESI-MS (m/z): calculated 897.5 [M+4H] ⁴⁺ , 718.2 [M+5H] ⁵⁺ , 598.7 [M+6H] ⁶⁺ , observed 897.3 [M+4H] ⁴⁺ , 718.0 [M+5H] ⁵⁺ , 598.5 [M+6H] ⁶⁺).	190

Figure S7.74: LC trace at UV210nm (left) of isolated L-TNRCD2 (Ser ⁷² -Asn ¹¹⁶). Right - (ESI-MS (m/z): calculated 1334.5 [M+4H] ⁴⁺ , 1067.8 [M+5H] ⁵⁺ , 890.0 [M+6H] ⁶⁺ , 763.0 [M+7H] ⁷⁺ , 667.7 [M+8H] ⁸⁺ , observed 1334.6 [M+4H] ⁴⁺ , 1067.7 [M+5H] ⁵⁺ , 889.8 [M+6H] ⁶⁺ , 762.9 [M+7H] ⁷⁺ , 667.5 [M+8H] ⁸⁺).	191
Figure S7.75: LC trace at UV210nm (left) of isolated D-TNRCD2 (Ser ⁷² -Asn ¹¹⁶) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1334.5 [M+4H] ⁴⁺ , 1067.8 [M+5H] ⁵⁺ , 890.0 [M+6H] ⁶⁺ , 763.0 [M+7H] ⁷⁺ , 667.7 [M+8H] ⁸⁺ , observed 1334.2 [M+4H] ⁴⁺ , 1067.6 [M+5H] ⁵⁺ , 889.8 [M+6H] ⁶⁺ , 762.9 [M+7H] ⁷⁺ , 667.7 [M+8H] ⁸⁺).	192
Figure S7.76: LC trace at UV210nm (left) of isolated D-TNRCD2 (Ser ⁷² -Asn ¹¹⁶ -Gly ¹¹⁷ -Gly ¹²¹ -Lys(biotin) ¹²²) using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1207.7 [M+5H] ⁵⁺ , 1006.6 [M+6H] ⁶⁺ , 862.9 [M+7H] ⁷⁺ , 755.2 [M+8H] ⁸⁺ , observed 1207.9 [M+5H] ⁵⁺ , 1006.5 [M+6H] ⁶⁺ , 862.7 [M+7H] ⁷⁺ , 755.2 [M+8H] ⁸⁺).	193
Figure S7.77: LC trace at UV210nm (left) of L-TCPB-E-C1 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 [M+2H] ²⁺ , 481.9 [M+3H] ³⁺ , observed 721.4 [M+2H] ²⁺ , 481.4 [M+3H] ³⁺).	194
Figure S7.78: LC trace at UV210nm (left) of isolated L-TCPB-E-C1 (prepared by solid-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 [M+2H] ²⁺ , 481.9 [M+3H] ³⁺ , observed 721.9 [M+2H] ²⁺ , 481.4 [M+3H] ³⁺).	194
Figure S7.79: LC trace at UV210nm (left) of isolated L-TCPB-E-C4 (prepared by solid-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 [M+2H] ²⁺ , 481.9 [M+3H] ³⁺ , observed 721.4 [M+2H] ²⁺ , 481.4 [M+3H] ³⁺).	195
Figure S7.80: LC trace at UV210nm (left) of L-TCPB-E-C11 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 [M+2H] ²⁺ , 481.9 [M+3H] ³⁺ , observed 721.4 [M+2H] ²⁺ , 481.4 [M+3H] ³⁺).	196
Figure S7.81: LC trace at UV210nm (left) of L-TCPB-E-G1 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 814.7 [M+2H] ²⁺ , 543.5 [M+3H] ³⁺ , observed 814.4 [M+2H] ²⁺ , 543.4 [M+3H] ³⁺).	197

Figure S7.82: LC trace at UV210nm (left) of L-TCPB-E-G11 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 814.7 [M+2H] ²⁺ , 543.5 [M+3H] ⁺³ , observed 814.8 [M+2H] ²⁺ , 543.4 [M+3H] ⁺³).	198
Figure S7.83: LC trace at UV210nm (left) of isolated L-TMPB-T2 using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1007.3 [M+5H] ⁵⁺ , 839.6 [M+6H] ⁶⁺ , 719.8 [M+7H] ⁷⁺ , 629.9 [M+8H] ⁸⁺ , 560.1 [M+9H] ⁺⁹ , observed 1007.2 [M+5H] ⁵⁺ , 839.5 [M+6H] ⁶⁺ , 719.8 [M+7H] ⁷⁺ , 629.9 [M+8H] ⁸⁺ , 560.1 [M+9H] ⁺⁹).	199
Figure S7.84: LC trace at UV210nm (left) of isolated L-TMPB-T3 using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1011.9 [M+5H] ⁵⁺ , 843.4 [M+6H] ⁶⁺ , 723.1 [M+7H] ⁷⁺ , 632.8 [M+8H] ⁸⁺ , 562.6 [M+9H] ⁺⁹ , observed 1011.7 [M+5H] ⁵⁺ , 843.3 [M+6H] ⁶⁺ , 723.0 [M+7H] ⁷⁺ , 632.8 [M+8H] ⁸⁺ , 562.6 [M+9H] ⁺⁹).	200
Figure S7.85: LC trace at UV210nm (left) of isolated L-TMPB-T4 using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1000.9 [M+5H] ⁵⁺ , 834.3 [M+6H] ⁶⁺ , 715.2 [M+7H] ⁷⁺ , 625.9 [M+8H] ⁸⁺ , 556.5 [M+9H] ⁺⁹ , observed 1000.8 [M+5H] ⁵⁺ , 834.2 [M+6H] ⁶⁺ , 715.2 [M+7H] ⁷⁺ , 626.0 [M+8H] ⁸⁺ , 556.5 [M+9H] ⁺⁹).	201
Figure S7.86: LC trace at UV210nm (left) of isolated L-TMPB-Sc using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1036.8 [M+5H] ⁵⁺ , 864.1 [M+6H] ⁶⁺ , 740.8 [M+7H] ⁷⁺ , 648.3 [M+8H] ⁸⁺ , 576.4 [M+9H] ⁺⁹ , observed 1036.6 [M+5H] ⁵⁺ , 864.0 [M+6H] ⁶⁺ , 740.8 [M+7H] ⁷⁺ , 648.4 [M+8H] ⁸⁺ , 576.5 [M+9H] ⁺⁹).	202

Table of tables

Table 1.1 A decade of racemic protein crystallography (at the time of writing).	8
Table 1.2 D-peptide binders identified through mirror-image phage display of D-protein targets in the last decade. (at the time of writing).	18
Table 1.3. Synthetic methods used in chemical protein synthesis, indicating use in reported D-protein synthesis and potential issues encountered with use.	32
Table 2.1: Minimum inhibitory concentrations ^[a] (MIC) of L- and D-bacteriocins and melittin control against bacterial strains.	48
Table 2.2: Minimum inhibitory concentrations ^[a] (MIC) of bacteriocin variants and control antibiotic agents against bacterial strains.	51
Table 5.1: Variables in Methods A and B for one-pot ligation desulfurization of AucA.	85
Table S7.1: Multiple sequence alignment (MSA) of homologous bacteriocins ($\geq 34\%$ homology to AucA) aligned using Clustal Omega and constructed in Jalview, showing conservation of tryptophan (red) and lysine (blue) residues, along with other potential H-bond donor residues arginine (green) and glutamine (yellow).	126
Table S7.2: Conditions used during screening of TNRC2 refolding. First screening oxidation method (GSSG/GSH best), then buffer pH and temperature (0.1 M NaPi, pH 6.5, 25 °C best), then finally reagent concentrations (all worse than previous). The conditions with the highest yield of folded TNRC2 are highlighted.	127
Table S7.3: Results of bacteriophage biopanning experiment 1 with the sequence format CX ₉ C. Enrichment factor for each round is defined as the amount of phage eluted from the target well containing D-TNRC2 divided by the amount of phage eluted from the target well without immobilized target (streptavidin/neutravidin only).	128
Table S7.4: Next generation sequencing of enriched CX ₉ C bacteriophage library following screening experiment 1. A total of 20282 sequences were identified, with sequences in abundance >40 listed below, in decreasing order.	128
Table S7.5: Results of bacteriophage biopanning experiment 2 with the sequence format CX ₄ WLGX ₂ C. Enrichment factor for each round is defined as the amount of phage eluted from the target well containing D-TNRC2 divided by the amount of phage eluted from the target well without immobilized target (streptavidin/neutravidin only).	129
Table S7.6: Sequencing of enriched CX ₄ WLGX ₂ C bacteriophage library following screening experiment 2. A total of 22 monoclonal sequences were randomly selected, with sequences listed below, in decreasing order of abundance.	129
Table S7.7: X-ray data collection, processing, and refinement statistics. [a] TNRC2 not yet deposited in PDB.	203

Table of abbreviations

Amino acid abbreviations

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tyrosine	Tyr	Y
Tryptophan	Trp	W
Valine	Val	V

All other abbreviations

[θ]MR	Mean residue ellipticity
AC	American type culture collection
Acm	Acetamidomethyl
ACPC	<i>trans</i> -(1S,2S)-2-aminocyclopentyl-1-carboxylic acid
APC	<i>trans</i> -(3R,4S)-4-aminopyrrolidine-3-carboxylic acid
ASFV pol X	African swine fever virus polymerase X
AucA	Aureocin A53
BHI	Brain heart infusion
BTD-2	β -defensin-2
BurResTNF	Surface area of TNF- α binding residues buried by scaffold docking
CANWARD	Canadian antimicrobial resistance alliance nosocomial pathogen surveillance
CD	Circular dichroism
CFU	Colony forming units
COVID-19	Coronavirus disease 2019
CPU	Central processing unit
CRD	Cysteine-rich domain
CSLI	Clinical and laboratory standards institute
CW	CANWARD
D-	Dextrorotatory
Dbz	3,4-diaminobenzoic acid
DCM	Dichloromethane
Ddae	1-(4,4-dimethyl-2,6-dioxocyclo-hexylidene)-3-[2-(2-aminoethoxy)ethoxy]-propan-1-ol
Ddap	2-(7-amino-1-hydroxyheptylidene)-5,5-dimethylcyclohexane-1,3-dione
Ddg	Rosetta simulated binding energy
DIC	<i>N,N'</i> -Diisopropylcarbodiimide
DIPEA	<i>N,N</i> -Diisopropylethylamine
DMF	<i>N,N</i> -dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid

Dpo4	DNA polymerase IV
EDT	1,2-ethanedithiol
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ESI+ MS	Positive electrospray ionisation mass spectrometry
fa _{atr}	Lennard-Jones attractive force
FDA	United States food and drug administration
Fmoc	Fluorenylmethyloxycarbonyl
FRET	Fluorescence resonance energy transfer
G3P	L-glycerol 3-phosphate
GCI	Grating-coupled interferometry
GHz	Gigahertz
GiB	Gigabyte
Gn·HCl	Guanidinium chloride
gp41	Glycoprotein 41
GSH	Glutathione
GSSG	Glutathione disulfide
HBS	HEPES buffered saline
HBS-EP	HBS with EDTA and P20
HBTU	(2-(1 <i>H</i> -benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HDMS	High-definition mass spectrometry
HEK	Human embryonic kidney
HEPES	<i>N</i> -2-hydroxyethylpiperazine- <i>N</i> -2-ethane sulfonic acid
HIV	Human immunodeficiency virus
HOBt	Hydroxybenzotriazole
HPLC	High-performance liquid chromatography
HT	High throughput
IC ₅₀	Half-maximal inhibitory concentration
IgV	Immunoglobulin-like variable
ITIM	Immunoreceptor tyrosine-based inhibitory motif
KAHA	α -Ketoacid-Hydroxylamine
K_D	Equilibrium dissociation constant
kDa	Kilodalton
L-	Levorotatory

LC-MS	Liquid chromatography (HPLC) – electrospray ionization mass spectrometry
LCP	Monoolein lipid cubic phase
LigA	<i>Haemophilus influenzae</i> DNA ligase
LnqQ	Lacticin Q
M2-TM	Transmembrane helix of the <i>influenza</i> M2 ion channel protein
MHB	Mueller-Hinton Broth
MHC	Major histocompatibility complex
MIC	Minimum inhibitory concentration
miPCR	Mirror-image polymerase chain reaction
MIPD	Mirror-image phage display
MPAA	4-mercaptophenylacetic acid
mRNA	Messenger RNA
Mtt	4-methyltrityl
NaPi	Sodium phosphate
NCL	Native chemical ligation
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance
NRPS	Non-ribosomal peptide synthetases
OG	β-octyl-glucoside
P20	Polysorbate 20
PDB	Protein data bank
PD-L1	Programmed-cell death protein ligand 1
PEG	Polyethylene glycol
PEG-PS	Polyethylene glycol-polystyrene
Pen	Penicillamine
Pfu	<i>Pyrococcus furiosus</i> DNA polymerase
PISA	Protein interfaces, surfaces, and assemblies
PLP	Pyridoxal phosphate
PTFE	Polytetrafluoroethylene
PTM	Post-translational modification
PTM'd	Post-translationally modified
RAM	Random access memory
RAPID	Repeated analyte pulses of increasing duration
RBM	Removable backbone modification

RI	Retro-inversion
Rif	Rotamer interaction field
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RP	Reversed phase
rRNA	Ribosomal RNA
SEA	bis(2-sulfanylethyl)amido
SEC	Size exclusion chromatography
SPE	Solid-phase extraction
SPPS	Solid-phase peptide synthesis
STA	Streptavidin
STL	Serine/Threonine ligation
sTNFR-1	Soluble Tumour necrosis factor receptor I
tBuSH	tert-butylthiol
TCEP	tris(2-carboxyethyl)phosphine
TCPB	TNFR-1 cyclic peptide binder
TFA	Trifluoroacetic acid
TFE	2,2,2-Trifluoroethanol
TFET	2,2,2-Trifluoroethanethiol
TGIT	T-cell immunoglobulin and ITIM
Thz	Thiazolidine
T_m	Melting temperature
TMPB	TNFR-1 mini protein binder
TNFR-1	Tumour necrosis factor receptor I
TNFR-2	Tumour necrosis factor receptor II
TNF- α	Tumour necrosis factor α
TNRCD2	Tumour necrosis factor receptor I - cysteine rich domain 2
tRNA	Transfer RNA
Trt	Trityl
Ub	Ubiquitin
UPLC	Ultra-performance liquid chromatography
UV	Ultra-violet
VEGFA	Vascular endothelial growth factor A
XML	Extensible markup language

Acknowledgements

This thesis summarizes the research that I have conducted throughout the duration of my PhD studies. However, the work presented here would not be possible without the extensive support and collaboration of my colleagues, spanning from MSc students to senior academics.

I extend great thanks to my primary PhD supervisor, Dr. L.Y.P. Luk, for providing me with the opportunity to undertake this work, the consistent motivation and feedback provided, and for allowing me the freedom to pursue my own ideas. I am indebted to his support not only as the principal investigator of this work, but as a career mentor, providing me with new opportunities for collaboration, international conference attendance (APS2022 & EPS2022), and motivating me to achieve my full potential.

I also thank Dr. Y. Jin, serving as the co-investigator of the work in this thesis, who has provided me with extensive training in all aspects of protein crystallography. Additionally, thanks to YJ for her research insights and ideas, where together with LYPL, we conceived the idea to investigate drug targets in COVID-19 (Chapter 3) during a zoom call in the lockdown of 2020.

I thank Dr. Y. Tsai and co-supervisor, Prof. R.K. Allemann, for providing access to essential shared resources and equipment used in this work and thank Dr. B. Bax for access to the GCI equipment.

Research discussed in this thesis was also conducted with the help of some highly talented collaborators. I thank Dr. B. Findlay and L. Domínguez of Concordia University, Canada, for conducting the antibiotic activity assays against nosocomial pathogens (Chapter 2). I also thank Prof. C. Wu and Y. Kong of Xiamen University, China, for carrying out the phage display experiments (Chapter 3). Finally, thanks to Dr. F. Parmeggiani of Bristol University, UK, for allowing me the opportunity to work in his group (11/2021-12/2021), providing invaluable training to conduct Rosetta *de novo* protein design, and to his group for the support during my stay (Chapter 3).

I also thank members within Cardiff University who assisted directly in this project: Dr. X. Li, for optimization of the initial synthetic route to the lactacin Q protein (Chapter 2), and Dr. P. Rizkallah and Dr. P. Baumann, for their assistance with the X-ray diffraction data collection.

I extend thanks to other members of our department, for the training, inspiration, and support during my PhD studies: to Dr. T. Williams, for training in mass spectrometry; to Dr. A. Nodling, for training in chemistry, biochemistry, and general mentorship; to Dr. N. Santi, for training in circular dichroism spectroscopy; to Dr. H. Bell, for training in protein crystallization; to Dr. K.

Elvers, for training in GCI; to V. Barlow, for training in antibiotic activity assays. I also thank members of our peptide group at Cardiff: D. Cardella, for the consistent feedback and inspirations during this PhD; L. Shepherd, for preparing the scaffold miniprotein (TMPB-Sc, Chapter 3); G.E.S. Smith and M. Übler, for the support in general operation of the group.

I also thank those members of the group, past and present, whose feedback during group meetings and general enjoyable company in the lab stimulated my productivity: Dr. T.M.S. Tang, Dr. E. Mills., Dr. Z. Wu, Dr. W. So, L. Spear, D. Zappala, H. Hayes, M. Dong, J. Chen, A. Coricello, S. Patel, M. Ma, L. Hatton and M. Dickerson.

This PhD would not be possible without the support of funding bodies. Thanks to EPSRC for the PhD studentship (2107414), and to Universities Wales for the research mobility grant (UNIW/RMF-CU/19) which facilitated my research stay at Bristol University and attendance of APS2022. I also thank Diamond Light Source for access to beamlines iO3 and iO4-1 (proposal numbers: 21047 and 29990).

Finally, I extend my thanks to those people not directly involved with this research, but whose emotional support provided me with the motivation to pursue this PhD: to my mother, G. Lander, for her constant support and enthusiasm to allow me to pursue my career in science; to my brother, P. Lander, and close friends, L. O'leary and L. O'Leary, for always motivating me to achieve my career aspirations; to my grandparents, H. Lander and H.J. Lander, for providing financial support to pursue an invaluable MSc degree prior to my PhD studies.

Chapter 1 : Introduction

A review manuscript has submitted (14/09/2022) for peer-review, with A.J. Lander serving as the first author. Figures and text contained within this chapter have been reproduced/used in the submitted manuscript, with the permission of co-authors Dr. L.Y.P Luk and Dr. Y. Jin.

1.1 Research surrounding D-protein enantiomers

Proteins, like other biomolecules, are composed of chiral building blocks.^{1, 2} Ribosomes recruit L- (Levorotatory) amino-acids for catalysis, and hence recombinant proteins are largely refrained in the L-framework unless engineered ribosomes are used.³ In contrast, incorporation of D- (Dextrorotatory) amino acids into L-polypeptides requires the use of engineered ribosomes,^{4, 5} post-translational modification systems (PLP-dependent enzyme)⁶ or non-ribosomal peptide synthetases (NPRS).⁷ At the time of writing, proteins entirely comprised of D-amino acids have yet to be found in nature and must be synthesized via chemical routes. Though being more difficult to prepare, D-amino acid proteins that fold into reciprocal chirality possess extraordinary potentials in scientific research, spanning from the creation of mirror-image life, mechanistic investigations of natural proteins to the isolation of ultra-stable binders.⁸

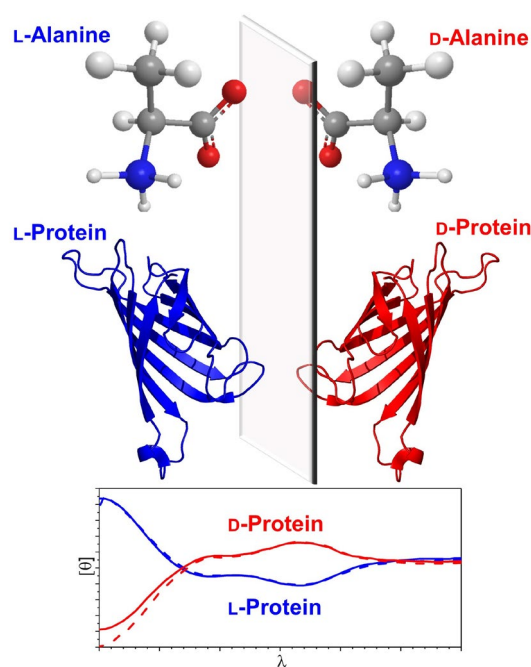


Figure 1.1 Proteins comprised entirely of D-amino acids fold into the mirror image of the corresponding L-protein.

1.1.1 Towards ‘mirror-image life’

Mirror-image life which was first proposed by Louis Pasteur in 1860⁹ refers to the creation of an artificial biosystem with all macromolecules presented in their opposite enantiomeric forms. In these self-replicating systems, L-nucleic acids serve to store genetic information creating D-protein workforce for biological function following the mirror-image central dogma.¹⁰ Indeed, the *de novo* design of living entities has gained significant attention because of our fundamental interest in understanding the origin of life.¹¹⁻¹³ While preparation of an entirely self-replicating living entity in mirror image form is a major challenge, D-enantiomers of key enzymes involved in the central dogma of molecular biology have been prepared (Figure 1.2).⁸ These enantiomeric enzymes hold potentials in research. For example, mirror-image polymerases can be used to generate L-nucleotide aptamer libraries¹⁴ or L-genomes for bioorthogonal information storage.¹⁵ Also, an enantiomeric ribosome could facilitate access to D-proteins through mirror-image translation. Furthermore, creation of a self-replicating mirror-image entity can offer access to D-proteins, L-sugars and/or other enantiopure pharmaceutical compounds. Recent examples of preparing D-enzymes towards mirror-image life are summarized below:

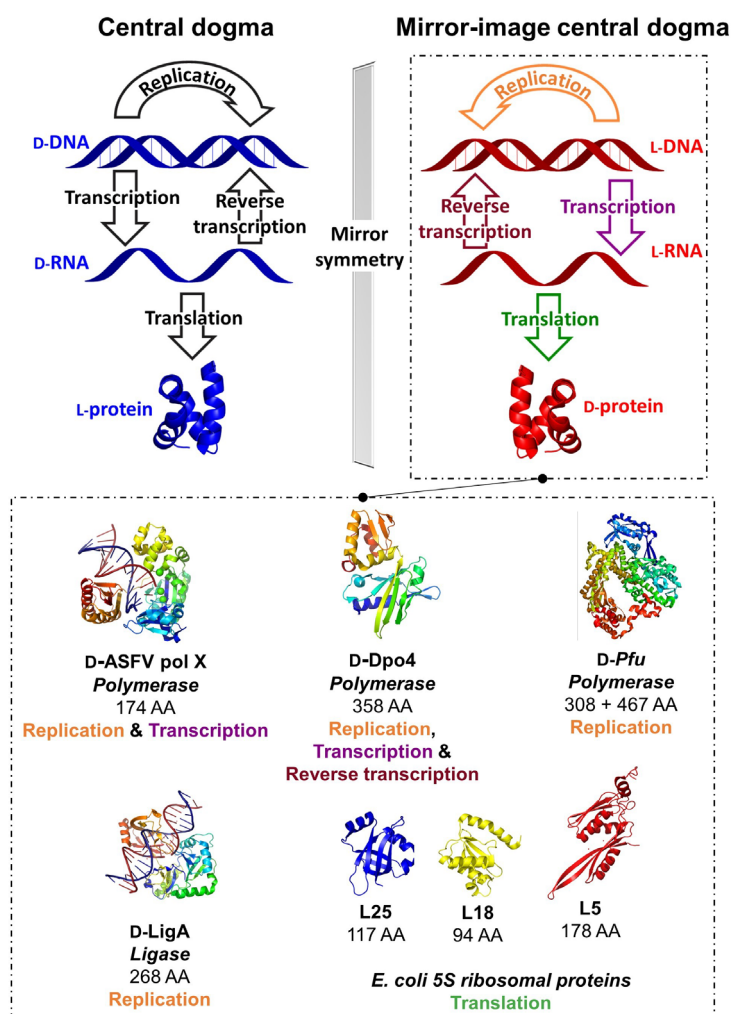


Figure 1.2 Recent advances in realizing mirror-image synthetic biology using D-protein enantiomers. Representative structures for illustrative purposes as D-ASFV pol X (PDB: 1JQE), D-Dpo4 (PDB: 3PR4), D-Pfu (PDB: 2JGU), D-LigA (PDB: 2Q2T), L5, L18 and L25 (PDB: 4YBB) constructed in PYMOL.¹⁶

Creation of artificial L-polynucleotides by use of mirror-image nucleic acid polymerases

Template-directed polymerizations of the enantiomeric L-DNA and -RNA were first carried out using the D-protein enantiomer of African swine fever virus polymerase X (D-ASFV pol X).¹⁷ Composed of only 174 residues, ASFV pol X is the smallest DNA polymerase known, hence an ideal candidate for total chemical synthesis. The longest polynucleotide successfully replicated by D-ASFV pol X was 44 nucleotides in length but required fresh enzymes at each cycle due to its weak thermal stability.¹⁷ A significant advancement was achieved by preparing the 358-residue enantiomeric P2 DNA polymerase IV (Dpo4) from *S. solfataricus*, which is sufficiently

stable to temperature flux and could perform mirror-image polymerase chain reaction (miPCR).^{14, 18} The D-Dpo4 could successfully create a 120-bp L-DNA sequence encoding the *E. coli* 5S ribosomal RNA gene *rrfB*.¹⁴ Its thermostable variant D-Dpo4-3C could assemble a full L-DNA gene encoding protein Ssoo7d.¹⁸ Interestingly, a further-engineered variant D-Dpo4-5m-Y12S was reported that is capable of both transcription and reverse transcription, laying a strong foundation for enabling mirror-image life.¹⁹

In order to create a lengthy enantiomeric gene with high fidelity, access to polymerase enzymes with a low error rate is essential, thus an enantiomeric derivative of *Pyrococcus furiosus* (Pfu) DNA polymerase has been deduced.²⁰ Pfu is composed of 775 amino acids reaching 90 kDa in molecular weight, rendering its chemical synthesis challenging. To circumvent this issue, a split version of Pfu was prepared consisting of N- (467-residue) and C- (308-residue) fragments. Due to the significant cost of D-isoleucine and its association with the aggregation of peptide fragments, most of them were replaced by other bulky residues including valine and leucine.¹⁵ The split polymerase D-Pfu could synthesize a 1.5-kb mirror-image gene from short, synthetic oligonucleotides. Interestingly, because of the inherent stability towards enzymatic cleavage, a trace amount of artificial L-DNA preserved in water from a local pond remained amplifiable by D-Pfu after one year, whereas D-DNA could not be amplified by L-Pfu after one day. This work is a clear leap forward in the pursuit of mirror-image biology. In addition, the miPCR platform is potentially useful in molecular discovery programs generating nuclease-resistant L-nucleotide aptamers for critical drug targets.¹⁴

Other life-essential, mirror-image proteins

Enantiomeric ligase is another critical enzyme that can be used to create long stretches of L-DNA.²¹ Preparation of a mirror-image ribosome is exceptionally challenging as it composes of multiple protein and nucleotide subunits.²² Currently, three out of the ~50 enantiomeric *E. coli* ribosomal proteins have been reported, including L5, L18 and L25.²³ The D-ribosomal proteins were prepared with native post-translational modifications and interacted specifically with L-5S RNA to form a mirror-image ribonucleoprotein complex. On the other hand, eukaryotic ribosomes consist of 79-80 proteins and four rRNAs,²² requiring approximately 200 non-ribosomal factors for assembly.²⁴ Both the ribosomal proteins and rRNA itself also bear post-translational modifications,²⁵ adding further challenges to their preparation.

Remarks

The genome of the laboratory strain *E. coli* K12 encodes for approximately 4300 proteins,²⁶ but only a small fraction of their enantiomeric counterparts have been reported.²⁷ Many of these proteins bear intrinsic synthetic challenges, because of their size, post-translational modifications and folding (see Section 1.2). Construction of the necessary mirror-image oligonucleotides is also a major challenge. Whilst the synthesis of relatively short oligonucleotides is possible from L-xylose or L-arabinose,²⁸ the resulting oligonucleotides suffer from lower purity.^{14, 17, 18} With the advent of the high fidelity D-Pfu capable of assembling complex genes,¹⁵ one might argue that this challenge is within reach. Given the significant efforts, it is also expected that a mirror-image ribosome from bacteria will soon be reported. In addition, mirror-image tRNAs, tRNA synthetases, and translation factors will need to be prepared to enable the translation of L-mRNA to D-polypeptides. When made available, the mirror-image translation system will be game-changing in the landscape of enantiomeric protein synthesis. Finally, pursuit of a truly self-replicating system will require an approach of devising a minimal cell and assembly of each essential component in mirror-image form.

1.1.2 Racemic protein crystallography

Racemic protein crystallography utilizes synthetic protein enantiomers for crystallization and is a technology particularly useful at yielding atomistic structural insights.²⁹ Unlike native L-proteins, racemic proteins can crystallize into achiral space groups possessing higher symmetry and order (Figure 1.3).³⁰ Solving structure based on racemic proteins can be advantageous. As illustrated in the first example, the phase issue in solving the structure of rubredoxin was vastly simplified, because the space group was found to be centrosymmetric with the crystal unit cell containing a center of inversion (Figure 1.3 B).³¹ The off-diagonal phases of the X-ray diffraction data obtained from a centrosymmetric crystal cancel out, restricting the phases to 0° or 180°, as opposed to the possible 0° to 360° arising from a homochiral crystal.²⁹ Additionally, the centrosymmetric crystal has high dimensionality. This generally results in rapid protein crystallization and structure solving with high-resolution detail, in addition to unveiling solute and ligand interactions.^{29, 32, 33}

Considering the higher order of crystal symmetry and favorable crystal growth, racemic protein crystallography has been used to resolve X-ray structures of numerous proteins up to ultra-high (sub-angstrom) resolution (Table 1). Indeed, 13 (20% so far) of the reported racemic crystal structures were resolved in the centrosymmetric space group $P-1$. Many quasi-racemic crystal structures, in which the enantiomers differ slightly, are reported to adopt the space group $P1$ due to a lack of true symmetry.³⁴ However, it is more appropriate to classify these crystals as *pseudo-centrosymmetric* or “*pseudo- $P-1$* .”^{30, 35, 36} While almost half of the racemic structures were resolved in centrosymmetric/pseudo-centrosymmetric space groups, the other half were also resolved in chiral space groups (Table 1.1). Notably, there are no reported instances of a single enantiomer crystallizing into a chiral space group from a racemate. Thus, crystallization of proteins from racemates may have advantages beyond that explained by the achiral space group theory (for review - see ref²⁹).

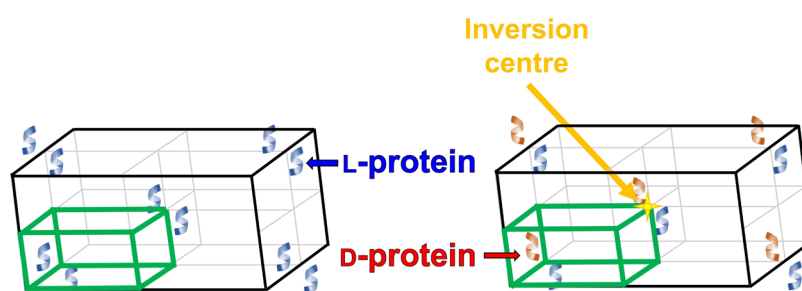


Figure 1.3 Illustrative comparison of (A) a homochiral crystal unit cell and (B) a racemic crystal unit cell containing an inversion center (e.g. $P-1$).

Racemic protein crystallography has been most frequently applied to study mini proteins containing fewer than 100 residues. These proteins are known to be difficult to crystallize because of their globular morphology which disfavors crystal packing. Meanwhile, their small sizes render the chemical synthesis of these proteins feasible (see Section 1.2 below for synthetic approaches).³⁷ Some of the unique insights generated by racemic crystallography are discussed in this section.

Table 1.1 A decade of racemic protein crystallography (at the time of writing).

Protein	Function	Length ^a	Res. (Å)	Space group	PDB accession	Structural insights	Ref
Lacticin Q	Bacteriocin	53	0.96	<i>P</i> 1	7P5R	First reported crystal structure	38
CyO2	Bracelet cyclotide	30	1.17	<i>P</i> 12 ₁ 1	7RMQ	First reported crystal structures of bracelet cyclotides	39
CyO2 (I11L)			1.04	<i>P</i> 12 ₁ 1	7RMR		
CyO2 (I11G)			1.10	<i>P</i> 12 ₁ 1	7RMS		
Hyen D			1.35	<i>P</i> 12 ₁ 1	7RIH		
Hyen D (I11L)			1.22	<i>P</i> -1	7RII		
Hyen D (I11G)			1.30	<i>P</i> 1	7RIJ		
Calciclude	Kunitz-type serine protease inhibitor homolog	60	2.52	<i>I</i> 4 ₁	6KZF	Confirmation of novel disulfide surrogate bridge strategy (DADA)	40
rC5a-desArg	Rat anaphylatoxin	76	1.80	<i>P</i> 1	^g	New insights on C-terminal conformation	41
Chimeric-rC5a		77	1.31	<i>P</i> 2 ₁ 2 ₁ 2 ₁	^g	Chimeric protein probes conjugated to small-molecule antagonist	42
rC5a			1.58	<i>P</i> -1	^g		
M2-TM ^b	Ion channel TM helix	24	2.00	<i>P</i> 2 ₁ / <i>c</i>	4RWB	Investigations of a heterochiral coiled coil	33
M2-TM ^c			1.05	<i>P</i> -1	4RWC		
M2-TM (I39A) ^c			1.55	<i>P</i> -1	6MPL		32
M2-TM (I42A) ^c			1.40	<i>P</i> -1	6MPM		
M2-TM (I42E) ^c			1.40	<i>P</i> -1	6MPN		

Melittin	Honeybee venom	27	1.27	C2	6O4M	Retention of native quaternary structure	43
Ribifolin	Orbitides from <i>Jatropha</i>	8	0.99	P12 ₁ /n1	6DKZ	Unveil structures of <i>Jatropha</i> orbitides	44
Pohlianin C		8	1.20	Pcab	6LD0		
Jatrophidin		8	1.03	P12 ₁ /n1	6DL1		
GsMTx4	Spider venom	34	1.75	P-3	^g	First reported crystal structure	45
BTD-2	Baboon θ -defensin	18	1.45	P-1	5INZ	Novel oligomeric state resembles mechanistically relevant assembly	46
Snakin-1	Potato snakin	63	1.50	P1	5E5Y	Novel use of radiation damage induced phasing of quasi-racemic crystals	47
			1.60	P21/c	5E5Q		
			1.57	P21/c	5E5T		
Ubiquitin	Ubiquitin	76	1.95	^d	^g	Confirm folding of synthetic protein	48
M1-linked tri-Ubs		76 ^e	1.80	P1	5GO7	D-monomeric Ub can facilitate Ub oligomer crystallisation	34
M1-linked tetra-Ub			2.18	P2 ₁	5GO8		
K6-linked di-Ub			1.15	P1	5GOB		
K11-linked di-Ub			1.73	P1	5GOC		
K27-linked di-Ub			1.15	P1	5GOD		
K29-linked di-Ub			1.98	P2	5GOG		
K33-linked di-Ub			1.95	P1	5GOH		
K48-linked di-Ub			1.59	P1	5GOI		
K63-linked di-Ub			1.55	P2 ₁ 2 ₁ 2	5GOJ		

K11/K63-linked tri-Ub			1.84	<i>P22₁2₁</i>	5GOK		
K27-linked di-Ub		152	1.55	<i>C2</i>	5J8P	Largest true synthetic racemic proteins to be crystallized	49
K27-linked tri-Ub		228	2.10	<i>H3</i>	5JBV		
VHP	Vinillin headpiece domain	35	2.10	<i>P-1</i>	3TRW	Investigation of pentafluoro phenylalanine (F ₅ Phe) amino acids on protein structure	50
VHP			2.30	<i>I-4c2</i>	3TRY		
VHP (F ₅ Phe10)			1.46	<i>F222</i>	3TJW		
VHP (F ₅ Phe17)			1.00	<i>P1</i>	3TRV		
VHP (β3-hGln26)			1.30	<i>P1</i>	5I1N	Investigation of beta amino acids on protein structure	51
VHP (ACPC26)			1.35	<i>P1</i>	5I1O		
VHP (β3-hLys30)			1.40	<i>P1</i>	_g		
VHP (APC30)			1.12	<i>P1</i>	5I1S		
Ts3	Scorpion venom	64	1.93	<i>P-1</i>	5CY0	First reported structure of Ts3	52
Magainin 2 (L-1)	Amphibian HDP	23	1.75	<i>I-42d</i>	4MGP	Beta amino acid variants investigating phenylalanine zipper motif	53
Magainin 2 (L-2)			2.20	<i>P2₁2₁2</i>	5CGN		54
Magainin 2 (L-3)			1.50	<i>P1</i>	5CGO		
ShK	Sea anemone venom	35	0.97	<i>P12₁/c</i>	4LFS	Structure variation to NMR and enantiospecific activity	55
ShK analogue			1.20	<i>H-3</i>	4Z7P	Structure activity relationships	56
ShK (allo-Thr13)			0.90	<i>P1</i>	5I5B	Investigation of side chain chirality on protein structure	57
ShK (allo-Thr31)			1.30	<i>P12₁1</i>	5I5C		
ShK (allo-Ile7)			1.20	<i>C2</i>	5I5A		

Rv1738	<i>M. tuberculosis</i> protein	94	1.50	C12/c1	4WPY	First reported structure, unknown function	58
STFI-1	Sunflower trypsin inhibitor	14	1.25	P-3	4TTK	Disulfide-rich scaffolds for drug design	59
cVc1.1	Cone snail venom	22	1.70	Pbca	4TTL		
kB1	Plant cyclotide	29	1.90	P-1	4TTM		
kB1 (G6A)			1.25	P-1	4TTN		
kB1(V25A)			2.30	P-1	4TTO		
Ser-CCL1	Chemokine	73	2.15	P1	4OIJ	Crystal structure of a homogenous, glycosylated chemokine	60
Glycosylated Ser-CCL1		73 ^f	2.10	P1	4OIK		
DKP Ester Insulin	Synthetic hormone + derivatives	51	1.60	P-1	4IUZ	Confirm folding of synthetic derivative	61
Ester insulin			1.50	I2 ₁ 3	5EN9	Confirmation of correctly folded synthetic protein for isotope experiments	62
Human insulin			1.35	I2 ₁ 3	5ENA		
VEGF-A/antagonist complex	Vascular endothelial growth factor A + D-protein binder	102 + 56	1.60	P2 ₁ /n	4GLN	First reported structure of a heterochiral protein complex by racemic crystallography	63
Crambin analogue	Thionin protein	46	1.08	P12 ₁ 1	3UE7	Novel linear-loop peptide chain topology	35
Kalioxin	Scorpion venom	38	0.95	P-1	3ODV	Basis for structure activity relationships	64
Omwaprin	Snake venom	50	1.30	P2 ₁ /c	3NGG	First reported structure	37

^a Total amino acid length of synthetic protein enantiomer, ^bcrystallized from monoolein lipidic cubic phase, ^ccrystallized from racemic lipids ^d data unavailable, ^e residues in D-protein enantiomer, ^f 73 residues + oligosaccharide, ^g Not reported/deposited

Quaternary states of protein

Oligomeric assemblies are thought to play a common role in the activity of a variety of antimicrobial peptides, particularly those acting on the bacterial membrane.⁶⁵ In an aim to elucidate its mechanism of action, the β -sheet antimicrobial peptide originated from Baboons (BTD-2) was chemically synthesized in both L- and D-forms. Interestingly, racemic protein crystallography of BTD-2 revealed a novel anti-parallel trimeric form (Figure 1.4 B). This supramolecular discovery is fibril-like and is postulated to have critical roles in membrane disruption.⁴⁶ In another example, melittin is an α -helical antimicrobial peptide isolated from honeybee venom which is known to exert its activity by disrupting the bacterial cell membrane.

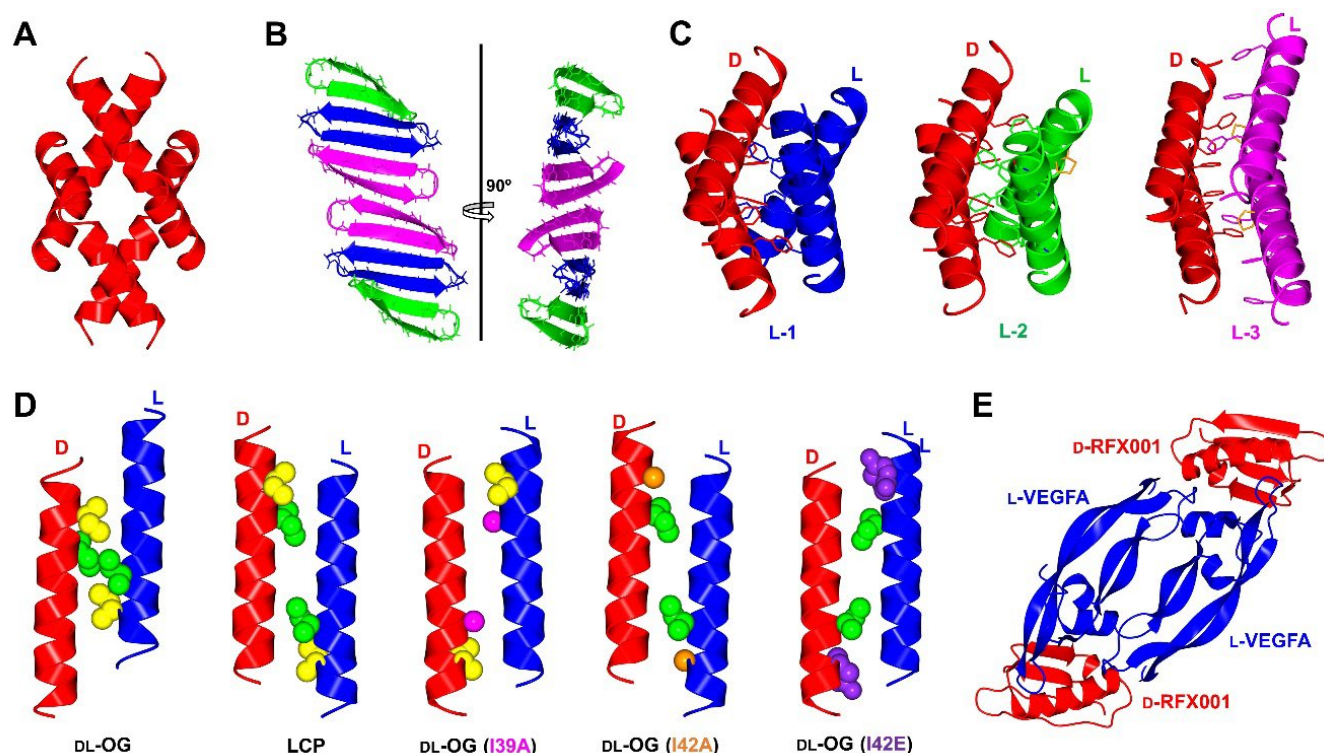


Figure 1.4 Quaternary structures by racemic protein crystallography; **(A)** Melittin tetramer (PDB: 6O4M); **(B)** BTD-2 extended fibril-like structure (PDB: 5INZ); **(C)** Magainin 2 phenylalanine zipper motif unaffected by β -amino acid substitutions. D-magainin 2 is shown in red and mutants L-1 (Ala) shown in blue (PDB: 4MPG), L-2 (APC) in green (PDB: 4CGN) and L-3 (ACPC)(PDB: 5CGO) in magenta. β -amino acids highlighted in orange; **(D)** M2-TM helix forms heterochiral coiled coils, with a hendecad repeat identified in lipidic cubic phase (LCP, PDB: 4RWB) but absent in racemic β -octylglucoside (DL-OG, PDB: 4RWC). Mutation of sterically disruptive isoleucine residues to alanine (DL-OG(I39A), PDB: 6MPL), (DL-OG(I42A), PDB: 6MPM) or glutamate (DL-OG(I42E), PDB: 6MPN) favored hendecad repeat motifs in chiral lipids; **(E)** Quaternary structure of VEGF-A dimer bound to two D-protein antagonist molecules (PDB: 4GLN). All structures were modelled in CCP4MG⁶⁶ with data obtained from the Protein Data Bank.

The tetrameric assembly observed in the solution state is also present in the racemic X-ray crystal structure (Figure 4A).⁴³ Similarly, the tetrameric nature of magainin 2 was suggested to be critical for the activity of this amphibian host defense peptide (Figure 4C).⁵⁴

Heterochiral interactions between L- and D-protein isomers observed during structure elucidation may also lead to fruitful development in binder creation. In the studies of the transmembrane helix of the influenza M2 ion channel protein (TM-M2),³³ a heterochiral coiled-coil association was observed between the two peptide enantiomers in the presence of detergent octyl-glucoside (DL-OG) or within the monoolein lipid cubic phase (LCP) (Figure 4D). The LCP structure shows an 11-residue helical repeat (hendecad, 3,4,4 spacing) in the coiled-coil, which differs from homochiral coiled-coils that adopts a 7-residue helical repeat (heptad, 3,4 spacing). The crystals grown in DL-OG do not form a hendecad repeat, as steric clashes involving Ile39 and Ile42 prevent proper 3,4,4 interaction. Substitution of these residues with alanine or glutamate produced the hendecad repeat coiled-coil in the racemic DL-OG structure, thus reinforcing the argument that hendecad repeats are a feature of heterochiral coiled coils.³² Such heterochiral interactions can be used to design D-proteins drugs, which are generally non-proteolytic and non-immunogenic (see section 2.3). The resulting drug-target complexes can also be resolved using racemic protein crystallography to aid in rational optimization (Figure 4E).

Post-translationally modified proteins

While obtaining homogenous, post-translationally modified (PTM'd) proteins through recombinant methods remains a major technical challenge,⁶⁷ chemical protein synthesis offers exquisite atomistic control and thus ensures homogeneity (see also Section 3). Racemic protein crystallography of PTM'd proteins was first applied to the glycosylated chemokine Ser-CCL1 protein, for which no structure was reported.⁶⁰ The protein was synthesized in the native L-form, followed by site-specific asparagine *N*-glycosylation with the native biantennary D-glycan. Synthesis of the corresponding D-enantiomer without glycosylation enabled the co-crystallization of the quasi-racemic protein (Figure 5A).

Another application involves the study of branched ubiquitin chains, where the folding of the branched protein molecules could be solved through racemic protein crystallography. Similarly, D-ubiquitin was prepared in unmodified form⁴⁸ and used to facilitate crystallization of the resulting branched L-ubiquitin proteins (Figure 5B).³⁴ A further example involved the preparation of branched ubiquitin proteins in both enantiomeric forms for racemic protein crystallography (Figure 5C), but the iso-peptide linkages of the D-proteins contained a non-native cysteine residue scar to facilitate ligation.⁴⁹

A key challenge, as presented in the former examples, is the preparation of PTM'd proteins in all D-form. Glycosylated proteins possess glycans in native D-chirality which would require complex synthesis from the corresponding L-carbohydrates for a true racemic crystal. In addition, preparation of branched ubiquitin chains requires the use of non-natural amino acids as auxiliaries for attachment of the ubiquitin, but they often suffer from poor ligation efficiency (see section 3.3.2).⁴⁹

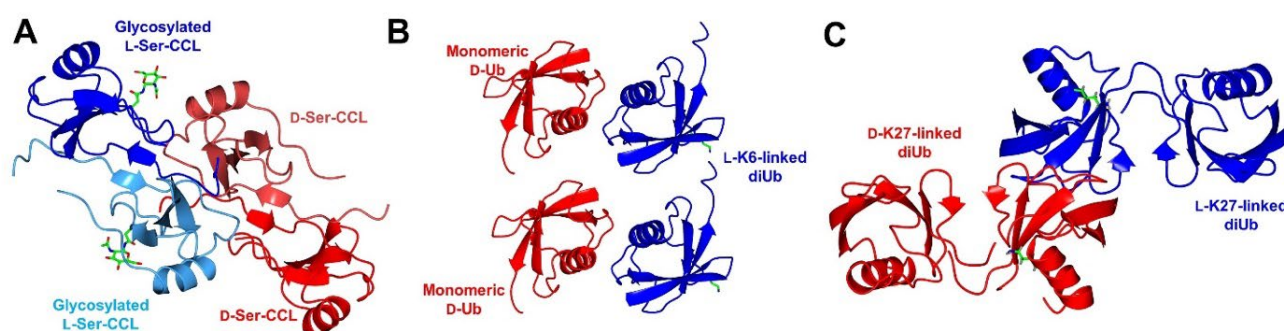


Figure 1.5 Quasi-racemic protein crystallography of homogenous, post-translationally modified proteins; (A) D-Ser-CCL facilitated crystallization of glycosylated L-Ser-CCL1 (PDB: 4OIK); (B) Monomeric D-ubiquitin (D-Ub) facilitated crystallization of L-K6-linked diUb (PDB: 5GOB); (C) Quasi-racemic protein crystallography of L- and D-K27-linked diUb (PDB: 5J8P). All structures modelled in CCP4MG⁶⁶ with data obtained from the protein data bank.

Remarks

Challenges associated with D-protein synthesis, folding and PTMs hamper the application of racemic protein crystallography. The average size of a protein ranges from 283-438 residues in length⁶⁸ with many bearing PTMs. Obtaining enough D-protein for crystallization screening (generally in milligram range) remains labor-intensive and uneconomical, typically requiring multiple chemical steps and protein refolding. Nevertheless, racemic protein crystallography of miniproteins remains an excellent method for deciphering molecular interactions, particularly serendipitous intermolecular interactions that deem difficult to obtain using homochiral protein crystallography, solution state NMR and/or computational structure prediction.^{32, 33, 46, 51, 57} Perhaps, a more promising avenue is to conduct quasi-racemic crystallography where a minimal D-protein is used to facilitate crystallization of a larger L-protein with (pseudo-)repeated domains.³⁴

1.1.3 Identification of drug candidates through mirror-image phage display and related screening technologies

Polypeptide binders can offer significant selectivity and potency, and hence are excellent candidates for the treatment of various diseases and human disorders. One major bottleneck is that many peptide candidates suffer from proteolytic degradation, both limiting the option of the delivery methods and eliciting unwanted immune responses caused by major histocompatibility complex (MHC) presentation by immune cells.⁶⁹ Peptides comprised of D-amino acids are a viable approach as they are non-recognizable by endogenous proteases.⁷⁰ It has been suggested that D-peptide binders can be made by retro-inversion (RI) which relies on flipping the entire peptide chain from the N- to C- termini to offset the flip in the side chain chirality.⁷¹ Though some success has been seen in short binders, it was quickly discovered that this double-flip approach did not reinstate the true peptide structure and, in some cases, could drastically weaken the peptide binding.^{72, 73} Another approach was to use D-peptides to mimic the shape of the L-peptide agonist when presented in an MHC, without reference to the L-peptide sequence, acting as a stable vaccine candidate.⁷⁴ Recently, binders composed of both D- and L-residues have been developed through ribosomal engineering⁷⁵⁻⁷⁸ or *in silico* protein design.⁷⁹ In order to create binders entirely composed of D-amino acids, the most routine approach is mirror-image phage display (MIPD).⁸⁰ In MIPD, D-enantiomers of protein targets are synthesized and subjected to L-peptide screening (Figure 1.6). Due to the nature of mirror-image symmetry, the same-sequence D-peptide will bind to the native L-protein target with equal affinity, thus yielding an inherently non-proteolytic peptide binder. MIPD has discovered D-peptide binders for a range of targets (Table 1.2). Key examples are summarized below:

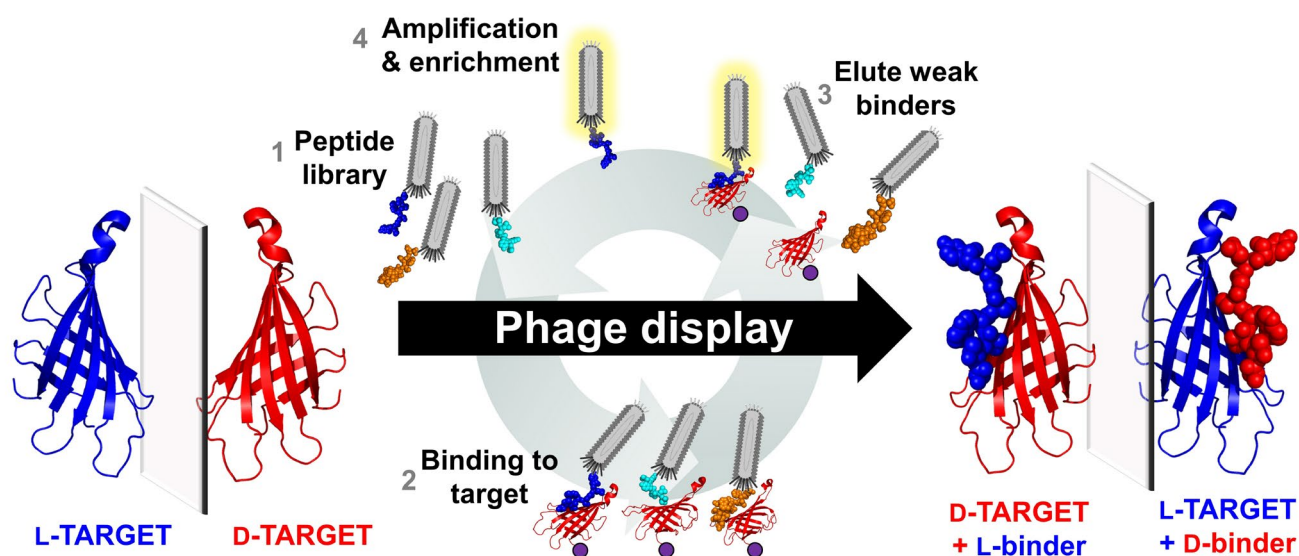


Figure 1.6 Process of mirror-image phage display. Natural chirality L-proteins are represented in blue, and mirror-image synthetic D-proteins are represented in red. Purple beads represent protein immobilization.

Table 1.2 D-peptide binders identified through mirror-image phage display of D-protein targets in the last decade. (at the time of writing).

Protein	Function	Length ^b	Name	Type	Length ^c	K _d (μM) ^d	Ref
ARQ23	Androgen receptor	46	QF2D-2	Linear	16	11	81
Annexin A1 NTD	Surface marker in malignant tumour vasculature	16	D-TIT7	Linear	7	8.5 x10 ⁻³	82
Tau PHF6*	Microtubule-associated protein	6	MMD3	Linear	12	<i>e</i>	83
			MMD3rev	Linear	12	<i>e</i>	
			p-NH	Linear	12	<i>e</i>	84
			TD28	Linear	12	<i>e</i>	85
Immunoglobulin variable domain of TIGIT	Immune checkpoint	119	D-TBP-3	Linear	12	5.6	86
Epidermal growth factor	Mitogenic factor	53	D-PI_4	Linear	12	54	87
Fibroblast growth factor-inducible 14 CRD	TWEAK (tumour necrosis factor-like weak inducer of apoptosis) receptor	43	D-FNB	Linear	12	0.28	88
Aβ-1-42	Monomeric precursor of AB oligomers and fibrils	42	Mosd1	Linear	12	<i>e</i>	89
Immunoglobulin variable domain of PD-L1	Programmed cell death protein 1	124	D-PPA-1	Linear	12	0.51	90

VEGF-A	Vascular endothelial growth factor	102	RFX-V1a2a	Bivalent scaffold	53 + 58	8×10^{-4}	91
			RFX001	GB1 scaffold	56	8.5×10^{-2}	63
gp41 N-trimer pocket mimic	HIV envelope protein ectodomain	42	PIE12-trimer	Flanking disulfide cyclic	8 ^f	^e	92
MDM2	Oncogenic E3 ubiquitin ligase	85	D-PMI α	Linear	12	5.3×10^{-2}	93

^a Most potent binder from phage panning experiments presented, ^b Total amino acid length of synthetic protein enantiomer target, ^c Amino acid length of peptide binder identified through phage display, ^d reported dissociation constant of D-peptide binder to native L-target used in phage display, ^e K_d not reported, ^f length of original, un-crosslinked peptide identified through phage display.

D-Peptides as potential anticancer lead candidates

Growth factor proteins and/or their receptors are overexpressed in many types of cancer,⁹⁴ and hence development of their antagonists can hinder malignant tumor growth as a form of treatment in cancer therapy.⁹⁵ Consequently, enantiomeric segments of both the epidermal growth factor (EGF) and vascular endothelial growth factor (VEGF-A) were synthesized for MIPD.^{63, 87} A 12-residue linear D-peptide ligand for EGF, D-PI₄, was identified with both binding affinity and half-maximal inhibitory concentration (IC₅₀) in micromolar range.⁸⁷ In the case of VEGF-A, the mini-protein GB1 was used as a template scaffold to create a 56 residue D-mini-protein RFX001.D that has a binding affinity as low as 85 nM.^{63, 96} A heterochiral protein complex between a vascular endothelial growth factor (L-VEGF-A) and a D-protein antagonist was also solved using racemic protein crystallography (Figure 1.4 E), providing a foundation for structure-based optimization of the D-protein antagonist.⁶³ Upon optimization, the binder RFX037.D was created, increasing both binding affinity ($K_D = 6$ nM vs 85 nM) and thermal stability ($T_m > 95$ °C vs 33 °C).⁹⁷ Of note, RFX037.D was non-immunogenic in mice, whereas the L-enantiomer generated a strong immune response. In an extension of the MIPD against VEGF-A, bivalent D-protein ligands were also developed using orthogonal MIPD assays with two different scaffold mini-proteins (53 and 58 residues).⁹¹ The two best scaffolds were connected via a covalent linkage to yield the bivalent D-protein RFX-V1a2a, with sub-nanomolar ($K_D = 0.8$ nM) affinity for VEGF-A.

Other key targets for cancer treatment are immune checkpoints,⁹⁸ which are often suppressed by cancer cells to avoid recognition by the innate immune system. The immunoglobulin-like variable (IgV) domains are known to govern immune checkpoints, and thus enantiomeric counterparts of the IgV domains of the programmed-cell death protein ligand 1 (PD-L1, 124 residues) and the T-cell immunoglobulin and ITIM domain (TGIT, 119 residues) were synthesized for MIPD.^{86, 90} After five rounds of bio-panning, binders with micromolar affinity and IC₅₀ were achieved, presenting themselves as promising drug candidates.⁹⁰ One specific binder D-TBP-3 demonstrated proteolytic stability and, importantly, the ability to penetrate through tumor tissue in mice which resulted in tumor suppression.⁸⁶

D-peptides as lead preventive therapeutic candidates

The development of potent D-peptide antagonists of the HIV-1 envelope protein gp41 was shown to prevent viral fusion and entry into cells.^{92, 99} A trimeric version of one of the isolated candidates could block pocket-specific viral entry with an IC₅₀ as low as 250 pM.⁹⁹ Further pharmacokinetic optimization, synthetic scale-up and reduction of production cost yielded the cholesterol-conjugated trimeric D-peptide CPT31,¹⁰⁰ which is currently in Phase Ia clinical trials for the treatment of HIV.

In another study, MIPD was used to create D-peptides with high affinity for the microtubule-binding protein Tau, preventing self-aggregation in the treatment of tauopathies.^{84, 85} The hexapeptides PHF6 (VQIINK) and PHF6* (VQIVYK) were found to promote Tau aggregation in tauopathies such as Alzheimer's disease, and their enantiomers have been used as a target for MIPD.⁸³⁻⁸⁵ Notably, two peptide candidates, MMD3 and MMD3rev, demonstrated cell-penetrating properties, with the ability to cross the cell membrane of neurons.⁸³

Remarks

Despite all the research efforts, there is no D-peptide therapeutic that has yet reached the market. To our knowledge, CPT31 is the only D-peptide candidate that has entered early-stage clinical trials, and the estimated success rate of bringing a binder from phase I to approval is 14%.¹⁰¹ Discovery of D-peptide binders remains challenging hampering downstream clinical research and product development. The ultimate challenge of MIPD lies within the preparation of the enantiomeric protein target. Not only can size be a concern, but both the PTM and protein folding status can also pose major synthetic challenges (see section 3 for synthetic approaches). Except for the Tau targeting peptides, many targets are restricted to extracellular protein domains, as cell-penetrating properties of peptide binders are often weak. In addition, the use of MIPD to identify competitive antagonists is limited by the arbitrary selection of off-target binders. Efforts have been directed to computation-based approaches with the goal to replace the tedious synthesis with *in silico* studies, including the docking of mirror-image helices derived from the PDB;^{102, 103} screening of D-tri/tetra peptides against the target active site;¹⁰⁴ virtual affinity maturation based on existing heterochiral structures.¹⁰⁵ However, existing *in silico* methods suffer from a lack of library diversity and polypeptide binder size, limiting the best example to a binding affinity of 20 μ M.¹⁰⁴

1.2 The current state of the art in D-polypeptide preparation

The most common issue encountered in enantiomeric protein research surrounds their preparation. Since polypeptides entirely composed of D-amino acids cannot be made recombinantly at the time of writing, they must be prepared by chemical synthesis and the current state-of-the-art is summarized as below:

1.2.1 Solid phase peptide synthesis (SPPS)

Allowing stepwise addition of protected amino acids on an insoluble polymer support, solid phase peptide synthesis (SPPS) facilitates access to D-polypeptide chains.¹⁰⁶ Fmoc-protected amino acids have gained popularity over the past two decades as they facilitate the use of milder cleavage conditions.¹⁰⁷ Efficient reagents that allow high conversion of amino acid coupling have been reported.¹⁰⁸ The systematic nature of SPPS has also led to automated systems.¹⁰⁹ When paired with microwave irradiation, each amino acid coupling cycle can be performed in four minutes.¹¹⁰ Recently, a fully automated system for SPPS, where the assembly is complete in a flow system has been developed, yielding complete coupling cycles in less than two minutes, and a full protein up to 164 residues has been assembled.¹¹¹ However, a major drawback is its requirement for a large excess of amino acids (6-60 equivalents), a major financial burden when it comes to D-polypeptide synthesis. This is especially the case when the target proteins contain diastereomeric D-isoleucine, which is significantly higher in cost than other building blocks.

1.2.2 Chemical ligation

To bring down the cost, convergent synthesis of proteins through the assembly of smaller polypeptides by chemical ligation has been achieved.^{112, 113} In general, two peptide fragments are chemically brought together by reacting two latent reaction motifs located at the termini. Various methods for chemical protein ligation have been developed, although some suffer drawbacks when preparing polypeptides in opposite chirality due to requiring modified amino acids (Table 1.3, Entries 1-4). On the other hand, native chemical ligation (NCL) and serine-threonine ligation (STL) employ canonical cysteine or serine/threonine residues, though the latter is yet to be reported in D-protein synthesis. NCL is commonly used whereby the thiol group of an *N*-terminal cysteine peptide and a *C*-terminal thioester of a reacting pair undergo *trans*-

thioesterification, followed by an S-to-N acyl shift to yield a traceless native peptide bond (Figure 1.7).

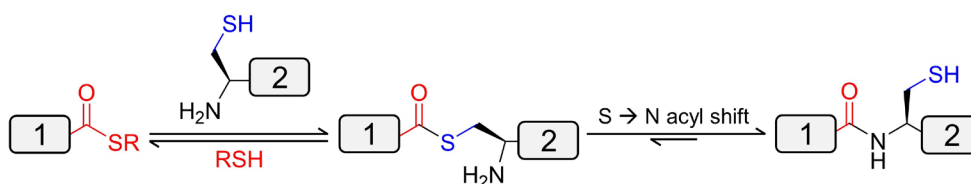


Figure 1.7 Reaction scheme of native chemical ligation.

Thioester preparation

Activated thioesters cannot be anchored at the C-terminus during Fmoc-SPPS due to their sensitivity to base treatment which is routinely used during deprotection. Therefore, numerous efforts have been directed to enable their preparation by Fmoc-SPPS,¹¹⁴ such as the use of thiol labile safety catch linkers.^{115, 116} A commonly employed approach utilizes the refined Dawson linker (Table 1.3, Entry 5). Following SPPS, the base-stable 3,4-diaminobenzoic acid (Dbz) derivative is activated and subjected to thiolysis to yield the thioester for native chemical ligation. More recently, a second generation Dawson linker recruits an *N*-methylated amino group minimizing unwanted acylation when an excess of glycine reagent is used.¹¹⁷ Assembly of peptides directly onto the linker amine is a simple and useful feature of the Dawson linker. Alternatively, the C-terminal peptide hydrazide can be used to create a C-terminal thioester (Table 1.3, Entry 4).¹¹⁸ Activated by the addition of sodium nitrite, the hydrazide can be oxidized into an acyl azide which can be subjected to thiolysis *in situ*.¹¹⁹ Hydrazine resins are prepared fresh before use, but it has recently been shown that resins can be prepared with Fmoc-hydrazine facilitating long term storage and facile loading quantification.¹²⁰

In the convergence of multiple peptide fragments to prepare larger synthetic proteins, some central fragments will be ligated at both *N*- and *C*-termini. To prevent intramolecular side-reactions (cyclization), the peptide can be prepared as a thioester surrogate, whereby a C-terminal functional group remains inert in NCL and can be subsequently activated for the next ligation step. The peptide hydrazide is the most common choice in this instance due to the facile *in situ* activation and has been used to prepare numerous D-proteins, including the mirror-image TIGIT domain discussed earlier.⁸⁶ Other methods for generating thioester surrogates, such as

SEA chemistry (Table 1.3, Entry 7),¹²¹ could also be used but have yet to be applied to D-protein synthesis.

N-terminal cysteine protection

In the pursuit of more complex protein targets, sequential NCL steps will require orthogonal *N*-terminal cysteine protecting groups for middle segments (Table 1.3, Entries 8-12). A common and reliable choice is the protection of cysteine in a thiazolidine ring (Thz), due to the facile conversion into cysteine, compatibility with ligation conditions and its commercial availability in both enantiomeric forms.¹²² However, Thz was found to be unstable under the oxidation conditions required for peptide hydrazide oxidation.¹¹⁸ Numerous efforts have been directed to the development of other orthogonal, *N*-terminal cysteine protecting groups to facilitate sequential peptide hydrazide ligations (Table 1.3, Entries 9-12). Preparation of their enantiomeric counterpart is theoretically simple. Indeed, TFA-Thz, which is stable to hydrazide oxidation, was used in the preparation of enantiomeric polymerase D-Dpo4¹⁴ and ubiquitin.⁴⁸

NCL beyond cysteine

Due to the low abundance of cysteine residues in proteins (<2%),¹²³ significant efforts have been directed to link peptides with alternative amino acids.¹²⁴ Desulfurization converts reactive cysteine thiol CH₂SH to the CH₃ group of alanine,¹²⁵ which is significantly higher in abundance (>8%).¹²⁶ Techniques include hydrogenation over Pd/Al₂O₃ and Raney nickel.¹²⁵ Given the issues associated with purity and the product recovery using metals, a metal-free desulfurization alternative has been developed through a free-radical mechanism (Table 3, Entry 14; Figure 1.8 A). However, native cysteines must be protected during desulfurization reactions. Acetamidomethyl cysteine (Cys(Acm)) is commonly employed in D-protein synthesis, and is commercially available in both L- and D- enantiomers for Fmoc-SPPS (Figure 1.8 B).¹²⁷ Following desulfurization, the Acm group can be removed using mercury acetate,¹³⁷ iodine,¹³⁸ or more recently, PdCl₂.¹²⁸

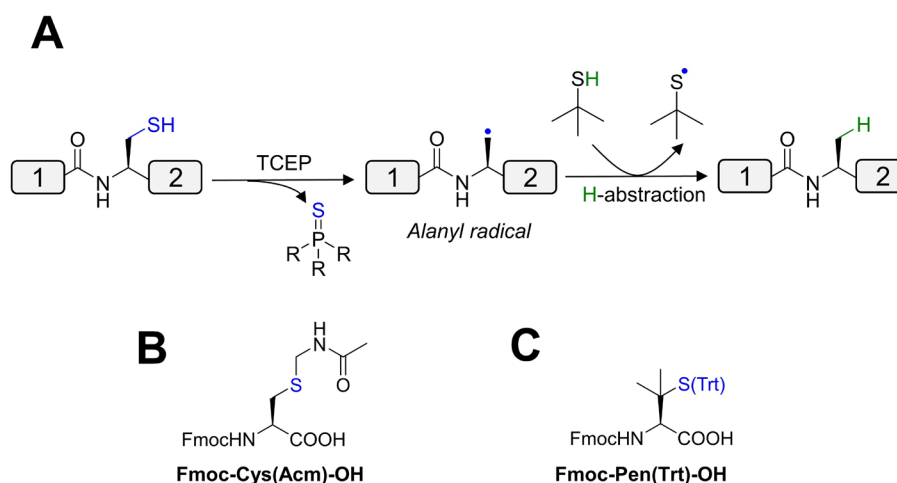


Figure 1.8 (A) Mechanism of metal-free radical desulfurization, (B) acetamidomethyl cysteine (Cys(Acm)) and (C) penicillamine buildings blocks for Fmoc-SPPS.

Thiols may also be inserted into other canonical amino acids and can be removed by desulfurization (For review, see ¹²⁹). However, this approach has not been applied in D-protein synthesis because of challenges associated with synthesizing the enantiomeric building blocks (Table 1.3, Entry 15). Commercially available D-penicillamine (D-Pen) may be a viable option for ligation at valine (Figure 1.8 C),¹³⁰ but this residue is associated with slow ligation kinetics with all C-terminal thioester sites other than glycine.¹³¹ NCL can also proceed by employing a temporarily inserted thiol auxiliary which can be removed by acid cleavage.¹³² Such auxiliaries enable the generation of the lysine isopeptide bonds in the synthesis of branched ubiquitin chains (Table 1.3, Entry 25).^{34, 49} Nevertheless, this approach has not yet been applied in the synthesis of the mirror-image D-counterparts (see Section 1.2.3).⁴⁹

One-pot approach

During the synthesis of large protein targets, it becomes advantageous to conduct steps in a ‘one-pot’ fashion minimizing lengthy and yield-reducing purification maneuvers. For example, directly after a NCL reaction, the desulfurization step can be performed in one pot, followed by Cys(Acm) deblocking without intermediate purification.¹²⁸ A primary limitation of performing desulfurization immediately after NCL is that: a large excess of thiol catalyst such as 4-mercaptophenylacetic acid (MPAA) is needed to improve ligation rates but they also quench the desulfurization reaction.¹³³ Efforts have been directed toward finding new thiol catalysts that are

compatible with desulfurization (Table 1.3, Entries 16-18). Methyl thioglycolate was found to not interfere with desulfurization whilst providing good catalytic properties, and it has been used in the synthesis of mirror-image ubiquitin for racemic protein crystallography studies.⁴⁸ Other efforts to reduce the number of HPLC purification steps include one-pot ligations of numerous fragments and performing chemical ligations on a solid support.^{64, 134}

Enhancing solubility of hydrophobic peptide fragments

The nature of protein has implications on the experimental design. Hydrophobic proteins such as membrane proteins can suffer from aggregation and poor solubility.¹³⁵ A solution that can address these issues is to recruit first-generation removable backbone modification (RBM), which minimizes aggregation whilst allowing conjugation to a poly-arginine solubility tag (Table 1.3, Entry 21).¹³⁶ However, the RBM is installed via a removable glycine auxiliary and thus has limited scope. A second-generation RBM was designed to be installed into all other amino acids, including the challenging Val-Ile junction, making this highly attractive for the synthesis of membrane proteins (Table 1.3, Entry 22; Figure 1.9).¹³⁷ The RBM tags have been employed in the synthesis of the mirror-image TIGIT membrane protein domain.⁸⁶ In addition to RBMs, removable solubilizing tags could also be incorporated onto lysine side chains (Fmoc-Ddae-OH) or (Fmoc-Ddap-OH) using the ‘helping-hand’ strategies (Table 1.3, Entries 19-20).^{138, 139} Whilst no D-proteins are currently reported using this method, these tags could possibly be installed onto commercially available D-amino acid building blocks. The lysine tags can also be employed to install click handles for templated chemical protein ligations.¹⁴⁰

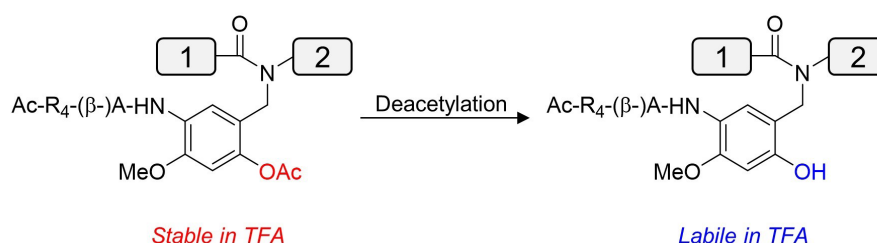


Figure 1.9 Second generation removable backbone modification. 4-methoxy-5-nitrosalicylaldehyde is installed onto backbone nitrogen during Fmoc-SPPS, followed by assembly of peptide main chain and desired tag sequence. Reversible acetylation of phenol group controls TFA lability for tag removal.

Remarks

Since the advent of chemical ligation methods, protein synthesis has transformed into a rapidly evolving field of research. Particularly, native chemical ligation has facilitated the synthesis of numerous D-proteins: enantiomeric venom toxins,^{37, 41-43, 45, 52, 55, 57, 59, 64} growth factors,^{63, 87, 91, 97} and enzymes,^{8, 10, 14, 17-19, 141, 142} including the 90 kDa split-enzyme D-*Pfu*.¹⁵ The pursuit of more complex D-proteins is perplexed by their size and post-translational modification status. Multi-segment, one-pot approaches can improve efficiency, but many recruit unique amino-acid reagents that need to be prepared in the laboratories.¹⁴³ Another issue involves the rates of ligation reactions, which can potentially be increased by adopting selenocysteine NCL,¹⁴⁴⁻¹⁴⁷ and template-directed chemical ligations.¹⁴⁸ Easier access to protected building blocks, and their commercial availability, will greatly improve the possibilities for D-protein chemical synthesis. Design of synthetic routes to D-proteins can also be assisted by computational approaches such as the open-source 'Alligator' tool.¹⁴⁹ Perhaps, the ultimate goal for D-polypeptide production is to completely circumvent chemical synthesis, with the entire mirror-image translational machinery (ribosome, rRNA, tRNA, tRNA synthetase, etc) served as a replacement.

1.2.3 Transformation from D-polypeptide to mirror-image protein

A somewhat overlooked challenge is the complexity of protein folding that researchers may need to address during synthetic protein preparation.¹⁵⁰ Typically, a solution of the protein in chaotropic conditions such as guanidine or urea is prepared and then diluted into a refolding buffer.^{151, 152} Many larger proteins require chaperones for efficient folding, particularly *in vivo* where direct control of refolding conditions is limited. Recently, it was shown that the GroEL/ES chaperone protein can efficiently fold both enantiomers (L- and D-) of synthetic DapA protein.¹⁴² This observation suggested that the protein chaperone activity can be achiral and may find broad utility in the pursuit of mirror-image life systems. Two specific challenges associated with protein folding that will be discussed here include correct oxidation of disulfide bonds and installation of post-translational modifications.

Disulfide bond formation

Correct folding of the disulfide bonds is case-dependent and often requires extensive screening and optimization for each protein. One common method involves diluting the reduced D-polypeptide chain from chaotropic agents in the presence of reduced and oxidized thiols as redox reagents, as reported in the synthesis of D-rC5a.⁴² However, misfolded D-protein often arises as thermodynamically trapped by-products containing mismatched disulfide bonds and adducts with thiol reagents.^{37, 52} Oxidation by air or DMSO has been used, following careful optimization of buffer additives, reagent concentrations, and pH. However, this method often results in low yields (typically <50%), requires large solvent volumes, and proceeds over several days.^{37, 52, 56} To gain additional control, orthogonal cysteine protection followed by pairwise cysteine oxidation may be used. Recently, an orthogonal cysteine protection scheme was developed, encoding a rapid system for disulfide oxidation (Table 1.3, Entry 23). Using palladium and UV light mediated deprotections, it was shown that up to three correctly paired disulfide bonds could be formed in less than 13 minutes.¹⁵³ The use of trityl (Trt) and acetamidomethyl (Acm) protecting groups has been reported for formation two disulfide bonds in a D-protein (Table 1.3, Entry 23).⁵⁹ Another protecting group that can be used is the 2-nitrobenzyl group but must be chemically synthesized (Table 1.3, Entry 23).¹⁵³ Disulfide bond formation can also be directed without the use of chaotropic agents or orthogonal protection schemes, such as cysteine-penicillamine pairings or repeat-proline (Cys-Pro-Pro-Cys) motifs.^{154, 155}

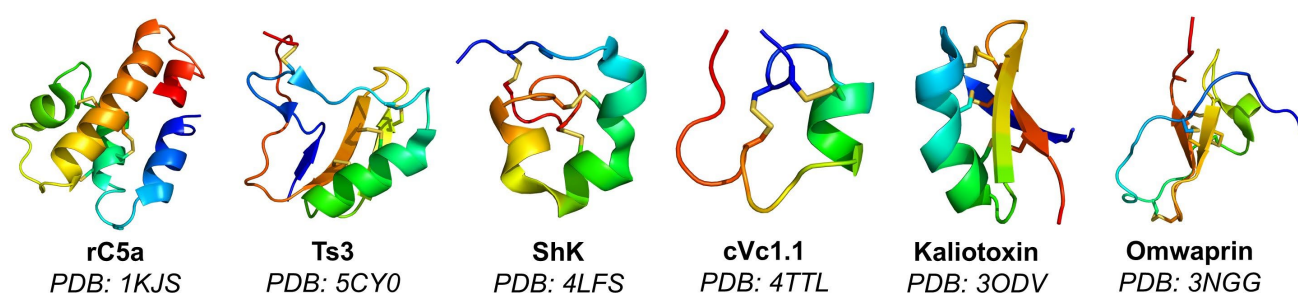


Figure 1.10: Examples of disulfide-rich D-proteins prepared by chemical synthesis, with structures resolved by racemic protein crystallography.

Post-translational modifications

To achieve a true, mirror-image protein with reciprocal stereospecific protein activity, PTMs must be incorporated into the D-polypeptide product. Serine and tyrosine phosphorylation are common PTMs.¹⁵⁶⁻¹⁵⁸ Whilst the L-serine equivalent (Table 1.3, Entry 34) is commercially available, the corresponding protected D-Ser building block must be accessed through chemical synthesis.²³ This has been demonstrated in the synthesis of mirror-image ribosomal proteins which are essential in the binding of L-RNA molecules.²³ Phosphorylation and sulfation of tyrosine residues have been incorporated into chemical synthesis of L-proteins^{159, 160} using similar protected building blocks (Table 1.3, Entry 33 & 35). It is reasonable that equivalent building blocks can be prepared in D-enantiomeric form. Indeed, tyrosine phosphorylation is a PTM observed in many transcriptional regulators and has been shown to impact DNA binding.¹⁵⁸ Glycosylation is another widespread PTM found in proteins.¹⁶¹ However, the sugars attached to the protein are chiral existing almost exclusively in D-form.¹ Mirror-image glycoprotein would require installation of L-sugar polymers onto the D-polypeptide and has not yet been reported. However, quasi-racemic protein crystallography of synthetic glycoprotein Ser-CCL1 could be facilitated using the un-glycosylated D-enantiomer.⁶⁰ In many cases, glycans can play important roles in protein or nucleotide binding,¹⁶¹ and therefore achieving mirror-image protein glycosylation is a key milestone in research surrounding D-proteins, particularly in the area of MIPD.

Other promising PTMs include lysine trimethylation or acetylation, both of which have been incorporated into the chemical synthesis of L-histones (Table 1.3, Entries 26-27).^{162, 163} Preparation of the corresponding D-histones would require the synthesis of the lysine building blocks in D-form. A more challenging lysine PTM is site-specific ubiquitination.¹⁶⁴ Chemical synthesis of branched ubiquitin chains in native L-form is well reported, typically employing a δ -mercapto lysine to facilitate ligation, followed by desulfurization.^{157, 164-167} However, preparation of the branched ubiquitin chains for racemic protein crystallography utilized an auxiliary for NCL, following subsequent removal with TFA to generate a native glycine at the branched ligation site (Figure 1.11 A).^{34, 49} The auxiliary facilitates NCL with a ubiquitin thioester, following subsequent removal with TFA to generate a native glycine at the branched ligation site. However, during preparation of the mirror-image branched ubiquitin proteins, the auxiliary was replaced with a cysteine residue to circumvent the low efficiency of the ligations, leaving a non-native cysteine as a scar at the ligation site (Figure 1.11B).^{34, 49}

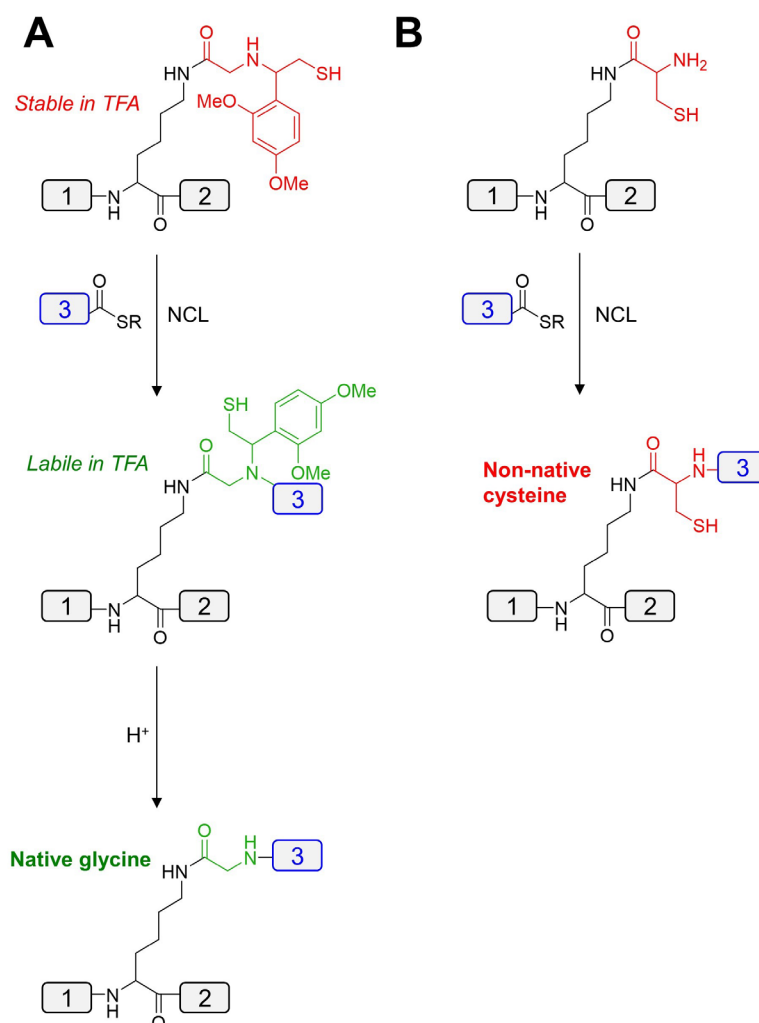


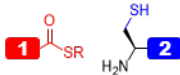
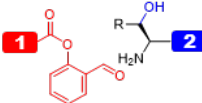
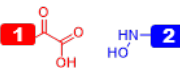
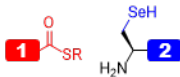
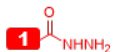
Figure 1.11. Methods for branched protein ligation used in preparation of poly-ubiquitins; **(A)** glycyl auxiliary mediated NCL for preparation of branched L-proteins and **(B)** cysteine mediated NCL for preparation of branched D-proteins.

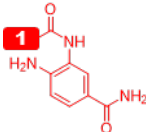
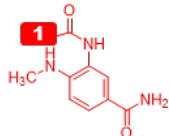
Palmitoylation also serves as a promising PTM for incorporation into D-proteins and has been applied to the synthesis of cysteine palmitoylated L-proteins (Table 1.3, Entries 31-32).^{168, 169} Palmitic acid is achiral and could be conjugated to D-cysteine or other D-amino acid residues. Synthesis of palmitoylated D-proteins may find use in racemic protein crystallography³¹ which has been applied to reveal the structure of transmembrane protein domains in racemic detergents.^{32, 33} In addition, palmitoylation is a useful modification for pharmacokinetic optimization of peptide drugs,¹⁷⁰ including the FDA-approved Liraglutide.¹⁷¹ This modification may find use in prolonging the half-life of D-peptide drug candidates, as encountered with the conjugation of cholesterol to the promising D-peptide drug candidate, CPT31.¹⁰⁰

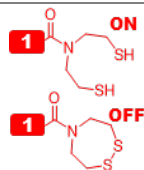
Remarks

Folding of D-polypeptide chains into the desired protein conformation is a challenging task, and often requires optimization in a case dependent manner. Protein refolding protocols can be pre-established using the native recombinant L-counterparts,¹⁷² and they are often sufficient to fold the corresponding D-enantiomer. In the case where disulfide bond formation is involved, oxidants such as cystine or glutathione disulfide can be used.⁵⁹ Alternatively, orthogonal protection schemes can be implemented.^{59, 153} Most PTM installation can be achieved, particularly the achiral components such as phosphorylation.²³ However, chiral PTMs such as glycosylation largely remains an unresolved synthetic challenge.⁶⁰ A plausible solution would be to obtain a mirror-image enzyme (such as *endo*-glycoside hydrolase),¹⁷³ capable of assembling the necessary polysaccharide building block from L-sugars, much like the enantiomeric polymerase enzymes discussed earlier.^{14, 15, 18, 19}

Table 1.3. Synthetic methods used in chemical protein synthesis, indicating use in reported D-protein synthesis and potential issues encountered with use.

#	Entry Synthetic method Ligation method	Reagents	Used in D-protein synthesis		Ref
			Y/N	Potential issues	
1	Native chemical ligation	 + thiol catalyst	Y	Dependence on suitable cysteine or alanine residues.	112
2	Serine/threonine ligation	 + acidolysis	N	Requires suitable Ser/Thr. Slower reaction kinetics than NCL.	174
3	KAHA ligation		N	Accessibility of enantiomeric reagent.	175
4	Selenocysteine NCL	 + thiol catalyst	N	Accessibility of enantiomeric reagent.	144
NCL reactive end					
Thioester surrogate					
4	Hydrazides	 <u>Activation</u>	Y	Oxidation incompatible with Thz. Low temperature activation needed (<-15 °C).	119

		+ NaNO ₂ + Thiol			
5	Dbz	 <p>Activation 4-nitrophenyl chloroformate or NaNO₂ or isoamyl nitrite</p>	Y	Di-acylation side product with excess Gly.	21, 176, 177
		+thiol			
6	MeDbz	 <p>Activation 4-nitrophenyl chloroformate</p>	N	Difficult to activate off-resin.	117
		+thiol			
7	SEA		N	SEA _{OFF} peptide can remain inactive in ligations for one-pot assembly of multiple fragments and activated by TCEP for thioester formation (SEA _{ON}). Therefore, TCEP should be avoided until the desired ligation step.	178



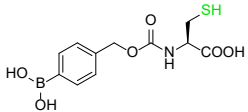
Activation of SEA_{OFF}

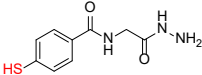

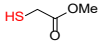
TCEP

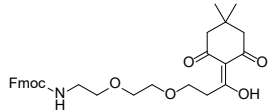
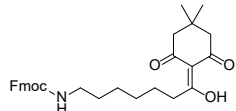
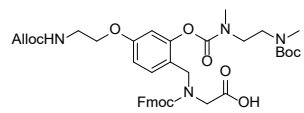
+thiol

N-cysteine protection

8	Thz	 <u>Deprotection</u> MeONH ₂	Y	Incompatible with hydrazide oxidation.	122
9	Cys(Tfacm)	 <u>Deprotection</u> pH 11.5	N	Accessibility of enantiomeric reagent.	179
10	TFA-Thz	 <u>Deprotection</u> Base then MeONH ₂	Y	Accessibility of enantiomeric reagent.	48
11	N ₃ -Cys		N	Accessibility of enantiomeric reagent.	180

		<u>Deprotection</u>			
		TCEP			
12	Cys(Dobz)		N	Accessibility of enantiomeric reagent. Harsh deprotection conditions.	181
		<u>Deprotection</u>			
		H ₂ O ₂			
Desulfurization					
13	Metal-based	Pd/Al ₂ O ₃ Or Raney Nickel + H ₂ (g)	Y	Removal of metal impurities can be problematic. Use of hydrogen gas. Potential side reactions with Trp and Met Quenched by thiol catalyst. Native Cys must be protected.	127
14	Metal-free radical based	VA-044 TCEP tert-butylthiol	Y	Quenched by thiol catalyst. Native Cys must be protected.	125
15	Beta/gamma thiol amino acids	β-thiol-Phe β-thiol-Val β-thiol-Leu β-thiol-Asp β-thiol-Asn β-thiol-Arg γ-thiol-Val γ-thiol-Thr γ-thiol-Ile	N	Accessibility of enantiomeric reagent. Commercially available D-Penicillamine (β-thiol-Val) could be used for D-peptide ligation at Val, if directly following a glycine residue.	182

		γ-thiol-Pro γ-thiol-Glu γ-thiol-Gln γ-thiol-Lys 2-thiol-Trp			
Thiol catalysts for one-pot ligation-desulfurization					
16	MPAA-hydrazide	 $pK_a = 6.6$ <u>Removal</u> Aldehyde-resin capture	N	Preparation of MPAA-hydrazide reagent coupled with use in large excess is uneconomical.	183
17	Trifluoroethanthiol	 $pK_a = 7.3$ <u>Removal</u> Evaporation (bp = 37 °C)	N	Malodorous and volatile, though could be used for D-protein synthesis.	184
18	Methyl thioglycolate	 $pK_a = 7.9$ <u>Removal</u>	Y	Slower kinetics with C-terminal beta-branched residue.	48

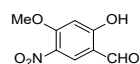
		none			
	Solubility enhancers				
19	Helping hand v1	 <p><u>Installation</u> Amine labelling with lysine side chain</p> <p><u>Removal</u> Hydrazine (aq)</p>	N	Additional steps to incorporate and remove tag. Potential issues with stability.	138
20	Helping hand v2	 <p><u>Installation</u> Amine labelling with lysine side chain</p> <p><u>Removal</u> Hydrazine (aq) Or Hydroxylamine (aq)</p>	N	Additional steps to incorporate and remove tag.	139
21	Removable backbone modification v1		Y	Limited to Gly only. Lengthy synthesis of building block. Additional steps to incorporate and remove tag.	136, 185

Installation

Standard Fmoc-SPPS

Removal

pH 7 then TFA



22

Removable backbone
modification v2Installation

Reductive amination

Acetylation

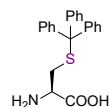
Y

Additional steps to incorporate
and remove tag.

86, 137

RemovalDeacetylation (Cys_(aq))
then TFA**Protein folding**

23

Cysteine orthogonal
protection**Disulfide #1**

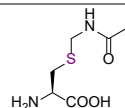
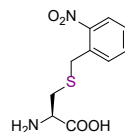
Y

Practically limited to two disulfide
bonds. Accessibility of a third,
orthogonally protected D-Cys
building block.

59, 153

Removal

TFA

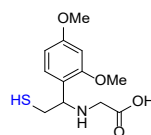
**Disulfide #2**RemovalIodine
or PdCl₂**Disulfide #3**Removal

UV light (350 nm)

24	“Ambidextrous” chaperone	GroEL/ES protein chaperone	Y	Mostly unnecessary for <i>in-vitro</i> protein folding. Limited scope reported.	142
----	--------------------------	----------------------------	---	---	-----

Post-translational modifications

25	Lys ubiquitination
----	--------------------

**N**

Low efficiency of ligation. Glycyl auxiliary replaced with Cys in preparation of enantiomeric di- and tri-ubiquitin proteins.

34, 49

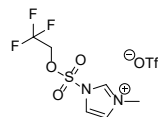
<p style="text-align: center;"><u>Installation</u> Coupling to lysine side chain <u>PTM</u> NCL to Ub-thioester <u>Auxiliary removal</u> TFA</p>					
26	Lys trimethylation	Fmoc-Lys(Me ₃)-OH	N	Accessibility of enantiomeric reagent.	162, 181
27	Lys acetylation	Fmoc-Lys(Ac)-OH	N	Accessibility of enantiomeric reagent.	163
28	Asn N-Glycosylation	Fmoc-Asn(Glycan)-OH or Boc-Asn(Xan)-OH (and) further glycosylation on-resin or in solution.	N	Accessibility of enantiomeric reagent (would also require L-sugars).	60, 186, 187
29		Oligosaccharide coupled directly to free Asn side chain during Boc-SPPS.	N	Accessibility of enantiomeric reagent (would also require L-sugars).	188
30	Thr O-Glycosylation	Fmoc-Thr(Glycan)-OH	N	Accessibility of enantiomeric reagent (would also require L-sugars).	160
31	Cys S-palmitoylation	Fmoc-Cys(Mmt)-OH Mmt removal on-resin with 2% TFA Reaction with palmitic anhydride	N	Incompatible with NCL. Potentially viable for D-protein synthesis via STL or Sec NCL.	168
32		Fmoc-Cys(palmityl)-OH	N	Incompatible with NCL. Fmoc-D-Cys(palmityl) must	169

be synthesized.

Fmoc-Tyr(OTBS)-OH

Deprotection and sulfation on-resin with:

33 Tyr sulfation



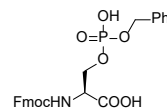
N

Accessibility of enantiomeric reagent.

160

+DIEPA

34 Ser phosphorylation

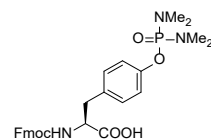


Y

Accessibility of enantiomeric reagent.

23, 157, 166,
189

35 Tyrosine phosphorylation



N

Accessibility of enantiomeric reagent.

159

1.3 Conclusion

Applications of synthetic D-protein enantiomers are vast but remains to be challenged by the difficulty in their preparation, particularly when their complexity increases with their size, folding and post-translational modifications. The ability to translate L-mRNA into D-proteins will be truly revolutionary. The first milestone will be the complete assembly of an enantiomeric ribosome, comprised of D-proteins and L-rRNAs. Since three out of ~50 mirror-image *E. coli* ribosomal proteins²³ and efficient L-nucleotide polymerases^{10, 14, 15, 17-19, 21} have been reported, it is anticipated that this ambition will be achieved in the future. Mirror-image translation could then be achieved *in vitro*,¹⁹⁰ using D-aminoacyl-tRNA synthetases to load the L-tRNAs along with the mirror-imaged translational factors. Subsequently, all D-proteins needed in mirror-image life, racemic protein crystallography and D-targets for mirror-image phage display may be obtained by supplying the exogenous L-nucleic acids encoding the desired protein. Perhaps, preparation of entire D-polypeptides can also be achieved using “Flexizyme” technology,¹⁹¹ which has been reported to incorporate D-amino acids into peptide chains without using enantiomeric translation components.^{5, 75, 76}

Chemical synthesis remains superior at atomistic control allowing researchers to incorporate building blocks without constraints associated with ribosome-based systems. Complex D-protein targets can be achieved via: engineering of split enzymes,¹⁵ mutational installation of suitable ligation sites¹⁵ and *in silico* design of accessible D-enzymes.²¹ For protein crystallography, small D-proteins can be readily accessed for facile investigations though racemic protein crystallography. To facilitate crystallization of complex L-proteins, smaller D-proteins may be used to facilitate crystallization of complex L-proteins by quasi-racemic protein crystallography.³⁴ In addition, mirror image phage display can be conducted against minimal protein targets, circumventing large enantiomeric target preparation.^{82-84, 92, 99, 100} Discovery of D-peptide binders could also be achieved via computational-based approaches.^{102-104, 192} Together, while there are unsolved challenges, D-polypeptide research remains to have strong potentials that can generate explosive impacts on numerous research topics. crystallography,

1.4 Thesis aims

The broader use of D-protein and D-peptide technology in applied research can be expedited by mitigating the challenges of their preparation. For example, racemic protein crystallography is a powerful tool at unveiling molecular insights at exquisite resolution, but obtaining sufficient quantities of D-enantiomer for crystallization screening can be inefficient. Additionally, preparing D-enantiomers of complex protein targets for mirror-image phage display screening remains a challenging task. The presented thesis aimed to harness both technologies in applied research, whilst mitigating inefficient D-protein preparation. Two key objectives were outlined:

1. **Enable in-depth racemic protein crystallography analysis to study membrane-active bacteriocins.** Highly efficient synthetic routes to the D-enantiomers of two key bacteriocins, aureocin a53 and lacticin Q were developed. The consumption of the costly D-amino acids was minimized (two to four equivalents), whilst producing sufficient quantity of D-protein in a one-pot procedure for in depth analysis by racemic protein crystallography. Novel insights were obtained into the roles of tryptophan in bacteriocin function, and subsequently, the new synthetic route could facilitate site specific substitutions of all tryptophan residues in the two bacteriocins for activity analysis.
2. **Develop an efficient route to discover D-peptide inhibitors of TNFR-1.** A minimal cytokine binding domain of the tumour necrosis factor receptor 1 (TNFR-1) was identified as a potential target for the discovery of D-peptide inhibitors. Chemical synthesis of the subdomain, followed by optimization and confirmation of the folding by racemic protein crystallography was conducted. The enantiomeric subdomain was first used as a target in mirror-image phage display, mitigating the need to prepare the challenging enantiomeric TNFR-1. Then, the use of computational methods to design D-protein inhibitors of TNFR-1 based only on the target structure was explored.

Chapter 2 : Investigations of the role of tryptophan in membrane-active bacteriocins guided by racemic protein crystallography

2.1. Preface

To understand how membrane-active bacteriocins have evolved to acquire their antibacterial properties, analysis based on racemic protein crystallography (0.89-1.21 Å resolution) of two key representatives, aureocin A53 (AucA) and lacticin Q (LnqQ) was conducted. Distinct salt bridges were revealed on the AucA surface, where tetrahedral oxyanions (sulfate or glycerol 3-phosphate) are hydrogen-bonded to the indole NH of Trp3, Trp31, and Trp40 and to neighbouring lysines. Antibacterial activity assays revealed that conserved Trp31 and Trp40 are critical for AucA activity. Also, enantiomeric D-AucA was found to be at least as active as its native L-counterpart. These observations suggest that Trp31 and Trp40 of AucA bind to the head groups of lipids and facilitate membrane destabilisation. Meanwhile, Trp3 and Trp22 associate at a homodimeric interface to confer proteolytic stability. The homologous bacteriocin LnqQ lacks a Trp3 and appeared to be monomeric, while enantiomeric D-LnqQ was up to 16-fold more active than L-LnqQ, suggesting sensitivity to proteolysis associated with a lack of oligomerization. However, Trp32 and Trp 41 of LnqQ, which are conserved with Trp31 and Trp40 of AucA, also proved important for antibacterial activity - thus reinforcing their importance in lipid interactions. The molecular interactions of tryptophans identified here show how a bacteriocin delivers its antibacterial properties.

2.2. Introduction

Leaderless bacteriocins have been produced by their hosts as a means of eliminating competitors, and so they must withstand harsh conditions and confer potent activity without discriminating against antibiotic-resistant and sensitive strains. Meanwhile, this family of small ribosomal proteins (<6 kDa) have been found in dairy products¹⁹³ and are considered safe for use as food preservatives.^{194, 195} Hence, understanding their mechanism of action can shed light on how to address the emergence of multidrug-resistant pathogens.¹⁹⁶

Two key representatives, aureocin A53 (AucA) and lacticin Q (LnqQ) share 48% homology and contain multiple conserved residues with other potent leaderless bacteriocins.¹⁹⁶⁻¹⁹⁸ Both bacteriocins exhibit broad-spectrum activity against Gram-positive pathogens.^{198, 199} Previous NMR analysis of isotopically-labelled, recombinant AucA and LnqQ revealed a globular morphology and an unusually high number of tryptophan residues (AucA = 5, LnqQ = 4).²⁰⁰

Although some of these residues are highly conserved, their roles remain elusive and hence impede the wider applications of bacteriocins in the medicine and food industries.

In this chapter, a cost-effective synthetic scheme for enantiomeric AucA and LnqQ is reported, such that atomistic insights of these leaderless bacteriocins could be revealed through racemic protein crystallography.^{29, 31} A racemic mixture of each DL-AucA or DL-LnqQ allows rapid crystal formation under a broad range of conditions, and thus their analysis could reveal molecular insights that may otherwise be difficult to obtain via other bioanalytical techniques.^{32, 33, 43, 46, 51, 53, 54} Furthermore, expedient synthetic schemes of AucA and LnqQ offer prompt access to site-specific variants for further investigations, bypassing issues related to inclusion bodies formation, enzymatic treatment and multiple chromatographic steps in the existing recombinant approach.²⁰⁰ The bioanalytical analyses conducted here revealed that the tryptophan residues in AucA and LnqQ have designated roles in conferring antibacterial activity and/or stability.

2.3. Results and discussion

2.3.1 Chemical synthesis and antibacterial activity of aureocin A53, lactacin Q and their enantiomers

Total chemical synthesis of AucA and LnqQ utilizing native chemical ligation was developed (Figure 2.1). *In situ* hydrazide activation, followed by thiolysis yielded the peptide thioester.^{112, 119} Reaction in stoichiometric quantities with a peptide bearing an N-terminal cysteine, followed by thiol removal and desulfurization,¹⁸² yielded the native bacteriocins. This approach facilitated an efficient route to the AucA and LnqQ variants, whereas linear solid-phase synthesis (SPPS) required a large excess of amino acids (10-20 equivalents, see Method 5.3.3, and ref²⁰¹ for AucA), a significant financial burden when D-amino acid building blocks are used. While an Ala11Cys substitution could facilitate both ligation and desulfurization resulting in native AucA, desulfurization at other Ala positions in AucA proved to be unsuccessful likely due to a lack of solvent accessibility (Figure S7.1).

Initial chemical synthesis of L- and D-LnqQ was optimized by Dr. Xuefei Li. Data pertaining to LnqQ presented in this thesis was conducted by Alexander Lander, following redesign of the synthetic route to a one-pot procedure.

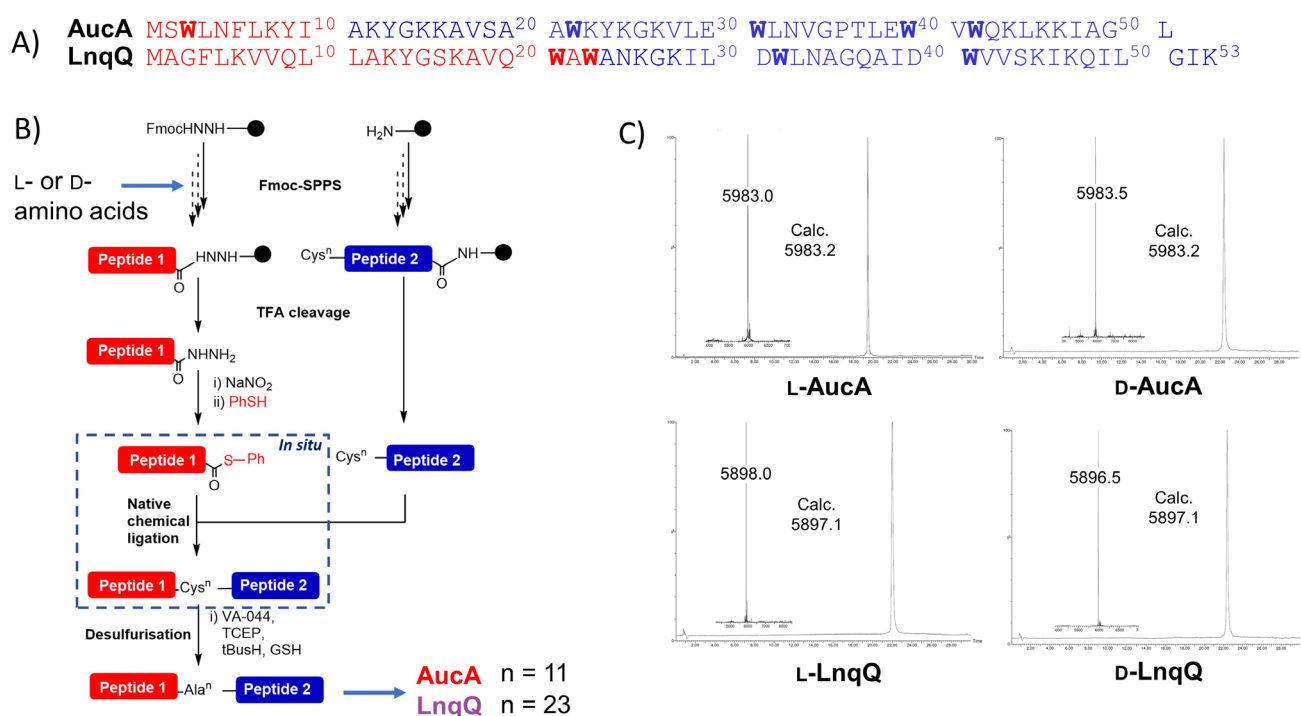


Figure 2.1: **A)** Sequence of AucA and LnqQ, **B)** Ligation scheme for synthesis of AucA and LnqQ, **C)** UPLC traces and deconvoluted high-definition ESI+ MS of isolated L- and D- LnqQ and AucA. ⁿ denotes the residue number of the Ala→Cys substitution used to facilitate ligation.

This synthetic scheme was used to prepare enantiomeric AucA and LnqQ entirely composed of D-amino acids (D-AucA and D-LnqQ) which yielded an opposite optical rotation to the native L-enantiomers by circular dichroism (CD) spectroscopy (Figure S7.2-S7.3). The antibacterial activities of both enantiomers were assessed by determining minimum inhibitory concentrations (MICs) against a panel of nosocomial pathogens isolated from hospitalised patients (CANWARD²⁰²) (Table 2.1). Contrary to a previous study which reported a decrease in activity of enantiomeric membrane-active antimicrobial proteins,²⁰³ D-AucA and D-LnqQ were at least as active as their L-enantiomer while neither enantiomer discriminated between antibiotic-resistant and -sensitive strains. Therefore, these bacteriocins likely target the bacterial membrane through non-stereospecific interactions.¹⁹⁹

Bacteriocin MIC assays were conducted by Laura Dominguez at Concordia University.

In the case of *S. aureus* CW115852, D-AucA displayed an eight-fold increase in activity over that of its native enantiomer. This effect is significantly more pronounced in the case of D-LnqQ, with activity increase (up to 16-fold) against all *S. aureus* and *S. epidermidis* strains. A likely explanation is that the L-bacteriocins are partially degraded during incubation, with L-LnqQ significantly more susceptible to proteolysis than L-AucA. Indeed, proteases are known to be secreted by *Staphylococci* as resistance mechanisms against antibacterial peptides.^{204, 205}

Table 2.1: Minimum inhibitory concentrations^[a] (MIC) of L- and D-bacteriocins and melittin control against bacterial strains.

	<i>S. aureus</i>				<i>S. epidermidis</i>		<i>E. faecalis</i>		<i>E. faecium</i>	
	AC2921 3 ^[b]	CW11412 5 ^[c]	CW11585 2 ^[c]	CW11337 9 ^[c]	CW13161 2 ^[c]	CW13050 0 ^[c]	CW13334 6 ^[c]	CW13300 3 ^[c]	CW13082 6 ^[c]	CW13112 6 ^[c]
L-AucA	4	4	16	4	4	4	4	4	2	2
D-AucA	4	4	2	4	4	4	4	4	2	4
L-LnqQ	16	32	64	32	8	16	4	2	2	2
D-LnqQ	4	8	4	4	4	4	2	2	2	2
Melittin	8	8	4	4	4	8	16	16	8	8

[a] MIC values expressed in µg/mL, [b] AC: American Type Culture Collection strain, [c] CW: CANWARD collection nosocomial strain.

	AC292 13	CW1141 25	CW1158 52	CW1133 79	CW1316 12	CW1305 00	CW1333 46	CW1330 03	CW1308 26	CW1311 26
L- AucA	4	4	16	4	4	4	4	4	2	2
D- AucA	4	4	2	4	4	4	4	4	2	4
L- LnqQ	16	32	64	32	8	16	4	2	2	2
D- LnqQ	4	8	4	4	4	4	2	2	2	2

Melitt in	8	8	4	4	4	8	16	16	8	8
--------------	---	---	---	---	---	---	----	----	---	---

2.3.2 Bacteriocin racemic protein crystallography

Since both L- and D-bacteriocins can be prepared on multi-milligram scale, we embarked on racemic protein crystallography and solved their X-ray crystal structures at atomic-resolutions (0.89-1.12 Å). Racemic mixtures of AucA gave crystals across greater than 50% of all crystallization conditions in less than 24 hours, whilst only one condition gave racemic LnqQ crystals suitable for X-ray diffraction. The racemic structure of LnqQ was solved at 0.96 Å resolution, confirming the globular arrangement and solvent exposed tryptophan residues reported through previous NMR investigation.²⁰⁰

The X-ray structures of a racemic AucA grown across a range of conditions were solved for an expanded investigation. First, various surface-exposed tryptophan residues including W3, W31 and W40 were found to form distinct salt-bridge networks. The NH in the indole motif and the side chain ammonium group of neighbouring lysines form hydrogen bonds (< 3.2 Å) with the sulphate oxygens (Figure 2.2 A-C). When a crystal was grown in the absence of sulfate, with the non-tetrahedral-oxyanion salts citrate and acetate, the resulting structure was devoid of the salt coordination networks. However, when the sulfate-free crystal was soaked with the isosteric glycerol 3-phosphate which is prominently seen in the phospholipid head group of Gram-positive bacterial membranes, the same Trp-Lys sidechain interaction was revealed for W31 (Figure 2.2 D). Indeed, W40 and W31 are highly conserved among leaderless bacteriocins, including LnqQ, suggesting that significant selection pressure resulted from these tryptophan residues being responsible for membrane interaction through forming a salt coordination network (Table S7.1).

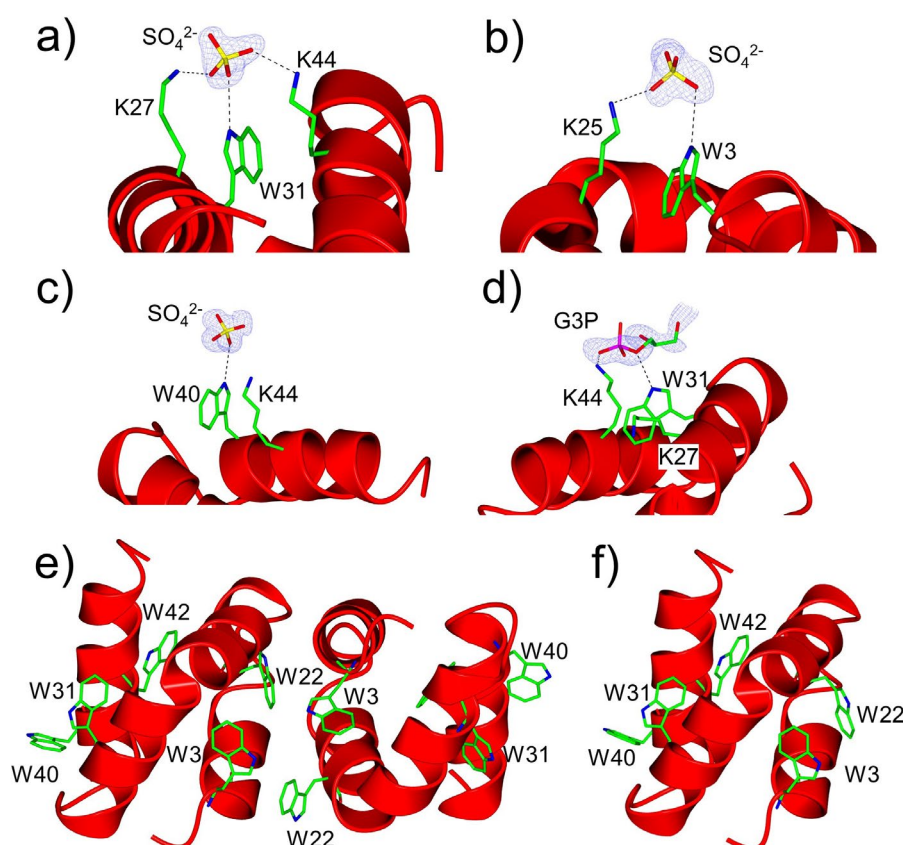


Figure 2.2: Insights from racemic protein crystallography of AucA; **a-c)** coordination of solvent exposed W31, W3 and W40 with sulfate (PDB: 8AVR) and **d)** coordination of W31 with glycerol 3-phosphate (G3P) (PDB: 8AVT); and comparison of **e)** dimeric and **f)** monomeric AucA. (PDB: 8AVU and 8AVS).

2.3.3 Effect of tryptophan substitutions on antibacterial activity

To further investigate the surface-exposed tryptophan residues in AucA, the corresponding site-specific variants were synthesized for analysis. Each of the salt-coordinating tryptophan residues (W3, W31 and W40) were initially replaced by a glutamate residue. When paired with neighbouring lysine residues, these Glu-Lys pairs are expected to result in locally decreased surface entropy disfavoured intermolecular interaction,²⁰⁶ for example, with an external tetrahedral oxyanion. Characterisations by LC-MS confirmed the synthesis of the variants, and CD spectroscopic analyses indicated that the alpha-helical folds are retained (Figure S7.2). Subsequently, the variants were assayed for their MICs with the hospital-isolated pathogens. The W40 and W31 appear to be critical as their individual mutation to glutamate abolished

antimicrobial activity. In contrast, a W3E mutation affects the activity of AucA in a strain-dependent manner.

Additional investigation was conducted by replacing each Trp in L-AucA with a leucine residue. The results show that the indole motif for W40 is essential for activity, as this substitution diminished the antibacterial activity for AucA-W40L. In contrast, the W31L variant remains essentially as active as the wild type, indicating that a hydrophobic isobutyl group can fulfil a similar role to that of this tryptophan residue. Interestingly, W3L and W22L variants have similar patterns in their MIC assays; however, unlike W3, W22 does not form a salt-bridge network component, despite being located at the protein surface. Finally, the W42L substitution had little effect on activity which is commensurate with it being buried in the peptide core. Collectively, these results suggest that W40 and W31 are key residues that facilitate AucA binding to the bacterial membrane, with the former primarily involved in coordination with lipid phosphate and the latter in hydrophobic interaction.

To confirm the results obtained with AucA, each tryptophan in L-LnqQ was also replaced with a leucine residue, without disruption of their alpha-helical folds (Figure S7.3). Similarly, the analogous W41L substitution in L-LnqQ diminished the antibacterial activity, confirming the necessity for an indole motif at this position. Meanwhile, the W32L and W23L variants affected activity strain-dependently, with lesser effect of the W21L substitution.

Table 2.2: Minimum inhibitory concentrations^[a] (MIC) of bacteriocin variants and control antibiotic agents against bacterial strains.

Antibiotic agent		<i>S. aureus</i>				<i>S. epidermidis</i>	<i>E. faecalis</i>	<i>E. faecium</i>
		AC2921 3	CW11412 5	CW11585 2	CW11337 9	CW131612	CW13300 3	CW13112 6
AucA variants	L-AucA	4	4	16	4	4	4	2
	D-AucA	4	4	2	4	4	4	4
	L-W3L	4	4	32	4	4	16	4
	L-W3E	16	32	64	16	16	64	4
	L-W22L	2	8	16	16	4	16	4
	L-W31L	4	4	8	4	4	8	2
	L-W31E	>64	>64	>64	>64	>64	32	8

	L-W40L	>64	>64	>64	>64	>64	32	8
	L-W40E	>64	>64	>64	>64	>64	>64	16
	L-W42L	2	4	8	8	4	16	2
LnqQ variants	L-LnqQ	16	32	64	32	8	2	2
	D-LnqQ	4	8	4	4	4	2	2
	L-W21L	16	32	32	16	16	8	4
	L-W23L	32	>64	64	64	32	4	4
	L-W32L	64	64	32	32	16	4	4
	L-W41L	>64	>64	>64	>64	>64	64	8
	Tetracycline	<0.125	32	8	16	1	1	>64
	Ampicillin	2	16	4	32	>64	2	>64
	Melittin	8	8	4	4	4	16	8

[a] MIC values expressed in $\mu\text{g/mL}$, [b] AC: American Type Culture Collection strain, [c] CW: CANWARD collection nosocomial strain.

	AC292 13	CW1141 25	CW1158 52	CW1133 79	CW1316 12	CW1305 00	CW1333 46	CW1330 03	CW1308 26	CW1311 26
L-AucA	4	4	16	4	4	4	4	4	2	2
D-AucA	4	4	2	4	4	4	4	4	2	4
L-LnqQ	16	32	64	32	8	16	4	2	2	2
D-LnqQ	4	8	4	4	4	4	2	2	2	2
Melittin	8	8	4	4	4	8	16	16	8	8

2.3.4 Investigation of bacteriocin oligomerization and its effect on stability

To investigate further why the W3 and W22 modifications affect AucA activity in a strain-dependent manner, racemic protein crystals generated in the absence of tetrahedral anion were investigated. The structure, solved at 0.89 Å, appeared to be homodimeric with two L-AucA molecules (or two D-AucA molecules) arranged in a head-to-tail fashion. Notably, the indole ring of W3 in one subunit is in proximity to the indole ring of W22 in the opposite subunit, forming a T-shaped arrangement at the dimer interface (ca. 5 Å, see Figure 2.2 E). Interestingly, W22 adopts a different orientation between monomeric and dimeric forms, switching from an intramolecular W3-W22 interaction to an intermolecular pair at the interface (Figure S7.4). Examination of the structure in protein interfaces, surfaces and assemblies (PISA)²⁰⁷ software showed that the dimeric interface is driven by the shielding of a hydrophobic patch on the surface of AucA from the solvent (Figure S7.5). Further analysis by size exclusion chromatography (SEC) was consequently conducted under physiological conditions (phosphate-buffered saline, pH 7.4). The W3L variant appeared to be monomeric having a retention time clearly longer than that of the wild-type AucA, but similar to that of LngQ with comparable monomeric molecular weight (5983 vs 5897 Da) (Figure 2.3). In contrast, AucA-W22L shared a similar retention time with those of the wild-type protein and other Trp→Leu variants (Figure S7.6), suggesting the dimeric interface was not completely disrupted upon replacement. However, the W22L substitution renders AucA considerably more susceptible to proteolysis, being completely degraded following 24 h incubation with promiscuous proteases (proteinase K or papain) at 37 °C and pH 7.4 (Figure S7.7). Accordingly, oligomerization that involves W3-W22 interaction may help keeping AucA intact preventing recognition by bacterial proteases, and hence activity of the corresponding variants appears to be strain-dependent in the MIC assay (Table 2.2).

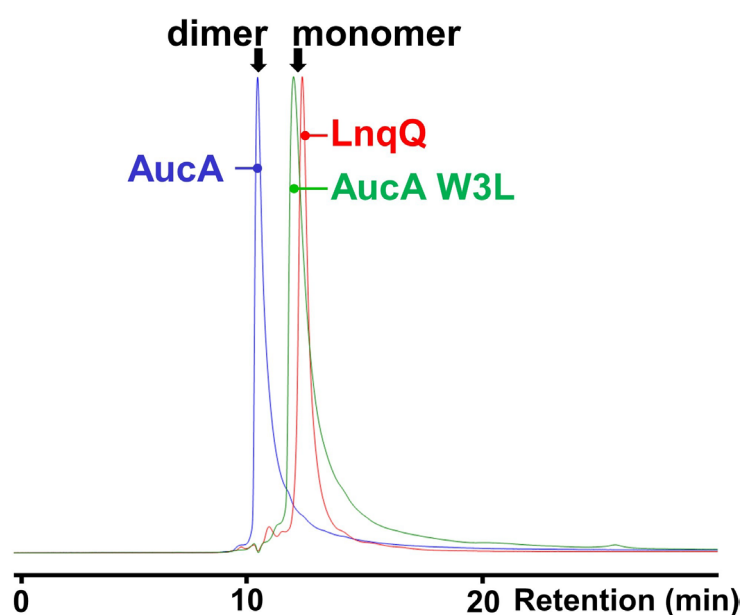


Figure 2.3: Analytical HP-SEC showing upfield retention of AucA W3L with respect to wild-type, dimeric AucA, and similar retention to monomeric LncQ. See Figure S6 for extended analysis.

2.4. Conclusion

Here, new insights into the physiological role of tryptophan residues located on the surface of leaderless bacteriocins are presented. Tryptophan contains an indole motif which has been known to play critical roles in both the stability and function of antimicrobial peptides.²⁰⁸ W31 and W40 in AucA are important for activity, forming a hydrogen-bonding network with membrane lipid phosphate and the neighbouring lysine residues. Similarly, the analogous W41 in LncQ also proved important for antibacterial activity, reinforcing the insights gained through racemic protein crystallography. These findings align with previous *in silico* analyses that have shown how the indole motifs of antibacterial polypeptides often locate at a membrane-water interface.²⁰⁹⁻²¹² In fact, W31 and W40 of AucA are highly conserved amongst AucA-like leaderless bacteriocins including lacticin Q,¹⁹⁸ lacticin Z,²¹³ epidermicin NI01²¹⁴, whereas positions 25 and 44 typically contain hydrogen-bond donor residues such as lysine or glutamine (Table S7.1) reinforcing their importance. On the other hand, solvent-exposed W3 and W22, which were previously proposed to be critical for activity,²⁰⁰ participate in forming an oligomeric

interface and increasing proteolytic stability. It has also been proposed that Trp is involved in a similar role in puroindoline A²¹⁵ and Tritrpticin,²¹⁶ stabilizing the folds of these antibacterial mini proteins. This work has illustrates how nature has evolved a scaffold that confers both activity and stability, and such knowledge will likely find uses in the design of novel antimicrobial agents.

Chapter 3 : Isolated domain of TNFR-1 for the discovery of D-peptide inhibitors via mirror image discovery technologies

3.1. Preface

In an aim to develop an efficient D-peptide discovery platform for the inhibition of tumour necrosis factor receptor 1 (TNFR-1) activation, a synthetic mid-chain cytokine binding domain (TNRC2) is reported. The D-TNRC2 provided a facile enantiomeric target for mirror image phage display, circumventing the need to synthesize D-TNFR-1 which contains 12 disulfide bonds. Phage display yielded a cyclic L-peptide candidate with affinity for D-TNRC2 ($K_D = 1.6 \mu\text{M}$), and its D-enantiomer binds to L-TNRC2 ($K_D = 1.0 \mu\text{M}$). The D-peptide candidate will be subject to biophysical analysis of TNFR-1 affinity and inhibition of cytokine activation. In addition, mirror image binder discovery was conducted *in silico*, using Rosetta to design mini protein binders for the D-TNRC2. However, flaws in the computational design disrupted the folding of the binder candidates, thus future optimization of the design protocol is required. Together, this chapter aims to circumvent the need for synthesis of the whole enantiomeric protein target in D-peptide inhibitor discovery.

3.2. Introduction

The cytokine tumour necrosis factor (TNF- α) activation of its receptor I (TNFR-1) is a key player in inflammation, with over activation - known as cytokine storm, associated with a variety of disease.²¹⁷⁻²¹⁸ Importantly, TNFR-1 overexpression correlated with mortality in COVID-19 patients, thus selective inhibitors are of profound interest.^{219, 220} Targeting TNFR-1 is advantageous, because activation of its sister receptor TNFR-2 by TNF- α is necessary for normal immunoregulatory function, limiting the use of anti-TNF- α therapeutics.²¹⁸ However, there are currently no reported small molecules capable of selective TNFR-1 inhibition.²¹⁸ Protein-based inhibitors could be used,²²¹ but they are amenable to proteolysis, triggering unwanted immunogenicity and limiting delivery routes. The development of D-peptide inhibitors of TNFR-1 pose an attractive solution, maintaining the necessary selectivity of polypeptides while being resistant to proteolysis.

Mirror image phage display (MIPD) remains the most routine approach for D-peptide binder discovery,⁸⁰ but requires the D-enantiomer of the protein target as a bait molecule for peptide screening. The extracellular portion of TNFR-1, involved in TNF- α binding, is 144 residues in length, consisting of four cysteine rich domains (CRD's) with 12 disulfide bonds.²²² Thus, preparation of its enantiomer would be a laborious challenge. Meanwhile, >80% of all residues involved in binding of TNF- α are contained within the mid-chain CRD2 (TNRCD2),^{222, 223} serving as a good candidate target for the screening of D-peptide inhibitors (Figure 3.2).

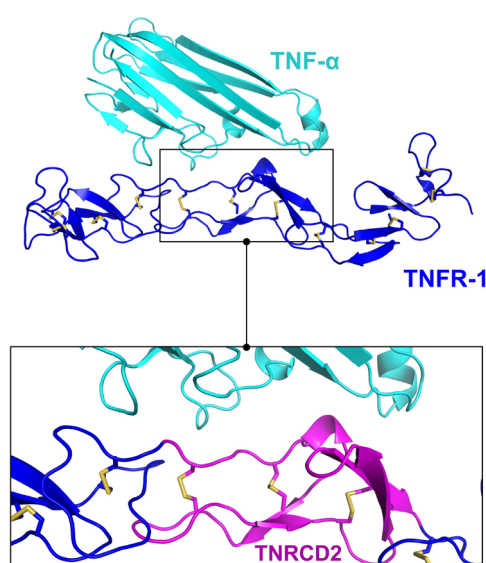


Figure 3.1: Interaction of TNFR-1 with TNF- α is significantly governed by the cysteine rich domain 2 (TNRCD2, magenta), serving as a good target for mirror-image phage display (MIPD).

TNRCD2 contains only 45 residues, presenting a significantly more accessible synthetic target, while the three disulfide bonds provide additional rigidity to the isolated domain.²²² The use of D-TNRCD2 as the target in MIPD not only simplifies its preparation, but also reduces the available protein area for off-target binding,²²⁴ increasing the probability of obtaining an inhibitor of TNFR-1/TNF- α interaction (Figure 3.2). Here, a synthetic route to TNRCD2 is reported, with folding into the correct conformation confirmed by racemic protein crystallography at 1.4 Å resolution. Preparation of D-TNRCD2 enabled MIPD for the discovery of a cyclic peptide

candidate with low micromolar affinity. Additionally, *de novo* design was used to create three potential D-TNRCD2 mini protein binders *in silico*, although further optimization is required.

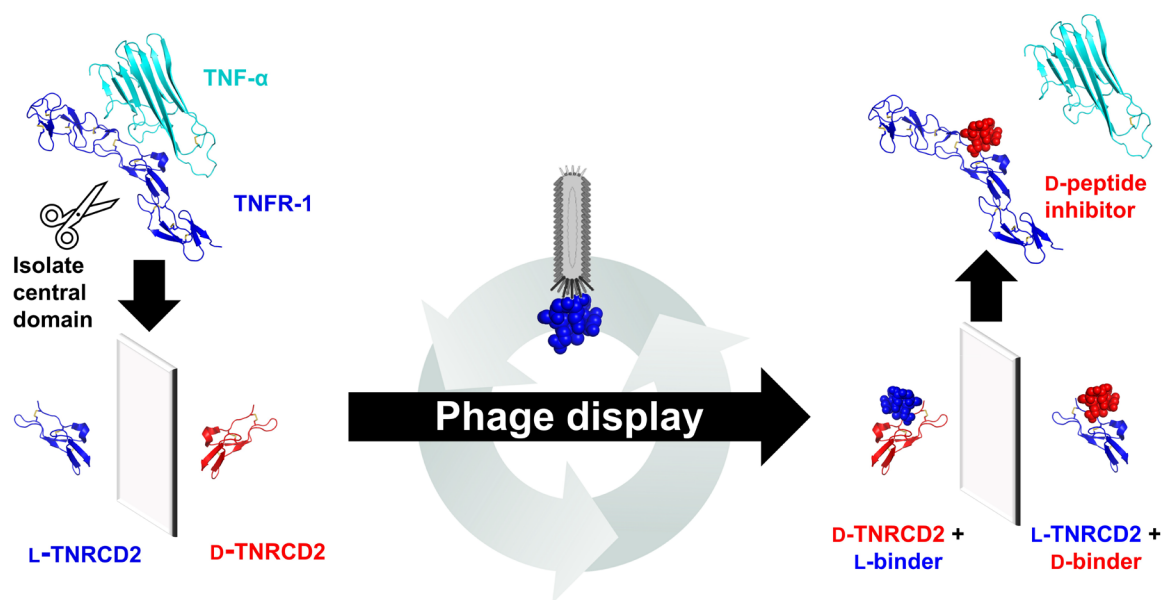


Figure 3.2: The use of TNFR-1 cysteine rich domain 2 (TNRCD2) as a target in mirror-image phage display (MIPD) to identify D-peptide inhibitors.

3.3. Results and discussion

3.3.1. Chemical synthesis of TNFR1 cysteine rich domain 2 (TNRCD2)

Synthesis of the entire TNRCD2 polypeptide chain (Figure 3.3 A) using automated SPPS required a large excess of amino acid reagents (10-20 equivalents), which is highly inefficient when using D-amino acids. Additionally, conjugation of a C-terminal linker (Gly-D-Ser-Gly-D-Ser-Gly-Lys(Biotin)-CONH₂) for streptavidin immobilization in phage display yielded multiple truncated products, regardless of amino acid excess. A likely explanation is that the additional C-terminal flexibility on the resin facilitated

peptide aggregation, decreasing *N*-terminus accessibility resulting in poor coupling efficiency. Therefore, synthesis via native chemical ligation (NCL) was developed.¹¹² The *N*-terminal peptide contains the first 16 residues as a peptide hydrazide,¹¹⁸ which is oxidized and converted to a trifluoroethanethiol (TFET) thioester *in situ*.¹⁸⁴ The *N*-terminal cysteine peptide contained the remaining 29 residues, either alone or with the C-terminal flexible biotin linker. NCL yielded the fully reduced TNRC2 peptide in good yield (72%), which was initially isolated by preparative HPLC for refolding optimisation. (Figure 3.3 A-B)

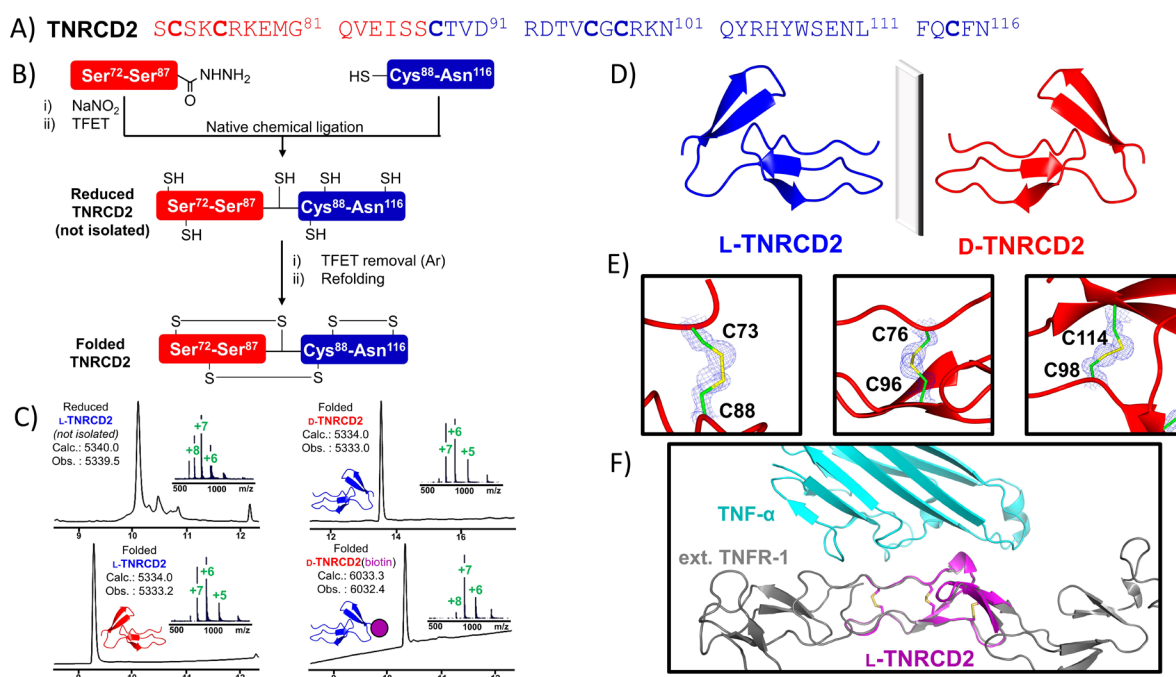


Figure 3.3: A) amino acid sequence of TNRC2, numbered based on full extracellular domain, B) synthetic scheme to prepare TNRC2, C) LCMS analysis of TNRC2 intermediate and enantiomer products, D) mirror image folding of L- and D- TNCRD2, E) σ_A -weighted $2F_o - F_c$ omit electron density maps contoured at 1σ (0.46 electrons per \AA^3) showing correct disulfide bond formation, F) structural overlay of the L-TNRC2 structure with TNF- α /TNFR-1 complex (PDB: 1TNR).

3.3.2. Optimization of TNRC2 *in vitro* protein folding

The six cysteine residues in TNRC2 must be oxidized into the correct disulfide bond pairings. Here, folding of the polypeptide from chaotropic agents is utilized to guide disulfide bond formation under thermodynamic control.⁴¹ The following optimisation

protocol was implemented for folding of the peptide: 1) oxidation method: DMSO, air (O_2) or Glutathione (GSH)/Glutathione disulfide (GSSG), then 2) buffer pH: tris (pH 8.5) or phosphate (pH 6.5) (TNRCD2 pI = 7.5), and finally 3) reagent concentrations: peptide and GSH/GSSG (Table S7.2). Refolding was monitored by LCMS, with up field retention of 1 min assumed to be correctly folded peptide, due to the burying of hydrophobic residues (Figure 3.3 C - left). The highest refolding yield (33% by LCMS) was obtained by dissolving TNRCD2 (0.5 mg/mL) in denaturation buffer (6 M Gn·HCl, 0.1 M NaPi, 6 mM GSSG, 60 mM GSH, pH 6.5), followed by a rapid five-fold dilution with phosphate buffer (0.1 M, pH 6.5). Refolding proceeded for four days at room temperature under inert atmosphere. Once refolding was established, the one-pot refolding protocol was developed (Figure 3.3 B). TFET was chosen for its volatility (bp = 37 °C), thus can be removed by bubbling Ar for 2 hours at 37 °C.¹⁸⁴ The NCL mixture, already in Gn·HCl/NaPi buffer, was adjusted to the correct concentration and pH followed by addition of GSH/GSSG for 10 mins, then rapid five-fold dilution. The reaction proceeded for four days at room temperature under inert atmosphere, and the folded peptide was isolated by preparative HPLC with 17% total yield (Figure 3.3 C). The protocol was subsequently used to prepare L- and D-TNRCD2, each with and without the C-terminal biotin linker.

3.3.3. Validation of TNRCD2 conformation by racemic protein crystallography

To confirm the correct folding of TNRCD-2, the structure was solved using racemic protein crystallography.³¹ Briefly, a mixture of DL-TNRCD2 was subject to sparse matrix crystallization screening using sitting drop vapour diffusion. The best crystal formed in 1.5 M Sodium chloride and 10% v/v ethanol, and its X-ray structure was solved at 1.4 Å resolution. Mirror-image folding of D- and L-TNRCD2 was observed in the crystal (Figure 3.3 D), an essential observation for MIPD. Well-defined electron density confirmed the correctly formed disulfide bonds (Figure 3.3 E), reinforcing assumptions made in the refolding assay. Finally, structural overlay of isolated L-TNRCD-2 with the TNF- α /TNFR-1 structure showed good agreement (Figure 3.3 F).²²² Therefore, TNRCD2 is an appropriate model for the design of D-peptide ligands, and ultimately, antagonistic ligands that can compete with TNF- α binding in cytokine storm.

3.3.4. Mirror-image phage display to identify D-peptide ligands for TNRC2

The biotinylated D-TNRC2 was delivered to the group of our collaborator, Prof. Chuanliu Wu at Xiamen University, and phage display experiments discussed in this section were conducted by PhD student, Yifu Kong. The peptide library presented on the surface of the bacteriophages consisted of the sequence format CX₉C, with the aim of obtaining 11-residue, terminal-disulfide cyclic peptides, where X could be any canonical amino acid. Following three rounds of biopanning, the peptide sequence CFHCVWLGMCEC was enriched (Table S7.3), and next generation sequencing showed that the WLG motif was relatively conserved (Table S7.4). However, the presence of a cysteine residue at position four perplexed the view of a terminal disulfide cyclic peptide. A second experiment of biopanning was conducted, this time keeping the WLG motif fixed in the format CX₄WLGX₂C. Following three rounds of selection, the library was enriched >8000 fold (Table S7.5), with the second library bearing multiple conserved residues (Figure 3.4 and Table S7.6). Similarly, Cys4 was highly conserved. The sequences and enrichment data (Tables S7.3 - S7.6) were provided back to the group at Cardiff University, and analysis resumed in this work.



Figure 3.4: Sequence consensus of the enriched peptide library following the second biopanning experiment, using a CX₄WLGX₂C library format.

3.3.5. Identification of the cyclic peptide active isomer

The most abundant peptide sequence from biopanning, CFHCIWLGDEC (TCPB-E), has three possible disulfide bond isomers, between cysteines 1-4, 1-11 and 4-11. Given the oxidizing environment of the *E. coli* periplasmic space,²²⁵ where the phage libraries are assembled,²²⁶ we envisaged that the cysteine not involved in intramolecular disulfide pairings would form an intermolecular disulfide bond with cysteine or glutathione. Thus, giving six possible peptide conformers that may be responsible for binding to D-TNRCD2 (Figure 3.5).

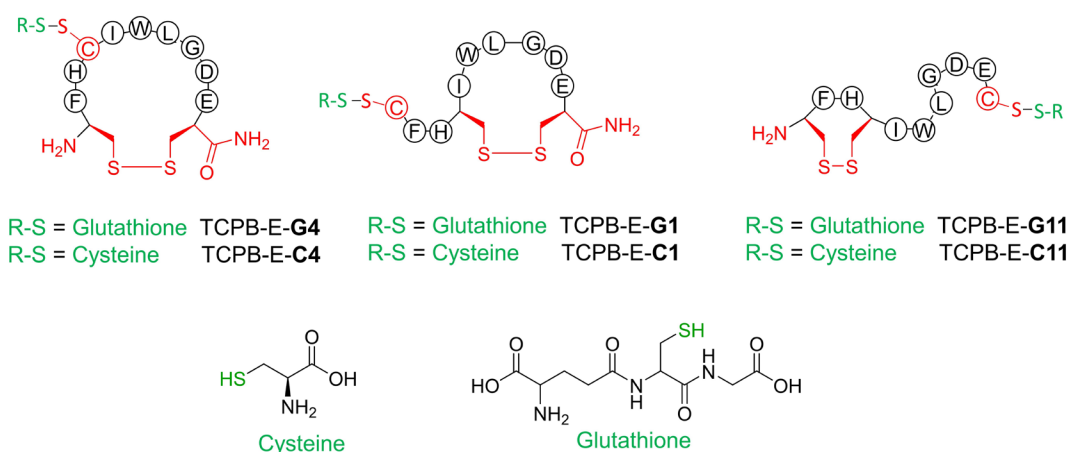


Figure 3.5: Six possible conformers of TCPB-E, oxidized with either glutathione (TCPB-E-**G**) or cysteine (TCPB-E-**C**).

To simulate these conditions *in vitro*, the reduced TCPB-E peptide prepared by SPPS, was dissolved at 50 μ M concentration in buffer (0.1 M NaPi, pH 6) containing either cystine (0.5 mM) or glutathione disulfide (0.5 mM). Two major peaks with each oxidant corresponded to the target masses of TCPB-E-**C** or TCPB-E-**G** isomers, suggesting an equilibrium of two of three of the possible disulfide conformers with each cysteine and glutathione (Figure S7.8). Each of the peaks were isolated by HPLC, and the conformers were identified by chymotrypsin digestion/LCMS disulfide bond mapping. In the case of each oxidant, the two major products consisted of intramolecular disulfide bond pairings between cysteine 1-4 (TCPB-E-**C11** or **G11**) or 4-11 (TCPB-E-**C1** or **G1**) (Figure 3.6 for **TCPB-E-C** peptide, S7.11–S7.12 for TCPB-E-**G** peptides).

Interestingly, no significant formation of the 1-11 disulfide product was observed, hence explained the bacteriophage enrichment of cysteine 4.

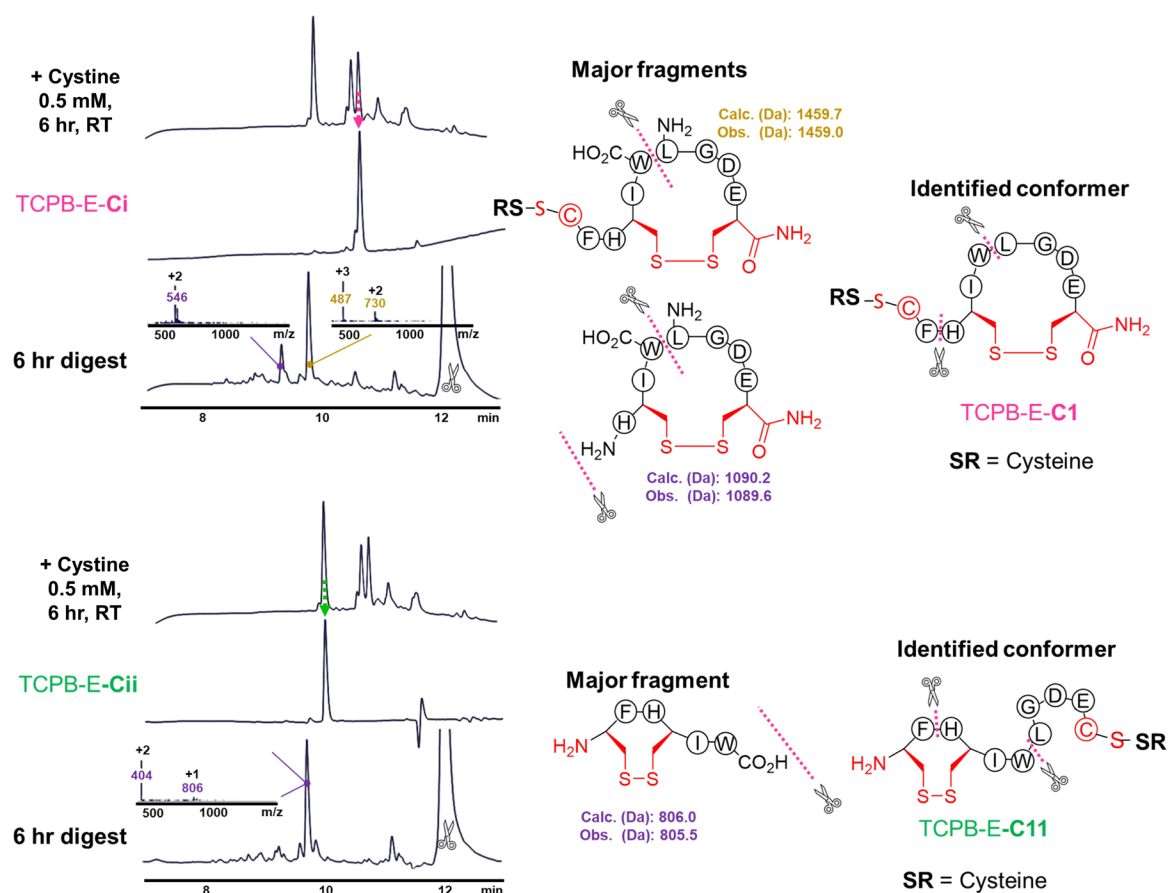


Figure 3.6: Solution-phase oxidation of reduced TCPB-E peptide (Cystine 0.5 mM) and isolation of each tautomer TCPB-E-Ci (top) and -Cii (bottom). Chymotrypsin digestion of the isolated oxidized peptide followed by LCMS analysis allowed assignment of the present conformer.

Next, the cyclic L-peptides were tested for binding to the phage display target by grating coupled interferometry (GCI), with biotinylated D-TNRCD2 immobilized on streptavidin coated sensor chips. Repeated analyte pulses of increasing duration (RAPID) assays were conducted to identify the peptide conformer with most significant binding, although quantities obtained were insufficient for full kinetic analysis at this point. TCPB-E-C1 demonstrated the highest sensor response with respect to concentration (Figure S7.13); hence, was proposed to be the most active conformer.

To enable further binding analysis, a synthetic route to the TCPB-E-**C1** peptide was developed for facile scale up synthesis. Orthogonal on-resin protection of the cysteine residues was used to control disulfide bond formation, based on previous work²²⁷ adapted here for on-resin synthesis (Figure S7.14 and Method 5.3.12). The new scheme was also used to prepare the TCPB-E-**C4** conformer, which was not isolated following the solution-phase oxidation. Indeed, comparison of HPLC retention times of the orthogonally prepared peptides with the crude solution-phase oxidised product supported the LCMS disulfide mapping experiments (Figure S7.15).

To further investigate the binding of each conformer, TCPB-E-**C1** and TCPB-E-**C4** (50 μM) were flowed over the GCI sensor chip containing D-TNRCD2, which reinforced that TCPB-E-**C1** is the desired binding conformation presented by the bacteriophage during biopanning (Figure S7.16). Subsequently, full analysis of the binding kinetics of TCPB-E-**C1** could be conducted using serial two-fold dilutions of concentration and demonstrated a binding affinity to D-TNRCD2 of 1.6 μM (Figure 3.7). The D-enantiomer of the candidate binder, D-TCPB-E-**C1** also demonstrates reciprocal binding to the L-TNRCD2 (K_D = 1.0 μM) (Figure 3.6).

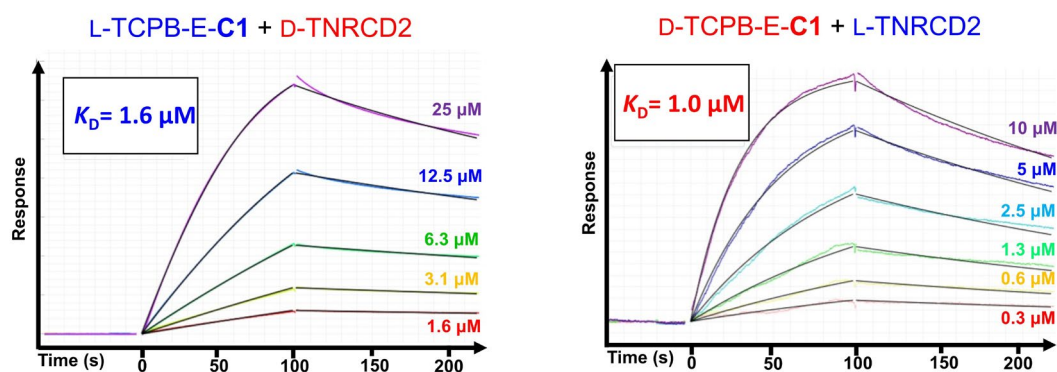


Figure 3.7: Multi-cycle kinetic analysis of L-TCPB-E-**C1** binding to D-TNRCD2 (left), and D-TCPB-E-**C1** binding to L-TNRCD2 (right) using grating-coupled interferometry (GCI). Equilibrium binding constant K_D is indicated in insets, using data fitted with a 1:1 Langmuir model (black lines).

D-TCPB-E-**C1** serves as a promising starting point for the optimization of a D-peptide binder for TNFR-1, and possibly, an inhibitor of TNF- α /TNFR-1 interaction. However,

further biophysical analysis is required to characterize the binding of the D-peptide native TNFR-1 receptor, followed by *in vitro* and *in vivo* inhibition assays (Chapter 4).

3.3.6. Mirror-image *de novo* protein design – a case study using TNRC2

The fundamental aim of this chapter was to develop D-peptide inhibitors of TNFR-1 by only preparing a smaller, enantiomeric subdomain for MIPD. However, if the mirror-image screening principle could be conducted *in silico*, the need for target chemical synthesis could be superseded. The method investigated here exploits *de novo* design of protein binders for the D-target structure (Figure 3.8), serving a similar role to that of MIPD.

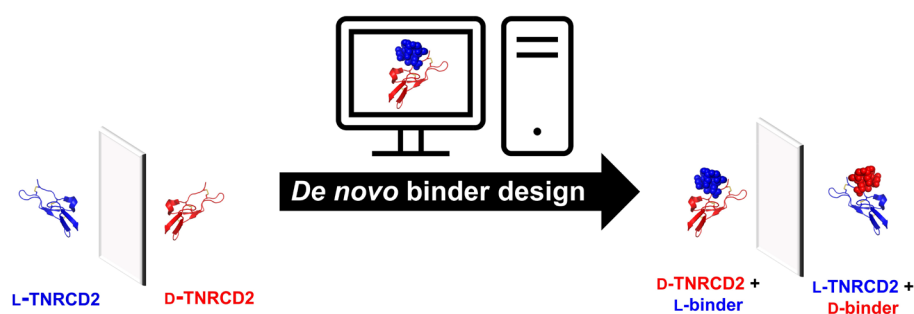


Figure 3.8: Proposed method of mirror-image *de novo* protein binder design

To investigate the feasibility of this approach, a D-protein target that can be accessed via synthetic routes was desired, so that both approaches can be easily cross-validated. With the D-TNRC2 at hand, and its structure solved by racemic protein crystallography, this protein was selected as the model target for this work. Various methods have been developed for *in silico* binder design.^{228, 229} This work adopts Rosetta protein design,²³⁰ due to the stability of the designed mini protein binders,^{79, 231} their high specificity and potency,²³² the tolerance of the software for D-amino acids,⁷⁹ the single requirement of an input target structure,²³³ and its relative simplicity of use.²³⁴

Initially, a previously reported mini protein scaffold is selected for the design of the TNFR-1 mini protein binder (TMPB). Four candidate scaffolds were chosen with

varying topologies,²³¹ all containing less than 64 amino acids (Figure 3.9). The scaffolds were first docked onto the D-TNRCD2, filtering out the scaffold “poses” which provided the best contact at the desired TNF- α binding interface. Following preliminary design of the scaffold binding interfaces, the scaffold 5UOI consistently produced better candidates, thus was selected as the protein scaffold for binder design.

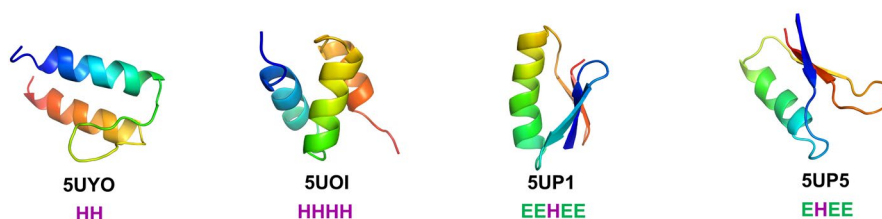


Figure 3.9: Mini protein scaffolds used for binder design in this work, denoting the PDB ID's and secondary structure compositions (H = α -helix, E = β -sheet).

The docking of the 5UOI protein to D-TNRCD2 was repeated, this time generating over 5500 poses of the docked scaffold. The top 20 poses, providing sufficient binding surface area and shape complementarity at the TNF- α interface were selected (Figure 3.10). The 20 docked poses were then each subjected to 100 rounds of interface design (2000 initial binders), and the best 100 candidates were sorted by their simulated binding energies. Following multiple sequence alignment of the best designs, and further refinement of the design aided by visual inspection in PyMOL, the candidate protein binders were identified.

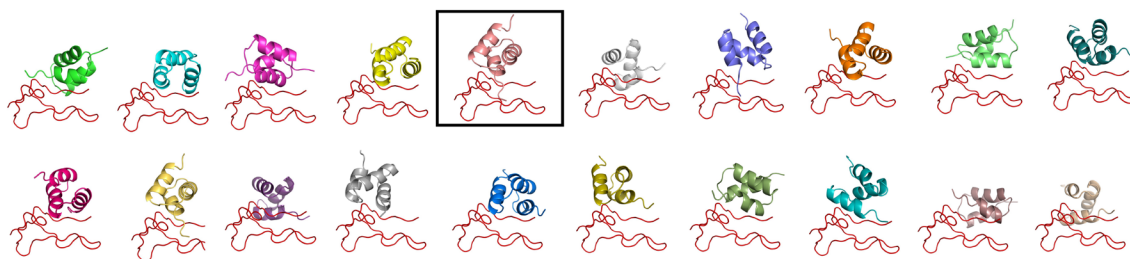


Figure 3.10: The top 20 poses of 5UOI docked to D-TNRCD2, used in the subsequent scaffold binding interface design. The pose giving rise to the best candidate binders (TMPB-TX) is highlighted.

The three top scoring designs arose from a single docked pose (Figure 3.10), where the C-terminal mini protein tail featuring VXXKK, is inserted into the D-TNRCD2 target, named here as TMPB-TX (Figure 3.11). Three candidate designs were initially validated by a full docking protocol of the TMPB proteins against the D-TNRCD2 using RosettaDock, followed by plotting of the predicted binding energies against the RMSD of the docked model to the designed model (Figure S7.17). The best designed binders gave the lowest binding energy at the lowest RMSD to the designed model, whereas a poorly designed model control (TMPB-TL) and the undesigned scaffold (TMPB-Sc) gave more dispersed results.

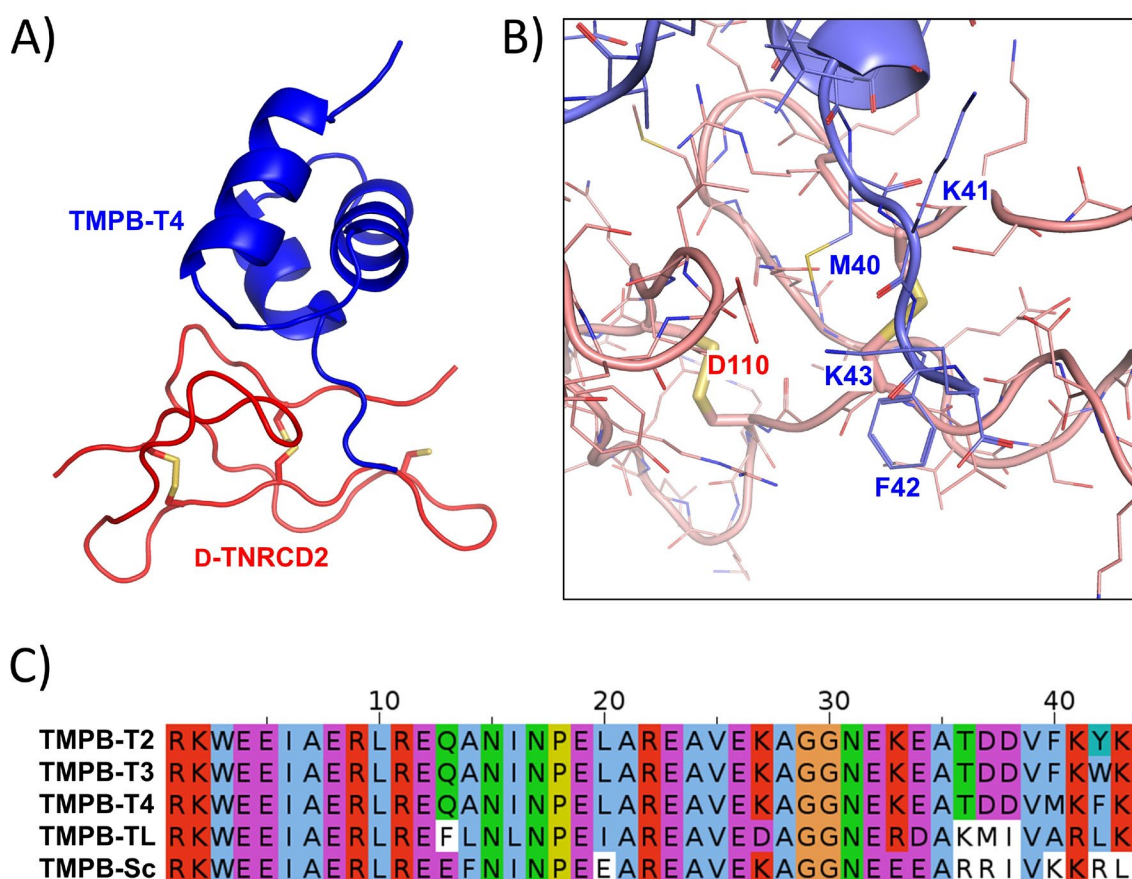


Figure 3.11: Candidate mini protein binders for the D-TNRCD2 target. A) Structure of designed TMPB-T4 bound to D-TNRCD2, B) interactions of the C-terminal tail of TMPB-T4 with D-TNRCD2, and C) sequence alignment of the three candidate binders (TMPB-T2, -T3, and -T4), a low scoring design TMPB-TL, and the undesigned scaffold protein (TMPB-Sc).

Next, the three designed proteins (TMPB-T2, -T3, and -T4) and the undesigned scaffold (TMPB-Sc) were prepared by SPPS, then tested for binding to the D-TNRCD2 target using GCI. Surprisingly, no binding to the target was observed. Circular dichroism spectra of the designed proteins were compared to the original protein scaffold, which showed that the alpha-helical folds of the protein are lost following the binder design (Figure S7.18). Subsequently, the mirror image *de novo* protein design platform must be subject to further optimization (see Chapter 4).

3.4. Conclusion

In this work, a mid-chain, ligand binding cysteine-rich domain of a key cytokine receptor has been prepared, with its D-enantiomer (D-TNRCD2) subject to mirror image binder discovery technologies. The D-TNRCD2 provided a facile enantiomeric target for mirror image phage display, circumventing the need to synthesize the challenging, D-TNFR-1, at 144 residues in length with 12 disulfide bonds. Phage display against the D-TNRCD2 target domain yielded a cyclic peptide candidate TCPB-E-C1 with low micromolar binding affinity, whose enantiomer will be tested for binding to the native TNFR-1 and subsequent cytokine inhibition. The use of central, cysteine rich domains (typically 30-50 residues) for discovering D-peptide inhibitors of protein-protein interactions holds promising applications. Indeed, many cell surface receptors contain extracellular cysteine rich domains,²³⁵ whose overexpression is linked to numerous diseases, including cancer,^{236 237} autoimmune diseases,^{217, 238} and atherosclerosis.²³⁹ Therefore, once D-TCPB-E-C1 inhibition of TNFR-1 cytokine activation is confirmed (see Chapter 4), this methodology could find profound impact in the discovery of D-peptide inhibitor candidates for numerous drug targets.

In addition, the mirror image binder discovery was conducted *in silico*, using the D-target structure alone. Three, 43-residue mini proteins were designed to bind at the cytokine receptor interface of D-TNRCD2. Whilst the functionality of the designs was confirmed *in silico*, no binding to the target was identified *in vitro*, attributed to disruption of the protein folding during design. Nevertheless, future optimization of the mirror-image *de novo* protein design platform has the potential to supersede synthesis of an enantiomeric protein target for D-peptide inhibitor discovery (see Chapter 4).

Through conducting design *in silico*, a greater number of D-polypeptide candidates can be discovered for a broader range of targets.

Collectively, this work aimed to address two major issues of mirror image phage display: D-target synthesis and off-target binder discovery. While the technologies presented in this chapter require further optimization, their application could greatly increase the feasibility of D-polypeptide inhibitor discovery, facilitating downstream clinical research and development.

Chapter 4 : Thesis summary and future directions

4.1. Summary

The applications of D-proteins are vast but remains to be challenged by their difficulty in preparation, particularly as protein complexity increases. While larger D-proteins have been reported through laborious synthetic protocols, the use of smaller, more accessible D-proteins can expedite their wider use in applied research.

Firstly, efficient synthetic routes to the D-enantiomers of two membrane-active bacteriocins, aureocin a53 (AucA) and lacticin Q (LnqQ) were reported. This enabled high-resolution racemic protein crystallography to unveil new insights into the roles of conserved surface tryptophan residues in bacteriocin activity and stability. While bacteriocins present as promising agents for tackling antibacterial resistance in the food and medical industries,^{194, 240} the lack mechanistic information impedes their application. The structural investigations presented here demonstrated how the tryptophan residues may be involved in membrane lipid coordination, which are known to locate near the water-lipid interface in membrane-active polypeptides.^{209, 212} Additionally, racemic protein crystallography unveiled a novel oligomeric state of AucA mediated by tryptophan residues, which was found to confer proteolytic stability. This work illustrated how nature has designed bacteriocin scaffolds that confer both activity and stability, and such knowledge can expedite their application as antibacterial agents.

Secondly, this work also aimed to develop a potent D-peptide candidate capable of targeting tumour necrosis factor receptor 1 (TNFR-1) activation, as a potential therapeutic candidate in the treatment of inflammation. A synthetic route to a central, cysteine rich domain (TNRC2) of TNFR-1 was reported, which contains the majority of the key binding residues of the TNFR-1 cytokine ligand. The use of its enantiomer,

D-TNRCD2, enabled mirror-image phage display (MIPD) to identify a cyclic peptide binder with low micromolar affinity, serving as a promising starting point for the development of anti-TNFR-1 D-peptide candidates. While further analysis is required to validate D-peptide inhibition of TNFR-1, the use of central, cysteine rich domains in MIPD holds promising applications. Indeed, many cell surface receptors contain extracellular cysteine rich domains,²³⁵ whose overexpression is linked to numerous diseases.^{217, 236-239} Therefore, this methodology could find broad use in the discovery of D-peptide inhibitors, whilst circumventing the challenge of preparing whole, D-protein targets.

Finally, a route to develop D-polypeptide ligands without any requirement of a synthetic D-target was explored. Conducting mirror-image binder design *in silico* could vastly widen the scope of possible targets, while controlling the site of binding and increasing the throughput of discovery. The method presented here utilized Rosetta to design three mini protein binders for the D-TNRCD2. However, the candidates did not show affinity for the target *in vitro*, attributed to a loss of protein folding during design. Nevertheless, future optimization of the mirror-image design platform could enable facile access to proteolysis-resistant D-polypeptide ligands, without the constraints imposed by D-target synthesis in MIPD.

Collectively, this thesis aimed to exploit the broad utility of D-protein technology in applied research, whilst mitigating the challenges arising from synthesis of large and complex D-proteins. New insights were generated in bacteriocin function enabled by an efficient synthesis of their enantiomers, and routes to facile D-peptide inhibitors were explored using a simplified domain for MIPD.

4.2. Further investigation into the mechanisms of bacteriocin activity

4.2.1. Probing the role of tryptophan-lipid interaction

In this work (Chapter 2), new insights were generated on the role of surface tryptophan in bacteriocins, where the indole NH of AucA was found to form hydrogen bonds with tetrahedral oxanions (sulfate and glycerol 3-phosphate) in conjunction with neighbouring lysine residues. Indeed, a co-ordinating Trp40 in AucA, and the analogous Trp41 in LnqQ were important for activity, with their substitution to leucine diminishing the MIC's. This work proposed that the indole NH hydrogen bonding observed in the crystal structures may be reminiscent of their interaction with the lipid phosphate in bacterial membranes. To further probe the role of the indole hydrogen bond, tryptophan derivatives can be incorporated into AucA and LnqQ using the efficient synthetic route reported here (Figure 4.1).

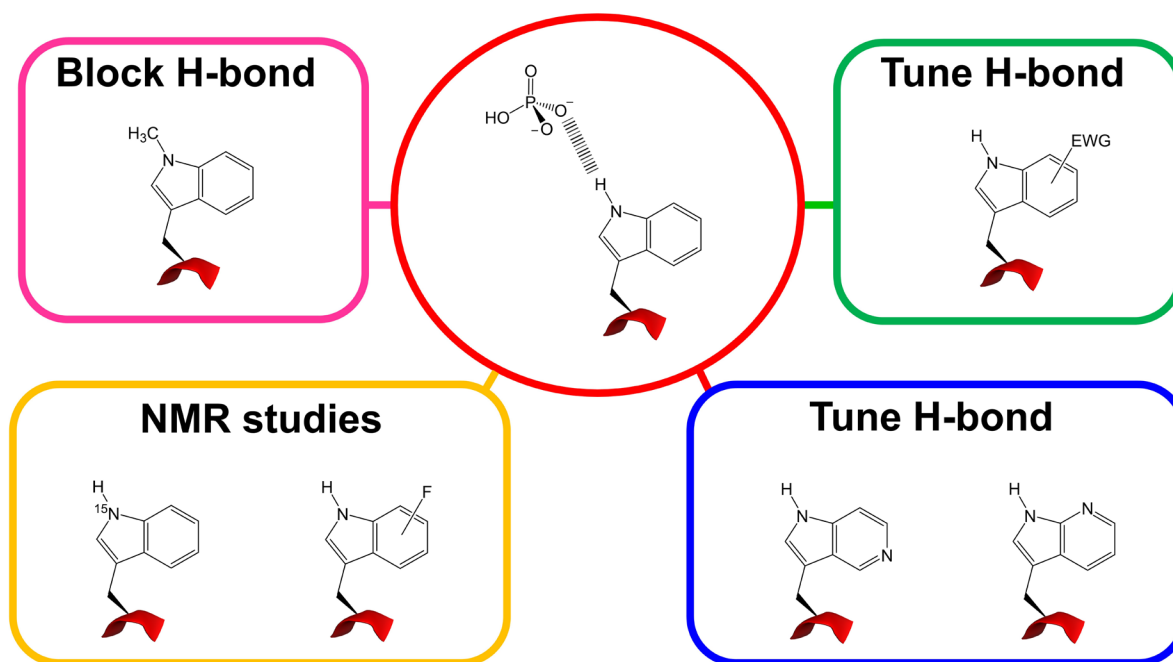


Figure 4.1: Tryptophan derivatives for incorporation into AucA and LpqQ, allowing further investigations into the role of the indole hydrogen bond (available from Aralez Bio²⁴¹).

Incorporation of 1-methyltryptophan derivatives at position 40/41 would be expected to decrease the MIC in a similar manner to the leucine substitution, due to loss of the indole hydrogen bond donor. Furthermore, the indole hydrogen bond donation can be fine-tuned. The introduction of electron withdrawing groups to the indole ring (halo-, cyano-, or nitro-indoles) should increase the acidity of the NH, resulting in stronger interaction with the lipid phosphate, and perhaps, better antibacterial activity. Similar indole NH acidification may also be obtained with the newly reported azo-tryptophans,²⁴¹ which would not impose additional steric bulk or electrostatics onto the indole ring.

Previously, UV spectroscopy of AucA has been used to probe the environment of tryptophan during lipid interaction.¹⁹⁹ However, this did not distinguish individual residues, which have been shown in this work to have vastly different roles in activity. Site specific incorporation of ¹⁵N on the indole, or fluorine-substituted indoles,²⁴² could be utilized for ¹⁵N or ¹⁹F NMR experiments (either in solution²⁴³ or solid state²⁴⁴) to provide information on the environment of each tryptophan in membrane mimetics.²⁴³

Collectively, these directions could provide a more comprehensive investigation into the role of tryptophan-lipid interaction.

4.2.2. Investigation of active bacteriocin conformations by racemic protein crystallography

The tryptophan-lysine-oxyanion coordination highlighted in this work provided useful insight into interactions that likely take place on the membrane surface. However, it is not known whether bacteriocins retain these conformations during the active state of membrane insertion.¹⁹⁶ Previously, the structures of transmembrane helices have been solved in racemic detergents by racemic protein crystallography.^{32, 33} Crystallization of DL-AucA and DL-LnqQ from racemic detergents, such as DL- β -octylglucoside or racemic monoolein, may provide novel structural information pertaining to the conformation of AucA and LnqQ within bacterial membranes.

4.3. Development of an expedient route to D-peptide binder discovery

4.2.1. Towards a D-peptide inhibitor of TNFR-1

Here, a lead candidate (D-TCPB-E-**C1**) has been identified that binds to L-TNRC2D with good affinity ($K_D = 1.0 \mu\text{M}$). However, steps towards development of a D-peptide capable of TNFR-1 inhibition remain to be conducted:

1. Analysis of D-TCPB-E-**C1** binding to TNFR-1 will be conducted, comparing the affinity to the TNRC2D as validation of the proposed method.
2. Investigations of D-TCPB-E-**C1** inhibition of TNFR-1/TNF- α should be conducted, first *in vitro*, using a commercially available FRET-based assay kit (AlphaLISA[®], Perkin Elmer), and then *in vivo*, using a HEK-2 blue cell assay²²¹

(TNF/TNFR-1 signalling activates NF- κ B,²⁴⁵ which subsequently regulates expression of the reporter protein, alkaline phosphatase).

3. Likely, affinity optimization will be required before D-TCPB-E-**C1** is capable of TNFR-1 inhibition at useful concentrations (low micromolar IC₅₀ or better).²²¹ Structure of the binder-target complex can be solved using racemic protein crystallography of DL-TCPB-E-**C1** with DL-TNRCD2,⁶³ or crystallography of D-TCPB-E-**C1** with either L-TNRCD2 or L-sTNFR1. Structure guided affinity optimization can then be conducted, exploring the use of non-natural amino acids and derivatization at position **C1**. In the absence of any structure, alanine-scan followed by structure activity relationships at key residues may prove useful.
4. Finally, D-TCPB-E-**C1** may be tested for selectivity of TNFR-1 over TNFR-2, and cysteine-rich domains of other targets may be explored using this methodology.

4.2.2. Optimization of mirror image *de novo* protein design

Three candidate mini proteins were designed to bind to D-TNRCD2 at the cytokine binding interface; however, the folding of the proteins was disrupted following the design. Further optimization of the Rosetta *de novo* design is needed:

1. New filters can be added into the Rosetta FlexDesign script (Appendix 7.5.6), to score proteins on their Lennard-Jones attractive force (fa_atr), thus searching for protein binder designs with more stable folds. Additionally, the layer design and favour sequence profile can be removed and/or tweaked, to allow scaffold redesign along with the binding interface to reinstate protein stability.
2. New scaffolds may be explored, particularly those with extra rigidity such as the disulfide bonded scaffolds.⁷⁹ It is envisaged that these will be more amenable to design, whilst retaining their three-dimensional folds.
3. The Rosetta design platform can be switched from scaffold docking and design, to a RifDock approach.²⁴⁶ This approach first screens for billions of

disembodied amino acids that can host interactions at the desired target interface (Rotamer interaction field – Rif). Then, a large library (typically >10000)²⁴⁷ of scaffolds is docked into the rotamer interaction field, selecting those that can host the desired amino acid residues. This approach has provided several high affinity *de novo* protein ligands,^{232, 233} although consumes significantly more computational resources.

Development of the *de novo* mini protein design platform holds vast potential in D-polypeptide binder discovery. In fact, virtually any protein with an available coordinate file may be used, by inverting the atom co-ordinates to generate a *pseudo*-D-protein target structure.

Chapter 5 : Materials and methods

5.1. Materials and instruments

Unless otherwise stated, chemicals and solvents were purchased from commercial suppliers (Sigma Aldrich, Fluorochem, Acros Organics, Alfa Aesar, Cambridge Bioreagents and Fisher Scientific) and used without further purification. HPLC grade (>99.8%) dimethylformamide was used for peptide synthesis. Rink amide ProTide resin was purchased from CEM and 2-Cl-trt Fmoc-hydrazine resin was prepared as previously described.¹²⁰ LCMS data was obtained using an Agilent Infinity 1260 II HPLC system fitted with an on-line Agilent 6120 quadrupole ESI-MS. Semi-preparative HPLC and analytical HP-size exclusion chromatography (SEC) was carried out using an Agilent 1260 infinity II HPLC system. Preparative HPLC was carried out using a Shimadzu Nexera preparative HPLC system. All HPLC systems used UV analyte detection at 210 nm and 280 nm. UPLC/high-definition mass spectrometry was obtained using a Waters SYNAPT G2-Si UPLC/HDMS system. Manual solid phase peptide synthesis (SPPS) was conducted in Telos Kinesis 15 mL solid-phase extraction (SPE) vessels fitted with PTFE frits and mounted onto a vacuum manifold. Automated SPPS was performed using a Liberty Blue microwave peptide synthesiser (CEM corp.). Peptide lyophilization was carried out by flash freezing the sample in liquid N₂ and drying on a Christ Alpha 2-4 LDplus freeze-dryer. Circular dichroism (CD) spectra were collected using an Applied Photophysics Chirascan CD spectrometer. Protein crystallization screening was conducted using a Douglas Instruments Oryx 4 crystallization robot. Minimum Inhibitory Concentration (MIC) assays were conducted in 96-Well Clear polypropylene corner notch microplates (Corning, #11313595). For the assay media, cation adjusted Mueller Hinton Broth 2 (Sigma Aldrich, #90922) was used for the *S. aureus*, and *S. epidermidis* strains and Brain Heart Infusion broth (BD, #221812) was used for the *E. faecalis* and *E. faecium* strains. As controls, ampicillin sodium salt (Biobasic, #AB0028) and tetracycline hydrochloride (Boehringer) were used, as well as melittin prepared by automated SPPS. Experimental bacterial strains used were procured from the American Type Culture collection (*S. aureus* 29213) and the Canadian Antimicrobial Resistance Alliance (CARA, CANWARD) strains with the following IDs: *S. aureus* (114125, 115852, 113379), *E. faecalis* (133003, 133346), *E. faecium* (130826, 131126) and *S. epidermidis* (130500, 131612).²⁰² Grating coupled interferometry experiments were conducted using a Creoptix WAVEsystem.

5.2. HPLC, UPLC and ESI mass spec procedures

5.3.1. Peptide LCMS analysis

Peptide samples were prepared at 0.1 mg/mL using 0.1% TFA in water and passed through a 0.22 μ M nylon filter. Unless otherwise stated, samples (10 μ L) were eluted with reversed mobile phase A (water + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid) at 0.3 mL/min over a RP-C18 column (ACE, 2.1 mm x 100 mm, 110 Å, 3 μ m) at 40 °C. A 5-70% gradient of A/B was applied over 30 minutes and analyte was detected using a UV detector at 210 nm and 280 nm, and positive electrospray ionisation mass spectrometry (ESI+ MS). ESI+ mass spectra are reported as the integrated spectra for the duration of the major peak in each UV210-nm chromatogram.

5.3.2. Peptide UPLC/HDMS analysis

Peptide samples were prepared at 0.1 mg/mL using 0.1% TFA in water and passed through a 0.22 μ M nylon filter. Samples (10 μ L) were eluted with reversed mobile phase A (water + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid) at 0.3 mL/min over a RP-C4 column (Waters Aquity BEH, 2.1 mm x 100 mm, 300 Å, 1.7 μ m) at 60 °C. A 5-95% gradient of A/B was applied over 50 minutes. Spectra were collected using high-definition electrospray ionisation in positive mode and analyzed using Waters MassLynx software version 4.1. ESI+ mass spectra are reported as the integrated spectra for the duration of the major peak in each total ion chromatogram.

5.3.3. Peptide semi-preparative HPLC

Samples were passed through a 0.22 μ M nylon filter. 1-4 mL of sample was eluted with reversed mobile phase A (water + 0.1% TFA) and B (acetonitrile + 0.1% TFA) at 4 mL/min over a RP-C18 column (ACE, 10 mm x 250 mm, 110 Å, 5 μ m) at room temperature. A 20-80% gradient of A/B was applied over 30 minutes and analyte was detected using a UV detector at 210 nm and 280 nm. Sample fractions were collected using an automated fraction collector, their identities were confirmed by LCMS and the fractions containing the target peptide were combined and lyophilized.

5.3.4. Peptide preparative HPLC

Samples were passed through a 0.22 μ m nylon filter. 3-10 mL of sample was eluted with reversed mobile phase A (water + 0.1% TFA) and B (acetonitrile + 0.1% TFA) at 18 mL/min over a RP-C18 column (Shimpack GIST, 20 mm x 150 mm, 100 Å, 5 μ m) at room temperature. A 20-60% gradient of A/B was applied over 40 minutes and analyte was detected using a photodiode array detector at 210 nm and 280 nm. Sample fractions were collected using an automated fraction collector, their identities were confirmed by LCMS and the fractions containing the target peptide were combined and lyophilized.

5.3. Synthetic procedures

5.3.1. Manual SPPS

Swelling: 0.1 mmol of resin was swollen in 50%/50% v/v DMF/DCM for 10 mins in an SPE column and drained.

Coupling: Fmoc-protecting group was removed by addition of 20% piperidine in DMF (3 mL) to the resin for 2 \times 5 mins. The resin was washed five times with DMF 3 mL. 2 equiv. of Fmoc-amino acid (200 mM), 1.95 equiv. of HBTU (195 mM), 2 equiv. of HOBT (200 mM) and 4 equiv. of DIPEA (400 mM) was dissolved in DMF and mixed for 0.5 min. The coupling mixture was transferred to the resin and allowed coupling to proceed for 30 mins at room temperature. Coupling was repeated for Val and Leu, whereas Ile single coupling was extended to 1 hr at room temperature. Final Fmoc-deprotection was carried out as above.

5.3.2. Automated SPPS with low amino acid excess

This SPPS procedure yields coupling cycles with 2-4 equivalents of amino acid and was used in preparation of fragments for native chemical ligation (LnqQ, AucA and TNRC2) as well as for TCPB peptides.

Automated SPPS was conducted at a 50 μmol scale using modified CarboMax coupling cycles.¹¹⁰ Fmoc-amino acid stock solutions, oxyma and DIC were used at 0.2 M concentration. Reactions were stirred by N_2 bubbling for 2 seconds on, 3 seconds off.

Fmoc deprotection: Piperidine in DMF (3mL, 20% v/v) was delivered to the reaction vessel. Microwave heating proceeded as follows: 0 W 20 ± 5 °C for 5 s, 100 W 78 ± 2 °C for 20s, 60 W 88 ± 2 °C for 10s, 20 W 90 ± 1 °C for 60s. The resin was then washed with DMF (4 x 2 mL).

Coupling cycle: Fmoc-amino acid (0.5 mL, 2 equiv.), Oxyma (0.5 mL, 2 equiv.) and DIC (1 mL, 4 equiv.) were delivered to the reaction vessel (Final concentrations: Amino acid 50 mM, Oxyma 50 mM, DIC 100 mM). Microwave heating proceeded as follows: 15 W 75 ± 2 °C for 15 s, 30 W 90 ± 1 °C for **X** s. The resin was then washed once with DMF (2 mL).

For single couplings, reactions proceeded for a total coupling time of 4 mins (**X** = **225** s). For double couplings (Trp, Leu and Val), the coupling was repeated. Due to high cost of diastereomeric D-Ile and its slower coupling rate (β -branched), a single 8 min coupling (**X** = **465** s) was implemented.

5.3.3. Automated SPPS with high amino acid excess

This SPPS procedure yields coupling cycles with 10-20 equivalents of amino acid and was used to prepare some full-length L-bacteriocins (L-AucA and L-AucA W \rightarrow E variants), L-TNRCD2 without C-terminal biotin linker, and L-TMPB proteins.

Automated SPPS was conducted at a 50 μmol scale using modified CarboMax coupling cycles.¹¹⁰ Fmoc-amino acid stock solutions were used at 0.2 M concentration; Oxyma and DIC were used at 1.0 M concentration. Reactions were stirred by N_2 bubbling for 2 seconds on, 3 seconds off.

Fmoc-deprotection was carried out as above.

Coupling cycle: Fmoc-amino acid (2.5 mL, 10 equiv.), Oxyma (0.5 mL, 10 equiv.) and DIC (1 mL, 20 equiv.) were delivered to the reaction vessel (Final concentrations:

Amino acid 125 mM, Oxyma 125 mM, DIC 250 mM). Microwave heating proceeded as follows: 15 W at 75 ± 2 °C for 15 s, 30 W at 90 ± 1 °C for 225 s. The resin was then washed once with DMF (2 mL).

Coupling was repeated for Ile, Val and all residues following 25 couplings.

5.3.4. Peptide hydrazide preparation

Peptides were assembled by either manual or automated SPPS onto a 2-chlorotrityl Fmoc-hydrazine resin¹²⁰ and subject to peptide cleavage. Because during automated SPPS the mildly acidic oxyma (pK_a 4.60) can cause premature release of the peptide from a 2-Cl-(Trt) resin at 90 °C, DIPEA (20 μ M) was added to the oxyma solution to minimize premature cleavage and increase yields of peptide hydrazide.¹¹⁰

5.3.5. C-terminal biotinylated peptide linker for TNRC2

Rink amide ProTide resin (0.1 mmol, CEM) was swollen in 50%/50% v/v DMF/DCM for 10 mins in an SPE column and drained. The fmoc-protecting group was removed by addition of 20% piperidine in DMF (3 mL) to the resin for 2 x 5 mins. The resin was washed five times with DMF (3 mL). 2 equiv. of Fmoc-Lys(Mtt)-OH (200 mM), 1.95 equiv. of HBTU (195 mM), 2 equiv. of HOBt (200 mM) and 4 equiv. of DIPEA (400 mM) was dissolved in DMF and mixed for 0.5 min. The coupling mixture was transferred to the resin and allowed coupling to proceed for 30 mins at room temperature. The remainder of the flexible linker containing Gly-D-Ser-Gly-D-Ser-Gly was assembled using manual SPPS procedure (Section 5.3.1), with the N-terminal Fmoc group left in place. The lysine side chain protecting group, 4-methyltrityl (Mtt) was removed using 1% TFA in DCM through 14 flow washes (3 mL each), monitored qualitatively by the intense yellow colour of the Mtt-OH group. The resin was washed three times with DCM (3 mL) and three times with DMF (3mL). Biotin-N-hydroxysuccinimide ester (0.2 mmol) was dissolved in DMF (10 mL) and added to the resin along with DIPEA (0.4 mmol). The resin slurry was heated by microwave irradiation in the Liberty Blue peptide synthesizer. Microwave heating proceeded as

follows: 15 W 75 ± 2 °C for 15 s, 30 W 90 ± 1 °C for 225 s. The resin was then washed once with DMF (2 mL) and the biotinylation reaction was repeated. The remainder of the TNRC2 fragment (Cys88-Asn116) was assembled onto the linker using automated SPPS (Section 5.3.2).

5.3.6. Peptide cleavage

The resin was washed with DMF (3×3 mL), DCM (3×3 mL) and Et₂O (3×3 mL). Cleavage cocktail was added to the resin and allowed to stir for 120 mins at room temperature.

Two cleavage cocktails were used in this work:

Cleavage reagent B containing 8.75 mL TFA, 0.25 mL triisopropylsilane, 0.5 g phenol and 0.5 mL water was used for TCPB-E's following solid-phase oxidation (5.3.12).

Cleavage reagent K containing 8.25 mL TFA, 0.25 mL EDT, 0.5 mL H₂O, 0.5 mL thioanisole and 0.5 g phenol was used for all other peptides.

The cleavage mixture was drained from the SPE column into a 50 mL centrifuge tube and the mixture was concentrated under a stream of N₂ to <3 mL. The peptide was precipitated using ice cold Et₂O and collected by centrifugation at 3500 RCF. The crude peptide was triturated twice with Et₂O, dissolved in 1% acetic acid (20 mL) and lyophilized. Crude peptides were analysed by LCMS and purified using preparative HPLC.

5.3.7. One-pot native chemical ligation and desulfurization of bacteriocins

Peptide hydrazide (2 μmol) was dissolved in 0.4 mL of 0.2 M sodium phosphate buffer containing 6 M Gn·HCl (pH 3.0-3.1) in a 2 mL centrifuge tube. The peptide solution was placed in an ice-salt bath at -15 °C and gently agitated for 10 mins. In a separate 2 mL centrifuge tube, the N-terminal cysteine peptide (2 μmol) was dissolved in 0.4 mL of 0.2 M sodium phosphate buffer containing 6 M Gn·HCl (pH 6.9-7.0). The peptide hydrazide was oxidized into the corresponding peptide azide by addition of 10 equiv.

NaNO₂ (aq. 0.5 M) and gently agitated at -15 °C for 20 mins.¹¹⁹ To convert the peptide azide to the thioester, the corresponding thiol was added* (2% or 4% v/v), the solution removed from the ice-salt bath and the pH adjusted to 5.0 at room temperature. Thioester conversion was allowed to proceed for 10 mins, followed by addition of the N-terminal cysteine peptide (2 μmol in 0.4 mL of 0.2 M phosphate, 6 M Gn·HCl, pH 6.9-7.0). The pH of the ligation mixture was adjusted to 6.8-6.9 and placed on a shaker at 37 °C for 4 hours. Reaction completion was confirmed by LCMS, by taking 10 μL of reaction mixture and quenching with 80 μL of pH 3.0 phosphate buffer and 10 μL of 0.1 M TCEP. Following ligation, excess thiol catalyst was removed**. The reaction mixture then was purged with Ar for 10 minutes. Desulfurization was carried out by addition of 0.8 mL desulfurization buffer (0.2 M phosphate, 6 M Gn·HCl, 400 mM TCEP, 40 mM GSH). The reaction was initiated by addition of VA-044 (80 mM) and tBuSH (80 mM) and heating on a shaker at 37 °C.¹⁸² The desulfurization was allowed to proceed for 16 hrs and complete conversion of cysteine to alanine was confirmed by LCMS. The reaction mixture was diluted 10-fold with water, passed through a 0.22 μm nylon filter and protein isolated by semi-preparative HPLC. (Final isolated yields: AucA variants, 45-70%, 5.3-8.3 mg; LngQ variants, 55-67%, 6.6-8.0 mg).

Table 5.1: Variables in Methods A and B for one-pot ligation desulfurization of AucA.

	*thiol catalyst	**removal of excess thiol
Method A	2% v/v thiophenol	Extraction with diethyl ether (8*0.5mL)
Method B	4% v/v trifluoroethanthiol ¹⁸⁴	Purging with Ar for 30 mins.

5.3.8. Native chemical ligation of TNRC2 for refolding optimization

Peptide hydrazide (Ser⁷²-Ser⁸⁷) (1.25 μmol) was dissolved in 0.5 mL of 0.2 M sodium phosphate buffer containing 6 M Gn·HCl (pH 3.0-3.1) in a 2 mL centrifuge tube. The peptide solution was placed in an ice-salt bath at -15 °C and gently agitated for 10 mins. In a separate 2 mL centrifuge tube, the N-terminal cysteine peptide (Cys⁸⁸-Asn¹¹⁶) (1 μmol) was dissolved in 0.5 mL of 0.2 M sodium phosphate buffer containing 6 M Gn·HCl (pH 6.9-7.0). The peptide hydrazide was oxidized into the corresponding

peptide azide by addition of 10 equiv. NaNO_2 (aq. 0.5 M) and gently agitated at $-15\text{ }^\circ\text{C}$ for 20 mins.¹¹⁹ To convert the peptide azide to the thioester, trifluoroethanethiol was added (2% v/v), the solution removed from the ice-salt bath and the pH adjusted to 5.0 at room temperature. Thioester conversion was allowed to proceed for 10 mins, followed by addition of the N-terminal cysteine peptide (1 μmol in 0.5 mL of 0.2 M phosphate, 6 M $\text{Gn}\cdot\text{HCl}$, pH 6.9-7.0). The pH of the ligation mixture was adjusted to 6.8-6.9, an additional quantity of trifluoroethanethiol was added (1% v/v, final conc. 2% v/v) and placed on a shaker at $37\text{ }^\circ\text{C}$ for 4 hours. Reaction completion was confirmed by LCMS, by taking 10 μL of reaction mixture and quenching with 80 μL of pH 3.0 phosphate buffer and 10 μL of 0.1 M TCEP. Following ligation, the reaction mixture was diluted 10-fold with water containing TCEP (50 μM), passed through a 0.22 μm nylon filter and reduced TNRC2 isolated by semi-preparative HPLC. (Final isolated yield: 72%, 4.3 mg).

5.3.9. Screening of TNRC2 refolding conditions

See Table S7.2 for summary of screening conditions and yields.

Optimization of oxidation method

Reduced TNRC2 (0.5 mg/mL – measured by UV absorbance²⁴⁸) was dissolved in 200 μL of denaturation buffer in a 2 mL centrifuge tube (i. 6 M $\text{Gn}\cdot\text{HCl}$, 0.25 M NaHCO_3 , 6 mM glutathione disulfide, 60 mM glutathione, pH 8.5 ; ii. 6 M $\text{Gn}\cdot\text{HCl}$, 0.25 M NaHCO_3 , 10% DMSO, pH 8.5 ; iii. 6 M $\text{Gn}\cdot\text{HCl}$, 0.25 M NaHCO_3 , pH 8.5 ; iv. 6 M $\text{Gn}\cdot\text{HCl}$, 0.1 M Tris, 6 mM glutathione disulfide, 60 mM glutathione, pH 8.5). Each reaction mixture i-iv was diluted five-fold with 800 μL renaturation buffer (i-iii 0.25 M NaHCO_3 , pH 8.5 ; iv. 0.1 M Tris, pH 8.5). Folding reactions proceeded at room temperature for 96 hours. 100 μL of reaction was removed, filtered through glass wool, and analyzed by LCMS. The refolding yield for each condition was estimated by the relative integration of the peak shifted up field by 1 min in the HPLC chromatogram at 210 nm (Table S7.2). Condition iv produced the highest yield (9%) and moved onto the next round of optimization below.

Optimization of buffer pH and temperature

Reduced TNRC2 (0.5 mg/mL – measured by UV absorbance²⁴⁸) was dissolved in 200 μ L of denaturation buffer in a 2 mL centrifuge tube (**i** & **iii**. 6 M Gn·HCl, 0.1 M Tris, 6 mM glutathione disulfide, 60 mM glutathione, pH 8.5 ; **ii** & **iv**. 6 M Gn·HCl, 0.1 M phosphate, 6 mM glutathione disulfide, 60 mM glutathione, pH 6.5). Each reaction mixture **i-iv** was diluted five-fold with 800 μ L renaturation buffer (**i** & **iii**. 0.1 M Tris, pH 8.5 ; **ii** & **iv**. 0.1 M phosphate, pH 6.5). Folding reactions proceeded for 96 hours, either at room temperature (**i-ii**) or at 4 °C (**iii-iv**). 100 μ L of reaction was removed, filtered through glass wool, and analyzed by LCMS. The refolding yield for each condition was estimated by the relative integration of the peak shifted up field by 1 min in the HPLC chromatogram at 210 nm (Table S7.2). Condition **ii** produced the highest yield (33%) and moved onto the next round of optimization below.

Optimization of reactant concentrations

Reduced TNRC2 (0.5 mg/mL **i-ii**, or 2 mg/mL **iii-v** – measured by UV absorbance²⁴⁸) was dissolved in 200 μ L of denaturation buffer in a 2 mL centrifuge tube (6 M Gn·HCl, 0.1 M phosphate, pH 6.5 containing ; **i**, **iii** & **v**. 6 mM glutathione disulfide, 120 mM glutathione ; **ii** & **iv**. 25 mM glutathione disulfide, 50 mM glutathione). Each reaction mixture **i-v** was diluted five-fold with 800 μ L renaturation buffer (0.1 M phosphate, pH 6.5). Folding reactions proceeded for 96 hours at room temperature. 100 μ L of reaction was removed, filtered through glass wool, and analyzed by LCMS. The refolding yield for each condition was estimated by the relative integration of the peak shifted up field by 1 min in the HPLC chromatogram at 210 nm (Table S7.2). Neither condition produced in improvement in refolding yield (<21%). Higher TNRC2 concentrations were desirable as this would result in smaller reaction volumes during scale-up preparations. However, folding at 0.4 mg/mL (**iii-v**) gave significantly reduced yield (5-13%). Therefore, the best condition (**ii**) from the previous round was used for TNRC2 refolding.

5.3.10. One-pot native chemical ligation and refolding of TNRC2

Peptide hydrazide (Ser⁷²-Ser⁸⁷) (12.5 μ mol) was dissolved in 5 mL of 0.1 M sodium phosphate buffer containing 6 M Gn·HCl (pH 3.0-3.1) in a 20 mL round bottomed flask. The peptide solution was placed in an ice-salt bath at -15 °C and gently agitated by magnetic stirring for 10 mins. In a 10 mL round bottomed flask, the N-terminal cysteine peptide (Cys⁸⁸-Asn¹¹⁶ or Cys⁸⁸-Asn¹¹⁶-Gly¹¹⁷-Gly¹²¹-Lys(Biotin)¹²²) (10 μ mol) was dissolved in 5 mL of 0.1 M sodium phosphate buffer containing 6 M Gn·HCl (pH 6.9-7.0). The peptide hydrazide was oxidized into the corresponding peptide azide by addition of 10 equiv. NaNO₂ (aq. 0.5 M) and gently stirred at -15 °C for 20 mins.¹¹⁹ To convert the peptide azide to the thioester, trifluoroethanethiol was added (2% v/v), the solution removed from the ice-salt bath and the pH adjusted to 5.0 at room temperature. Thioester conversion was allowed to proceed for 10 mins, followed by addition of the N-terminal cysteine peptide (10 μ mol in 5 mL of 0.1 M phosphate, 6 M Gn·HCl, pH 6.9-7.0). The pH of the ligation mixture was adjusted to 6.8-6.9 and an additional quantity of trifluoroethanethiol was added (1% v/v, final conc. 2% v/v). The flask was then placed in a heated water bath (37 °C) and stirred for 4 hours. Reaction completion was confirmed by LCMS, by taking 10 μ L of reaction mixture and quenching with 80 μ L of phosphate buffer (pH 3.0) and 10 μ L of TCEP (0.1 M). Following ligation completion, excess trifluoroethanethiol was removed by purging with argon for 2 hours at 37 °C. The reaction mixture (containing reduced ligation product) was transferred to a 2 L round bottomed flask, cooled to room temperature and diluted with 120 mL of buffer (0.1 M phosphate, 6 M Gn·HCl, pH 6.5). In a 250 mL conical flask, glutathione (120 mM) and glutathione disulfide (12 mM) were dissolved in 130 mL of buffer (0.1 M phosphate, 6 M Gn·HCl, pH 6.5)). The buffer containing glutathione/glutathione disulfide was added to the diluted ligation reaction mixture (final ligation product conc. = 0.5 mg/mL) and stirred for 10 mins at room temperature. The reaction mixture was then diluted five-fold with phosphate buffer (0.1 M phosphate, pH 6.5) and stirred for four days at room temperature under inert atmosphere. The folding reaction was then monitored by LCMS. The reaction mixture was filtered through a sintered glass funnel, and the folded protein was isolated by preparative HPLC (Final isolated yields: L-TNRC2, 17%, 9.1 mg; D-TNRC2, 15%, 8.0 mg; D-TNRC2(biotin) at 1/10th scale (1.0 μ mol), 14%, 0.84 mg).

5.3.11. TCPB-E solution-phase oxidation

TCPB-E was prepared by automated SPPS (Method 5.3.2) and isolated by preparative HPLC in fully reduced form. In a 10 mL round bottomed flask, the peptide (50 μ M) was dissolved in 5 mL of buffer (0.1 M phosphate, pH 6) containing 0.5 mM of either cystine or glutathione. The reaction mixtures were stirred for six hours at room temperature. Reaction completion was confirmed by LCMS, by taking 10 μ L of reaction mixture and diluting with 90 μ L of distilled water. The two disulfide bonded products for each cysteine and glutathione mixture were isolated by semi-preparative HPLC. LCMS analysis is provided in Figure S7.8.

5.3.12. TCPB-E solid-phase oxidation

The following procedure utilizes the differential reactivity of iodine towards orthogonally protected cysteine residues, based on a previously reported procedure in solution-phase.²²⁷ The method was optimized here for application on the solid-phase. Briefly, in non-polar solvents, iodine selectively removes Cys(Trt) groups and oxidizes cysteines to form disulfide bonds, with Cys(Acm) left intact. In the more polar DMF, Cys(Acm) is removed and enables formation of the second disulfide bond. This method circumvents the use of the commonly employed Cys(Mmt) and *N*-chlorosuccinimide, due to the high cost of the Fmoc-D-Cys(Mmt)-OH building block.

TCPB-E sequence was assembled by automated SPPS (Method 5.3.2) on a low-loading PEG-PS rink amide resin (0.19 mmol/g) to minimize unwanted intermolecular disulfide bond formation. For TCPB-E-**C1**, Fmoc-Cys(Trt)-OH was used at position 1, and Fmoc-Cys(Acm)-OH used at positions 4 and 11. For TCPB-E-**C4**, Fmoc-Cys(Trt)-OH was used at position 4, and Fmoc-Cys(Acm)-OH used at positions 1 and 11. First, the intermolecular disulfide bond with cysteine was formed. Boc-Cys(Trt)-OH (2 equiv., 60 mM) was dissolved in 50% TFE in DCM and added to the peptide resin. Then, an equal volume of iodine (1 equiv., 30 mM) in 50% TFE in DCM was also added to the resin, and the slurry stirred for 10 mins at room temperature. The resin was washed with DCM (3 mL) and the reaction was repeated. The resin was washed once with DCM (3 mL) and three times with DMF (3

x 3 mL). For intramolecular disulfide bond formation, ten equivalents of Iodine in DMF (0.5 M) was added to the resin, and the slurry stirred for 60 mins at room temperature. The resin was then washed three times with DMF (3 × 3 mL), once with 1 M aq. ascorbic acid (3 mL), three times with water (3 × 3 mL) and three times with DMF (3 × 3 mL). The peptide was cleaved from the resin and isolated by preparative HPLC.

5.4. Circular dichroism spectroscopy

Solutions of the proteins were dissolved in water at a concentration of 40 μ M. CD spectra were recorded from 250 to 190 nm at 20 °C. Each measurement was performed in triplicate using a sample cell with a 0.1 mm path, 1 nm bandwidth and 0.5 s per point. CD spectra of the solvents were subtracted from the CD spectra of the proteins to eliminate interference. The units of ellipticity are expressed as the mean residue ellipticity ($[\theta]_{MR}$) in $\text{deg}\cdot\text{cm}^2\cdot\text{dmol}^{-1}\cdot\text{res}^{-1}$. CD data are available in appendices: AucA and variants (Figure S7.2), LncQ and variants (Figure S7.3), TMPB proteins (Figure S7.18).

5.5. Racemic protein crystallography

5.5.1. AucA crystallization

L-AucA and D-AucA were dissolved in water to a final concentration of 80 mg/mL, as determined by UV absorbance at 280 nm with a calculated molar absorptivity of 31970 $\text{M}^{-1}\text{cm}^{-1}$.²⁴⁸ The peptide solutions were mixed 1:1 to yield an 80 mg/mL racemate of DL-AucA which was diluted two-fold with water to yield 40 mg/mL DL-AucA. Both 80 mg/mL and 40 mg/mL racemate concentrations were subject to sparse-matrix crystallization screening using Crystal Screen HT (HR2-130) and SaltRx HT (HR2-136) from Hampton research. 50 μ L of each precipitant condition solution was added into the wells of a SWISSCI 96-well plate. The two racemate concentrations were each mixed 1:1 with the precipitant in a 0.4 μ L sitting drop, yielding 384 crystallisation drops across two screens. The best conditions which produced single crystal blocks were selected for optimization to produce crystals suitable for X-ray diffraction.

AucA crystal 1 (AucA-SO₄²⁻ complex) was formed in the sitting drop made with 0.5 μ L 40 mg/mL DL-AucA and 0.5 μ L precipitant composed of 0.2 M ammonium sulfate, 0.1 M sodium acetate and 24.5% v/v PEG 4000 at pH 4.6.

AucA crystal 2 (AucA apo) was formed in the sitting drop made with 0.5 μ L 40 mg/mL DL-AucA and 0.5 μ L precipitant composed of 0.2 M ammonium acetate, 0.2 M sodium citrate and 29% PEG 4000 v/v at pH 5.6.

To explore the binding of AucA with L-glycerol 3-phosphate (G3P), L-G3P (Generon) was dissolved in 0.2 M ammonium acetate, 0.2 M sodium citrate and 29% PEG 4000 v/v to a final concentration of 334 mM. The G3P stock was used to soak the apo DL-AucA crystals formed in the condition for Crystal 2 at 1:1 v/v for 24 hours (50 equivalents). The resulting crystal is referred to herein as **AucA crystal 3** (AucA-G3P).

AucA crystal 4 (AucA dimer) was formed in the sitting drop made with 0.2 μ L 40 mg/mL DL-AucA and 0.2 μ L precipitant composed of 0.7 M sodium citrate, 0.1 M Tris at pH 8.5.

5.5.2. LnqQ crystallization

L-LnqQ and D-LnqQ were dissolved in water to a final concentration of 27 mg/mL, as determined by UV absorbance at 280 nm with a calculated molar absorptivity of 23490 M⁻¹ cm⁻¹.²⁴⁸ The peptide solutions were mixed 1:1 to yield a 27 mg/mL racemate of DL-LnqQ. Half of the solution was diluted two-fold with water to yield 13.5 mg/mL DL-LnqQ. The two racemate concentrations were subject to sparse-matrix crystallization screening using Crystal Screen HT (HR2-130) and SaltRx HT (HR2-136) from Hampton research. 50 μ L of each precipitant condition was added into the wells of a SWISSCI 96-well plate. The two racemate concentrations were each mixed 1:1 with the precipitant in a 0.4 μ L sitting drop, yielding 384 crystallisation drops across two screens. The best initial condition which produced single, three-dimensional crystals was selected for optimization to produce crystals suitable for X-ray diffraction. The best **LnqQ crystal** was formed with 4.0 M sodium formate as the precipitant at pH 7.0 in a 1 μ L sitting drop.

5.5.3. TNRC2 crystallization

L-TNRC2 and D-TNRC2 were dissolved in water to a final concentration of 25 mg/mL, as determined by UV absorbance at 280 nm.²⁴⁸ The peptide solutions were mixed 1:1 to yield a 25 mg/mL racemate of DL-TNRC2. Half of the solution was diluted two-fold with water to yield 12.5 mg/mL DL-TNRC2. The two racemate concentrations were subject to sparse-matrix crystallization screening using Crystal Screen HT (HR2-130 from Hampton research. 50 µL of each precipitant condition was added into the wells of a SWISSCI 96-well plate. The two racemate concentrations were each mixed 1:1 with the precipitant in a 0.5 µL sitting drop, yielding 192 crystallisation drops. The best single, three-dimensional crystals were sent for X-ray diffraction data collection. The best **TNRC2 crystal** was formed in the sitting drop made with 0.25 µL 25 mg/mL DL-TNRC2 and 0.25 µL precipitant composed of 1.5 M Sodium chloride, 10% v/v ethanol at pH 8.5.

5.5.4. X-ray diffraction data collection

The crystals were fished from the sitting drop, dipped into cryoprotectant (for **AucA crystals 1-3**: 20% PEG 400, for **AucA crystal 4**, **LnqQ crystal**, and **TNRC2 crystal**: 2.0 M Li₂SO₄,) before flash-frozen in liquid nitrogen. X-ray diffraction data were collected at the Diamond Light Source synchrotron, with beamline I04-1 using a Dectris Pilatus 6M-F detector (**AucA crystals 1 and 2**) or beamline I03 using a Dectris Eiger2 XE 16M detector (**all other crystals**). The collected datasets were processed with Xia2, and data scaling performed with Aimless.²⁴⁹ The crystal space groups for data reductions were validated using Zandua.²⁵⁰

5.5.5. AucA structure solution

AucA crystal 1: The structure of two L-AucA molecules in the protein crystal were first solved through molecular replacement in MOLREP, using state one of the solution NMR structure PDB 2N8O as a search model.²⁵¹ This led to the calculated phases delivering a clear electron density for two D-AucA protein molecules. With these

phases, the peptide backbones of D-AucA were partially built by ARP/wARP²⁵² as a poly-glycine chain. The full crystal structure was then built through iterative rounds of manual model building using COOT²⁵³ and anisotropic B-factor refinement via REFMAC.²⁵⁴

AucA crystals 2-4: The structure L-AucA molecules in the protein crystal were first solved through molecular replacement with MOLREP, using state one of the solution NMR structure PDB 2N8O as a search model. D-AucA chain from **AucA crystal 1** was truncated to a poly-glycine chain in PyMOL and used as a second search model in molecular replacement. The full crystal structure was then built through iterative rounds of manual model building using COOT and anisotropic B-factor refinement.

Data refinement statistics are available in appendices 7.4. The refined models of racemic AucA have been deposited in the Protein Data Bank²⁵⁵ with the PDB codes 8AVR, 8AVS, 8AVU and 8AVT.

5.5.6. LnqQ structure solution

The structure of two L-LnqQ molecules in the protein crystal were first solved through molecular replacement, using ideal 10-residue helices as search models in Fragon.²⁵⁶ This led to the calculated phases delivering a clear electron density for two D-LnqQ protein molecules. With these phases, the peptide backbones of D-LnqQ were partially built by Buccaneer²⁵⁷ as a poly-glycine chain. The full crystal structure was then built through iterative rounds of manual model building using COOT²⁵³ and anisotropic B-factor refinement via REFMAC.²⁵⁴

Data refinement statistics are available in appendices 7.4. The refined model of racemic LnqQ has been deposited in the Protein Data Bank²⁵⁵ with the PDB code 7P5R.

5.5.7. TNRC2 structure solution

The structure of the L-TNRC2 molecule in the protein crystal was first solved through molecular replacement in MOLREP, using residues 72-116 of the TNFR-1 X-ray crystal structure PDB 1EXT as a search model.²⁵⁸ This led to the calculated phases delivering a clear electron density for the D-TNRC2 protein molecules. The structure of D-TNRC2, and the remaining waters and solute, was then built through iterative rounds of manual model building using COOT²⁵³ and anisotropic B-factor refinement via REFMAC.²⁵⁴

Data collection statistics are available in appendices 7.4. Finalized data refinement statistics are not provided, as deposition in the Protein Data Bank²⁵⁵ is ongoing.

5.6. Antibiotic activity assay

The following assay was conducted by Laura Dominguez at Concordia University.

Minimum Inhibitory Concentrations (MIC) were determined for each of the leaderless bacteriocins and controls using a broth microdilution method, as per CSLI guidelines for microbial susceptibility testing.²⁵⁹ The peptides were dissolved in sterile deionized water to prepare the working stocks, with concentrations determined by UV absorbance at 280 nm using calculated molar absorptivity's.²⁴⁸ Overnight cultures from the strains of interest were incubated shaking overnight at 37 °C. Cation-adjusted Mueller Hinton Broth 2 (MHB) was used for the *S. aureus*, and *S. epidermidis* strains and Brain Heart Infusion (BHI) broth was utilized for the *E. faecalis* and *E. faecium* strains. The cultures were prepared and diluted to the turbidity of a 0.5 McFarland standard using MHB or BHI fresh media. Increasing concentrations of the leaderless bacteriocins were added to a 96-well polypropylene plate. The range tested for each bacteriocin was 64-0.125 µg/mL. The bacterial culture was further diluted and mixed in a 1:1 ratio with the bacteriocin solution to yield a final concentration in the wells of 5×10^5 CFU/mL. Ampicillin, tetracycline and melittin were used as positive controls and included in each assay. A 1:1000 dilution of the growth control was prepared in the respective media and plated onto agar (100 µL) to ensure the correct inoculum

concentration (5×10^5 CFU/mL indicated by ≈ 50 bacterial colonies). The 96-well and agar plates were incubated at 37 °C, and growth was assessed by the formation of a pellet observable with the naked eye after 20 h. Uninhibited bacterial growth at 64 μ g/mL was denoted as MIC >64 μ g/mL and for the purposes of this work was classified as loss of activity. Each assay was performed in triplicate.

5.7. Analytical size exclusion chromatography

Analytical size exclusion chromatography was performed using an Agilent Bio SEC-3 column (3 μ m, 150 Å, 4.6 x 500 mm) using an Agilent 1260 infinity II system fitted with an autosampler. The column was held at 37 °C employing phosphate buffered saline at pH 7.4 over a 30 min isocratic run. First, a gel filtration standard from BIO-RAD containing Thyroglobulin (670 kDa, **T**), γ -globulin (158 kDa, **G**), Ovalbumin (44 kDa, **O**), Myoglobin (17 kDa, **M**) and Vitamin B12 (1.35 kDa, **V**) was analyzed by HP-SEC, with a chromatogram recorded by UV absorbance at 210 nm. The log(MW) of the proteins in the standard were plotted against the retention time to create a calibration plot, allowing the conversion of HP-SEC retention time to an approximate MW of the protein solution state. AucA and variants were run with 10 μ g column loadings (20 μ L of 0.5 mg/mL protein in mobile phase). AucA eluted with a higher up-field retention than anticipated (HP-SEC estimated MW: 2 kDa, actual dimer MW: 12 kDa), likely due to the compact, globular protein folding. The monomeric bacteriocin LnqQ was used as a reference sample to distinguish dimeric AucA variants from the monomeric variant. Data available in Figure S7.6.

5.8. AucA proteolytic stability assay

Proteolytic digestions were performed in the following buffers:

- Buffer **A**: 50 mM Sodium phosphate, pH 7.4, 10 mM EDTA, 10 mM Cysteine.
- Buffer **B**: 50 mM Sodium phosphate, pH 7.4, 5 mM CaCl₂, 100 mM NaCl.

Protease stocks were prepared at 80 μM concentration in the appropriate buffer (buffer **A** – papain and buffer **B** – proteinase K). Stocks of AucA, AucA-W22L and AucA W3L were also prepared in the corresponding buffer **A** or buffer **B** (89 μM). Proteolysis reactions were initiated by the addition of protease stock (10 μL) to AucA stock (90 μL) yielding a 1:10 ratio of protease to AucA. Reactions were shaken at 37 °C for 12 hr. 5 μL of reaction mixture was quenched in 45 μL of 50% Acetonitrile in water + 0.1% trifluoroacetic acid and analyzed by LCMS with a 20 μL injection volume by autosampler (Agilent 1260 infinity II). UV chromatogram at 210 nm was used for integration (Figure S7.7).

5.9. Bacteriophage bio-panning

The bacteriophage bio-panning experiments were performed by the group of Prof. Chuanliu Wu at Xiamen University, China. Briefly, a library of peptide sequences with nine randomized amino acid positions flanked by two cysteines (CX_9C) were displayed on the M13 phages, where X is any amino acid encoded by NNK. The peptides were linked to the gene-3 coat protein (pIII) *via* a tri-alanine (Ala-Ala-Ala) spacer. Gel purified DNA encoding the libraries were ligated into the pCantab 5E phagemid vector, and transformed into chemically competent *E. coli* K12 ER2738. 20-fold excess of helper phage M13KO7 was used during rescue and amplification of the bacteriophages. The target protein (D-TNRCD2) containing C-terminal GsGsGK(biotin) was immobilized (30 μg) onto streptavidin- or neutravidin-coated beads, alternating between panning rounds to reduce enrichment of non-specific binders. Phage enrichment factor for each round is defined as the amount of phage eluted from the target well containing D-TNRCD2 divided by the amount of phage eluted from the target well without immobilized target (see Tables S7.3 and S7.5). For enriched sequences, see Tables S7.4 and S7.6.

5.10. LCMS disulfide bond mapping of TCPB-E's

To a solution containing the oxidized TCPB-E peptides (isolated by HPLC in 5.3.11, ~1-10 μg) in 90 μL of buffer (0.1 M phosphate, pH 6), 10 μL aqueous solution of

chymotrypsin (1 mg/mL) was added. The reaction mixture was shaken in a 2 mL centrifuge tube at room temperature for 6-12 hours until complete digestion of the oxidized peptide was observed. The digested fragments were analyzed by LCMS (Figure S7.9 – S7.12), and/or compared to the orthogonally prepared TCPB-E peptides (Figure S7.15).

5.11. Grating coupled interferometry

Grating-coupled interferometry (GCI) experiments were performed on a Creoptix WAVEsystem using 4PCH-STA WAVE sensor chips (polycarboxylate surface, streptavidin coated). Chips were first conditioned with borate buffer (0.1 M sodium borate, pH 9, 1M Sodium chloride). D-TNRCD2 containing the C-terminal flexible biotinylated linker (100 µg/mL) in HBS EP buffer (10 mM HEPES, 150 mM sodium chloride, 3 mM EDTA, 0.05% surfactant P20) was immobilized onto the sensor chip by injection over the sensor surface at 10 µL/min flow rate, followed by passivation of the surface with blank HBS EP buffer. A final immobilization density of 2300 pg/mm² D-TNRCD2(biotin) was obtained. All subsequent GCI experiments were conducted in the HBS EP running buffer.

Initial binding analysis was performed using either (a) waveRAPID (Repeated Analyte Pulses of Increasing Duration) or (b) a single pass of association and dissociation over the sensor surface. Binding kinetics analysis of TCPB-E-**C1** was conducted using (c) multi-cycle kinetics experiment.

- a) For waveRAPID experiments, solutions of TCPB-E peptide conformers were prepared in HBS-EP buffer (TCPB-E-**C1** = 100 µM, TCPB-E-**C11** = 250 µM , TCPB-E-**G1** = 200 µM, TCPB-E-**G11** = 200 µM). The peptide solutions (122 µL) were passed over the immobilized D-TNRCD2 flow cell and a blank reference flow cell at a flow rate of 120 µL/min for 25 s total injection time (repeat pulses of association and dissociation), followed by a 300 s dissociation phase with the HBS EP buffer. GCI sensograms are shown in Figure S7.13.
- b) The single pass of association and dissociation were used to differentiate binding of TCPB-E-**C1** and TCPB-E-**C4**. The peptides were prepared in HBS-EP buffer at 50 µM concentration. The peptide solutions were passed over the

immobilized D-TNRCD2 flow cell and a blank reference flow cell at a flow rate of 30 $\mu\text{L}/\text{min}$ for 100 s total injection time, followed by a 120 s dissociation phase with the HBS EP buffer. GCI sensograms are shown in Figure S7.13.

- c) Solutions of TCPB-E-**C1** peptide were prepared at serial two-fold dilutions (as indicated in legend) in HBS-EP buffer. The peptide solutions were passed over the immobilized D-TNRCD2 flow cell and a blank reference flow cell at a flow rate of 30 $\mu\text{L}/\text{min}$ for 100 s total injection time, followed by a 120 s dissociation phase with the HBS EP buffer. Data were fitted with the one-to-one Langmuir binding model and binding affinity analysis conducted in the Creoptix Wave control software. GCI sensograms are shown in Figure 3.6. Analysis of binding of D-TCPB-E-**C1** binding to L-TNRCD2 was conducted in the same manner, with a L-TNRCD2 surface density of 1000 pg/mm^2 .

In all experiments, blank injections were used for double referencing, and a DMSO calibration curve was used for bulk correction. Initial binding experiments of the TMPB proteins were conducted using method a) at varying concentrations (10 – 100 μM), but no binding to D-TNRCD2 was observed (data not shown).

5.12. ROSETTA *de novo* protein binder design

The target structure D-TNRCD2 was obtained from the racemic protein crystal structure PDB file 5.5.7. Mini protein scaffold coordinate files (PDB ID's: 5UYO, 5UOI, 5UP1 and 5UP5) were downloaded from the Protein Data Bank²⁵⁵. Rosetta programs (Version 3.13) used in this work were run through the RosettaScripts interface, using an XML script file and an 'options' flag file (examples of the command line to run RosettaScripts, with scripts and flags used in this work, are included in Appendices 7.5).

Initial trial experiments below were conducted on a Lenovo IdeaPad 320S, with Intel DualCore i5-8250U CPU (1.6 GHz and 1.8 GHz) and 8.00 GiB of RAM.

The structures of the target and scaffold proteins were first relaxed using the Relax script (Appendix 7.5.2) with Relax flag (Appendix 7.5.3). Each scaffold protein was

individually combined with the target coordinate file using PyMOL, positioning the two proteins at some random distance apart.

For initial global docking (where each amino acid side chain is modelled as a unified pseudo-atom – centroid) trials of the four scaffold proteins, GlobalDock script (Appendix 7.5.4) with Dock flag (Appendix 7.5.5) was used. Approximately 1000 docked structures (-nstruct1000) was conducted for each scaffold. The GlobalDock script contained filters to define which docking results provided binders near to the TNF- α binding interface (BurResTNF). Structures which gave BurResTNF >1600 Å were considered as initial candidates.

Preliminary design rounds at the interface between docked 5UOI and D-TNRCD2 showed better binder design (lower energy) when allowing some flexibility of the D-protein target, thus target flexibility was implemented in all future designs (FlexDesign script, Appendix 7.5.6).

The docked structures of the four scaffold proteins were each subject to 20 rounds of interface design, using FlexDesign script (Appendix 7.5.6) and Design flag (Appendix 7.5.7). The FlexDesign script contains a layer design, adapted from the literature,²³² which defines the allowed residues for design in the miniprotein based on their solvent accessibility (i.e. charged residues only at the surface, no hydrophobic residues exposed to the solvent). In addition, proline, glycine and cysteine substitutions were disallowed. The FlexDesign script also contained a filter for simulated binding energy (Ddg), which provides an estimate of the binding energies between target and protein binder, subsequently used to identify potential candidates. Scaffold 5UOI consistently gave better simulated binding energies, thus was selected as the protein scaffold for binder design.

Subsequent Rosetta experiments below were conducted using the BlueCrystal4 at the University of Bristol, Advanced Computer Research Centre. One Lenovo nx360 m5 compute node was used per submitted experiment, which has two 14 core 2.4 GHz Intel E5-2680 v4 (Broadwell) CPUs, and 128 GiB of RAM.

The same workflow as previous was implemented for 5UOI only, to cover a broader scope of docking and design. GlobalDock (Appendix 7.5.4) with Dock flag (Appendix 7.5.5) was conducted for 5UOI, generating 5500 docked structures, with the top 20

filtered by BurResTNF. Each of the 20 docked poses were then subjected to 100 rounds of interface design (2000 initial binders) using the FlexDesign script (Appendix 7.5.6) with Design flag (Appendix 7.5.7), and the best 100 candidates were sorted by their simulated binding energies. Following multiple sequence alignment of the best designs, and further refinement of the design aided by visual inspection in PyMOL, the candidate protein binders were identified.

For docking validation of the designed proteins to D-TNRCD2, the FullDock script (Appendix 7.5.8) was used with FullDock flag (Appendix 7.5.9), generating 10000 docked structures. The coordinate files of each output from design were used as a reference to yield a root mean square deviation (RMSD) of the output docked structure to the output designed structure. When poorly designed binders are positioned away from the target, and re-docked using the FullDock script, the resulting binding energies will show poor correlation with RMSD. For the best binders, the lowest energy (D_{dg}) from docking is obtained at the lowest RMSD to the designed structure, as the binder is designed to bind tightly in this position only. The resulting plots of D_{dg} vs RMSD (for the 500 lowest RMSD structures), along with visual representations, are available in Figure S.17.

Chapter 6 : References

1. V. A. Davankov, *Symmetry-Basel*, 2018, **10**.
2. G. Carenzi, S. Sacchi, M. Abbondi and L. Pollegioni, *Amino Acids*, 2020, **52**, 849-862.
3. S. K. Kuncha, S. P. Kruparani and R. Sankaranarayanan, *J. Biol. Chem.*, 2019, **294**, 16535-16548.
4. L. M. Dedkova, N. E. Fahmi, S. Y. Golovine and S. M. Hecht, *Biochemistry*, 2006, **45**, 15541-15551.
5. Y. Goto, H. Murakami and H. Suga, *Rna*, 2008, **14**, 1390-1398.
6. J. Kobayashi, Y. Shimizu, Y. Mutaguchi, K. Doi and T. Ohshima, *J. Mol. Catal. B: Enzym.*, 2013, **94**, 15-22.
7. Y. Ogasawara and T. Dairi, *Front. Microbiol.*, 2018, **9**, 156.
8. R. Milton, S. Milton and S. Kent, *Science*, 1992, **256**, 1445-1448.
9. L. Pasteur, *Société Chimique de Paris*, 1860, **Reprint No. 14 (Alembic Club, 1905)**.
10. M. Peplow, *Nat. News*, 2016, **533**, 303.
11. J. W. Szostak, D. P. Bartel and P. L. Luisi, *Nature*, 2001, **409**, 387-390.
12. A. C. Forster and G. M. Church, *Mol. Syst. Biol.*, 2006, **2**, 45.
13. D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma and M. M. Moodie, *Science*, 2010, **329**, 52-56.
14. W. Jiang, B. Zhang, C. Fan, M. Wang, J. Wang, Q. Deng, X. Liu, J. Chen, J. Zheng, L. Liu and T. F. Zhu, *Cell Discov.*, 2017, **3**, 17037.
15. C. Fan, Q. Deng and T. F. Zhu, *Nat. Biotechnol.*, 2021, 1548-1555.
16. Schrodinger, LLC, *Journal*, 2015.
17. Z. Wang, W. Xu, L. Liu and T. F. Zhu, *Nat. Chem.*, 2016, **8**, 698-704.
18. A. Pech, J. Achenbach, M. Jahnz, S. Schülzchen, F. Jarosch, F. Bordusa and S. Klussmann, *Nucleic Acids Res.*, 2017, **45**, 3997-4005.
19. M. Wang, W. Jiang, X. Liu, J. Wang, B. Zhang, C. Fan, L. Liu, G. Pena-Alcantara, J.-J. Ling, J. Chen and T. F. Zhu, *Chem*, 2019, **5**, 848-857.
20. J. Cline, J. C. Braman and H. H. Hogrefe, *Nucleic Acids Res.*, 1996, **24**, 3546-3551.
21. J. Weidmann, M. Schnölzer, P. E. Dawson and J. D. Hoheisel, *Cell Chem. Biol.*, 2019, **26**, 645-651.e643.

22. J. de la Cruz, K. Karbstein and J. L. Woolford Jr, *Annu. Rev. Biochem.*, 2015, **84**, 93-129.
23. J.-J. Ling, C. Fan, H. Qin, M. Wang, J. Chen, P. Wittung-Stafshede and T. F. Zhu, *Angew. Chem. Int. Ed.*, 2020, **59**, 3724-3731.
24. S. Klinge and J. L. Woolford, *Nat. Rev. Mol. Cell Biol.*, 2019, **20**, 116-131.
25. Y. Liu, E. Holmstrom, J. Zhang, P. Yu, J. Wang, M. A. Dyba, D. Chen, J. Ying, S. Lockett and D. J. Nesbitt, *Nature*, 2015, **522**, 368-372.
26. Y. Ishihama, T. Schmidt, J. Rappsilber, M. Mann, F. U. Hartl, M. J. Kerner and D. Frishman, *BMC Genom.*, 2008, **9**, 102.
27. F. Rohden, J. D. Hoheisel and H.-J. Wieden, *Trends Biochem. Sci.*, 2021, **46**, 931-943.
28. W.-S. Song, S.-X. Liu and C.-C. Chang, *J. Org. Chem.*, 2018, **83**, 14923-14932.
29. T. O. Yeates and S. B. H. Kent, *Annu. Rev. Biophys.*, 2012, **41**, 41-61.
30. S. W. Wukovitz and T. O. Yeates, *Nat. Struct. Biol.*, 1995, **2**, 1062-1067.
31. L. E. Zawadzke and J. M. Berg, *Proteins: Struct. Funct. Genet.*, 1993, **16**, 301-305.
32. D. F. Kreitler, Z. Yao, J. D. Steinkruger, D. E. Mortenson, L. Huang, R. Mittal, B. R. Travis, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2019, **141**, 1583-1592.
33. D. E. Mortenson, J. D. Steinkruger, D. F. Kreitler, D. V. Perroni, G. P. Sorenson, L. Huang, R. Mittal, H. G. Yun, B. R. Travis, M. K. Mahanthappa, K. T. Forest and S. H. Gellman, *Proc. Natl. Acad. Sci.*, 2015, **112**, 13144-13149.
34. S. Gao, M. Pan, Y. Zheng, Y. C. Huang, Q. Y. Zheng, D. M. Sun, L. N. Lu, X. D. Tan, X. L. Tan, H. Lan, J. X. Wang, T. Wang, J. W. Wang and L. Liu, *J. Am. Chem. Soc.*, 2016, **138**, 14497-14502.
35. K. Mandal, B. L. Pentelute, D. Bang, Z. P. Gates, V. Y. Torbeev and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2012, **51**, 1481-1486.
36. B. L. Pentelute, Z. P. Gates, V. Tereshko, J. L. Dashnau, J. M. Vanderkooi, A. A. Kossiakoff and S. B. H. Kent, *J. Am. Chem. Soc.*, 2008, **130**, 9695-9701.
37. J. R. Banigan, K. Mandal, M. R. Sawaya, V. Thammavongsa, A. P. A. Hendrickx, O. Schneewind, T. O. Yeates and S. B. H. Kent, *Protein Sci.*, 2010, **19**, 1840-1849.

38. A. J. Lander, X. Li, Y. Jin and L. Y. P. Luk, 2020, ChemRxiv preprint DOI: 10.26434/chemrxiv.12444554.v12444551.
39. Y.-H. Huang, Q. Du, Z. Jiang, G. J. King, B. M. Collins, C. K. Wang and D. J. Craik, *Molecules*, 2021, **26**, 5554.
40. Q. Qu, S. Gao, F. Wu, M.-G. Zhang, Y. Li, L.-H. Zhang, D. Bierer, C.-L. Tian, J.-S. Zheng and L. Liu, *Angew. Chem. Int. Ed.*, 2020, **59**, 6037-6045.
41. C. Zuo, B. Zhang, M. Wu, D. Bierer, J. Shi and G.-M. Fang, *Chin. Chem. Lett.*, 2019, **31**, 693-696.
42. C. Zuo, W.-W. Shi, X.-X. Chen, M. Glatz, B. Riedl, I. Flamme, E. Pook, J. Wang, G.-M. Fang, D. Bierer and L. Liu, *Sci. China Chem.*, 2019, **62**, 1371-1378.
43. K. W. Kurgan, A. F. Kleman, C. A. Bingman, D. F. Kreitler, B. Weisblum, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2019, **141**, 7704-7708.
44. S. D. Ramalho, C. K. Wang, G. J. King, K. A. Byriel, Y.-H. Huang, V. S. Bolzani and D. J. Craik, *J. Nat. Prod.*, 2018, **81**, 2436-2445.
45. Q. Qu, S. Gao and Y.-M. Li, *J. Pept. Sci.*, 2018, **24**, e3112.
46. C. K. Wang, G. J. King, A. C. Conibear, M. C. Ramos, S. Chaousis, S. T. Henriques and D. J. Craik, *J. Am. Chem. Soc.*, 2016, **138**, 5706-5713.
47. H. Yeung, C. J. Squire, Y. Yosaatmadja, S. Panjekar, G. López, A. Molina, E. N. Baker, P. W. R. Harris and M. A. Brimble, *Angew. Chem. Int. Ed.*, 2016, **55**, 7930-7933.
48. Y.-C. Huang, C.-C. Chen, S. Gao, Y.-H. Wang, H. Xiao, F. Wang, C.-L. Tian and Y.-M. Li, *Chem. Eur. J.*, 2016, **22**, 7623-7628.
49. M. Pan, S. Gao, Y. Zheng, X. Tan, H. Lan, X. Tan, D. Sun, L. Lu, T. Wang, Q. Zheng, Y. Huang, J. Wang and L. Liu, *J. Am. Chem. Soc.*, 2016, **138**, 7429-7435.
50. D. E. Mortenson, K. A. Satyshur, I. A. Guzei, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2012, **134**, 2473-2476.
51. D. F. Kreitler, D. E. Mortenson, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2016, **138**, 6498-6505.
52. B. Dang, T. Kubota, K. Mandal, A. M. Correa, F. Bezanilla and S. B. Kent, *Angew. Chem. Int. Ed. Engl.*, 2016, **55**, 8639-8642.
53. Z. Hayouka, D. E. Mortenson, D. F. Kreitler, B. Weisblum, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2013, **135**, 15738-15741.

54. Z. Hayouka, N. C. Thomas, D. E. Mortenson, K. A. Satyshur, B. Weisblum, K. T. Forest and S. H. Gellman, *J. Am. Chem. Soc.*, 2015, **137**, 11884-11887.
55. B. Dang, T. Kubota, K. Mandal, F. Bezanilla and S. B. Kent, *J. Am. Chem. Soc.*, 2013, **135**, 11911-11919.
56. J. K. Murray, Y.-X. Qian, B. Liu, R. Elliott, J. Aral, C. Park, X. Zhang, M. Stenkilsson, K. Salyers, M. Rose, H. Li, S. Yu, K. L. Andrews, A. Colombero, J. Werner, K. Gaida, E. A. Sickmier, P. Miu, A. Itano, J. McGivern, C. V. Gegg, J. K. Sullivan and L. P. Miranda, *J. Med. Chem.*, 2015, **58**, 6784-6802.
57. B. Dang, R. Shen, T. Kubota, K. Mandal, F. Bezanilla, B. Roux and S. B. Kent, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 3324-3328.
58. R. D. Bunker, K. Mandal, G. Bashiri, J. J. Chaston, B. L. Pentelute, J. S. Lott, S. B. H. Kent and E. N. Baker, *Proc. Natl. Acad. Sci.*, 2015, **112**, 4310-4315.
59. C. K. Wang, G. J. King, S. E. Northfield, P. G. Ojeda and D. J. Craik, *Angew. Chem. Int. Ed.*, 2014, **53**, 11236-11241.
60. R. Okamoto, K. Mandal, M. R. Sawaya, Y. Kajihara, T. O. Yeates and S. B. H. Kent, *Angew. Chem. Int. Ed.*, 2014, **53**, 5194-5198.
61. M. Avital-Shmilovici, K. Mandal, Z. P. Gates, N. B. Phillips, M. A. Weiss and S. B. H. Kent, *J. Am. Chem. Soc.*, 2013, **135**, 3173-3185.
62. K. Mandal, B. Dhayalan, M. Avital-Shmilovici, A. Tokmakoff and S. B. H. Kent, *Chembiochem*, 2016, **17**, 421-425.
63. K. Mandal, M. Uppalapati, D. Ault-Riché, J. Kenney, J. Lowitz, S. S. Sidhu and S. B. H. Kent, *Proc. Natl. Acad. Sci.*, 2012, **109**, 14779-14784.
64. B. L. Pentelute, K. Mandal, Z. P. Gates, M. R. Sawaya, T. O. Yeates and S. B. H. Kent, *Chem. Commun.*, 2010, **46**, 8174-8176.
65. N. Marín-Medina, D. A. Ramírez, S. Trier and C. Leidy, *Appl. Microbiol. Biotechnol.*, 2016, **100**, 10251-10263.
66. S. McNicholas, E. Potterton, K. S. Wilson and M. E. M. Noble, *Acta Crystallographica Section D-Structural Biology*, 2011, **67**, 386-394.
67. G. L. Rosano and E. A. Ceccarelli, *Front. Microbiol.*, 2014, **5**, 172-172.
68. L. P. Kozlowski, *Nucleic Acids Res.*, 2017, **45**, D1112-D1116.
69. S. Lien and H. B. Lowman, *Trends Biotechnol.*, 2003, **21**, 556-562.
70. J. L. Lau and M. K. Dunn, *Biorg. Med. Chem.*, 2018, **26**, 2700-2707.
71. M. Goodman and M. Chorev, *Acc. Chem. Res.*, 1979, **12**, 1-7.

72. C. Li, C. Y. Zhan, L. Zhao, X. S. Chen, W. Y. Lu and W. Y. Lu, *Biorg. Med. Chem.*, 2013, **21**, 4045-4050.
73. C. Li, M. Pazgier, J. Li, C. Q. Li, M. Liu, G. Z. Zou, Z. Y. Li, J. D. Chen, S. G. Tarasov, W. Y. Lu and W. Y. Lu, *J. Biol. Chem.*, 2010, **285**, 19572-19581.
74. J. J. Miles, M. P. Tan, G. Dolton, E. S. J. Edwards, S. A. E. Galloway, B. Laugel, M. Clement, J. Makinde, K. Ladell, K. K. Matthews, T. S. Watkins, K. Tungatt, Y. Wong, H. S. Lee, R. J. Clark, J. M. Pentier, M. Attaf, A. Lissina, A. Ager, A. Gallimore, P. J. Rizkallah, S. Gras, J. Rossjohn, S. R. Burrows, D. K. Cole, D. A. Price and A. K. Sewell, *J. Clin. Investig.*, 2018, **128**, 1569-1580.
75. S. Imanishi, T. Katoh, Y. Yin, M. Yamada, M. Kawai and H. Suga, *J. Am. Chem. Soc.*, 2021, **143**, 5680-5684.
76. W. Liu, S. J. de Veer, Y.-H. Huang, T. Sengoku, C. Okada, K. Ogata, C. N. Zdenek, B. G. Fry, J. E. Swedberg, T. Passioura, D. J. Craik and H. Suga, *J. Am. Chem. Soc.*, 2021, **143**, 18481-18489.
77. T. Katoh and H. Suga, *J. Am. Chem. Soc.*, 2022, **144**, 2069-2072.
78. T. Katoh and H. Suga, *Annu. Rev. Biochem.*, 2022, **91**, 221-243.
79. G. Bhardwaj, V. K. Mulligan, C. D. Bahl, J. M. Gilmore, P. J. Harvey, O. Cheneval, G. W. Buchko, S. V. S. R. K. Pulavarti, Q. Kaas, A. Eletsky, P.-S. Huang, W. A. Johnsen, P. Greisen, Jr., G. J. Rocklin, Y. Song, T. W. Linsky, A. Watkins, S. A. Rettie, X. Xu, L. P. Carter, R. Bonneau, J. M. Olson, E. Coutsiyas, C. E. Correnti, T. Szyperski, D. J. Craik and D. Baker, *Nature*, 2016, **538**, 329-335.
80. T. N. M. Schumacher, L. M. Mayr, D. L. Minor, M. A. Milhollen, M. W. Burgess and P. S. Kim, *Science*, 1996, **271**, 1854-1857.
81. P. E. Kolkwitz, J. Mohrlüder and D. Willbold, *Biomolecules*, 2022, **12**, 157.
82. M. Nonaka, H. Mabashi-Asazuma, D. L. Jarvis, K. Yamasaki, T. O. Akama, M. Nagaoka, T. Sasai, I. Kimura-Takagi, Y. Suwa, T. Yaegashi, C.-T. Huang, C. Nishizawa-Harada and M. N. Fukuda, *PLoS One*, 2021, **16**, e0241157.
83. M. Malhis, S. Kaniyappan, I. Aillaud, R. R. Chandupatla, L. M. Ramirez, M. Zweckstetter, A. H. C. Horn, E. Mandelkow, H. Sticht and S. A. Funke, *Chembiochem*, 2021, **22**, 3049-3059.
84. X. C. Zhang, X. Y. Zhang, M. L. Zhong, P. Zhao, C. Guo, Y. Li, T. Wang and H. L. Gao, *ACS Chem. Neurosci.*, 2020, **11**, 4240-4253.

85. C. Dammers, D. Yolcu, L. Kukuk, D. Willbold, M. Pickhardt, E. Mandelkow, A. H. C. Horn, H. Sticht, M. N. Malhis, N. Will, J. Schuster and S. A. Funke, *PLoS One*, 2016, **11**, 18.
86. X. Zhou, C. Zuo, W. Li, W. Shi, X. Zhou, H. Wang, S. Chen, J. Du, G. Chen, W. Zhai, W. Zhao, Y. Wu, Y. Qi, L. Liu and Y. Gao, *Angew. Chem. Int. Ed.*, 2020, **59**, 15114-15118.
87. C. Díaz-Perlas, M. Varese, S. Guardiola, M. Sánchez-Navarro, J. García, M. Teixidó and E. Giralt, *Chembiochem*, 2019, **20**, 2079-2084.
88. Z. Li, J. Xie, S. Peng, S. Liu, Y. Wang, W. Lu, J. Shen and C. Li, *Bioconjugate Chem.*, 2017, **28**, 2167-2179.
89. S. Rudolph, A. N. Klein, M. Tusche, C. Schlosser, A. Elfgen, O. Brener, C. Teunissen, L. Gremer, S. A. Funke, J. Kutzsche and D. Willbold, *PLoS One*, 2016, **11**, e0147470.
90. H.-N. Chang, B.-Y. Liu, Y.-K. Qi, Y. Zhou, Y.-P. Chen, K.-M. Pan, W.-W. Li, X.-M. Zhou, W.-W. Ma, C.-Y. Fu, Y.-M. Qi, L. Liu and Y.-F. Gao, *Angew. Chem. Int. Ed.*, 2015, **54**, 11760-11764.
91. P. S. Marinec, K. E. Landgraf, M. Uppalapati, G. Chen, D. Xie, Q. Jiang, Y. Zhao, A. Petriello, K. Deshayes, S. B. H. Kent, D. Ault-Riche and S. S. Sidhu, *ACS Chem. Biol.*, 2021, **16**, 548-556.
92. B. D. Welch, J. N. Francis, J. S. Redman, S. Paul, M. T. Weinstock, J. D. Reeves, Y. S. Lie, F. G. Whitby, D. M. Eckert, C. P. Hill, M. J. Root and M. S. Kay, *J. Virol.*, 2010, **84**, 11235-11244.
93. M. Liu, C. Li, M. Pazgier, C. Li, Y. Mao, Y. Lv, B. Gu, G. Wei, W. Yuan, C. Zhan, W. Y. Lu and W. Lu, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 14321-14326.
94. E. Witsch, M. Sela and Y. Yarden, *Physiology (Bethesda)*, 2010, **25**, 85-101.
95. E. R. Verhaar, A. W. Woodham and H. L. Ploegh, *Semin. Immunol.*, 2020, **52**, 101425.
96. M. Uppalapati, D. J. Lee, K. Mandal, H. Li, L. P. Miranda, J. Lowitz, J. Kenney, J. J. Adams, D. Ault-Riché, S. B. Kent and S. S. Sidhu, *ACS Chem. Biol.*, 2016, **11**, 1058-1065.
97. M. Uppalapati, D. J. Lee, K. Mandal, H. Li, L. P. Miranda, J. Lowitz, J. Kenney, J. J. Adams, D. Ault-Riché, S. B. H. Kent and S. S. Sidhu, *ACS Chem. Biol.*, 2016, **11**, 1058-1065.

98. P. Sharma and J. P. Allison, *Science*, 2015, **348**, 56-61.
99. B. D. Welch, A. P. VanDemark, A. Heroux, C. P. Hill and M. S. Kay, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16828-16833.
100. J. S. Redman, J. N. Francis, R. Marquardt, D. Papac, A. L. Mueller, D. M. Eckert, B. D. Welch and M. S. Kay, *Mol Pharm*, 2018, **15**, 1169-1179.
101. C. H. Wong, K. W. Siah and A. W. Lo, *Biostatistics*, 2018, **20**, 273-286.
102. H. Engel, F. Guischard, F. Krause, J. Nandy, P. Kaas, N. Höfflin, M. Köhn, N. Kilb, K. Voigt, S. Wolf, T. Aslan, F. Baezner, S. Hahne, C. Ruckes, J. Weygant, A. Zinina, E. B. Akmeriç, E. B. Antwi, D. Dombrovskij, P. Franke, K. L. Lesch, N. Vesper, D. Weis, N. Gensch, B. Di Ventura and M. A. Öztürk, *Synth. Syst. Biotechnol.*, 2021, **6**, 402-413.
103. M. Garton, S. Nim, T. A. Stone, K. E. Wang, C. M. Deber and P. M. Kim, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 1505-1510.
104. J. E. Hernández González, R. J. Eberle, D. Willbold and M. A. Coronado, *Front. Mol. Biosci.*, 2022, **8**, 816166.
105. M. Garton, M. Sayadi and P. M. Kim, *PLoS One*, 2017, **12**, e0187524.
106. R. B. Merrifield, *J. Am. Chem. Soc.*, 1963, **85**, 2149-2154.
107. R. Behrendt, P. White and J. Offer, *J. Pept. Sci.*, 2016, **22**, 4-27.
108. E. Valeur and M. Bradley, *Chem. Soc. Rev.*, 2009, **38**, 606-631.
109. A. P. Tofteng, L. Malik, S. L. Pedersen, K. K. Sorensen and K. J. Jensen, *Chim. Oggi*, 2011, **29**, 28-31.
110. J. M. Collins, K. A. Porter, S. K. Singh and G. S. Vanier, *Org. Lett.*, 2014, **16**, 940-943.
111. M. D. Simon, P. L. Heider, A. Adamo, A. A. Vinogradov, S. K. Mong, X. Li, T. Berger, R. L. Polcarpo, C. Zhang, Y. Zou, X. Liao, A. M. Spokoiny, K. F. Jensen and B. L. Pentelute, *Chembiochem*, 2014, **15**, 713-720.
112. P. Dawson, T. Muir, I. Clark-Lewis and S. Kent, *Science*, 1994, **266**, 776-779.
113. S. B. Kent, *Chem. Soc. Rev.*, 2009, **38**, 338-351.
114. A. J. Camarero and R. A. Mitchell, *Protein & Peptide Letters*, 2005, **12**, 723-728.
115. B. J. Backes, A. A. Virgilio and J. A. Ellman, *J. Am. Chem. Soc.*, 1996, **118**, 3055-3056.
116. B. J. Backes and J. A. Ellman, *J. Org. Chem.*, 1999, **64**, 2322-2330.

117. J. B. Blanco-Canosa, B. Nardone, F. Albericio and P. E. Dawson, *J. Am. Chem. Soc.*, 2015, **137**, 7197-7209.
118. G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui and L. Liu, *Angew. Chem. Int. Ed.*, 2011, **50**, 7645-7649.
119. J.-S. Zheng, S. Tang, Y.-K. Qi, Z.-P. Wang and L. Liu, *Nat. Protoc.*, 2013, **8**, 2483-2495.
120. Y.-C. Huang, C.-C. Chen, S.-J. Li, S. Gao, J. Shi and Y.-M. Li, *Tetrahedron*, 2014, **70**, 2951-2955.
121. O. Melnyk, C. Simonneau and J. Vicogne, in *Chemical Ligation*, 2017, pp. 89-123.
122. D. Bang and S. B. Kent, *Angew. Chem. Int. Ed.*, 2004, **43**, 2534-2538.
123. J. M. Chalker, G. J. L. Bernardes, Y. A. Lin and B. G. Davis, *Chem. Asian J.*, 2009, **4**, 630-640.
124. V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoët, J.-C. M. Monbaliu and O. Melnyk, *Chem. Rev.*, 2019, **119**, 7328-7443.
125. L. Z. Yan and P. E. Dawson, *J. Am. Chem. Soc.*, 2001, **123**, 526-533.
126. R. F. Doolittle, in *Prediction of Protein Structure and the Principles of Protein Conformation*, ed. G. D. Fasman, Springer US, Boston, MA, 1989, pp. 599-623.
127. B. L. Pentelute and S. B. H. Kent, *Org. Lett.*, 2007, **9**, 687-690.
128. S. K. Maity, M. Jbara, S. Laps and A. Brik, *Angew. Chem. Int. Ed.*, 2016, **55**, 8108-8112.
129. K. Jin and X. Li, *Chem. Eur. J.*, 2018, **24**, 17397-17404.
130. C. Haase, H. Rohde and O. Seitz, *Angew. Chem. Int. Ed.*, 2008, **47**, 6807-6810.
131. J. Chen, Q. Wan, Y. Yuan, J. Zhu and S. J. Danishefsky, *Angew. Chem. Int. Ed.*, 2008, **47**, 8521-8524.
132. S. S. Kulkarni, J. Sayers, B. Premdjee and R. J. Payne, *Nat. Rev. Chem.*, 2018, **2**, 0122.
133. E. C. B. Johnson and S. B. H. Kent, *J. Am. Chem. Soc.*, 2006, **128**, 6640-6646.
134. M. Jbara, M. Seenaiiah and A. Brik, *Chem. Commun.*, 2014, **50**, 12534-12537.
135. D. Olschewski and C. F. Becker, *Mol Biosyst*, 2008, **4**, 733-740.
136. J.-S. Zheng, M. Yu, Y.-K. Qi, S. Tang, F. Shen, Z.-P. Wang, L. Xiao, L. Zhang, C.-L. Tian and L. Liu, *J. Am. Chem. Soc.*, 2014, **136**, 3695-3704.

137. J.-S. Zheng, Y. He, C. Zuo, X.-Y. Cai, S. Tang, Z. A. Wang, L.-H. Zhang, C.-L. Tian and L. Liu, *J. Am. Chem. Soc.*, 2016, **138**, 3553-3561.
138. M. M. Disotuar, M. E. Petersen, J. M. Nogueira, M. S. Kay and D. H. Chou, *Org. Biomol. Chem.*, 2019, **17**, 1703-1708.
139. J. M. Fulcher, M. E. Petersen, R. J. Giesler, Z. S. Cruz, D. M. Eckert, J. N. Francis, E. M. Kawamoto, M. T. Jacobsen and M. S. Kay, *Org. Biomol. Chem.*, 2019, **17**, 10237-10244.
140. P. W. Erickson, J. M. Fulcher, P. Spaltenstein and M. S. Kay, *Bioconjugate Chem.*, 2021, **32**, 2233-2244.
141. J. Weidmann, M. Schnölzer, P. E. Dawson and J. D. Hoheisel, *Cell Chem. Biol.*, 2019, **26**, 645-651.e643.
142. M. T. Weinstock, M. T. Jacobsen and M. S. Kay, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 11679-11684.
143. L. Raibaut, N. Ollivier and O. Melnyk, *Chem. Soc. Rev.*, 2012, **41**, 7001-7015.
144. N. Metanis, E. Keinan and P. E. Dawson, *Angew. Chem. Int. Ed. Engl.*, 2010, **49**, 7049-7053.
145. O. Firstova, V. Agouridas, V. Diemer and O. Melnyk, in *Chemical Protein Synthesis*, ed. X. Li, Springer US, New York, NY, 2022, pp. 213-239.
146. M. Cargoët, V. Diemer, L. Raibaut, E. Lissy, B. Snella, V. Agouridas and O. Melnyk, in *Peptide and Protein Engineering: From Concepts to Biotechnological Applications*, eds. O. Iranzo and A. C. Roque, Springer US, New York, NY, 2020, pp. 1-12.
147. V. Diemer, N. Ollivier, B. Leclercq, H. Drobecq, J. Vicogne, V. Agouridas and O. Melnyk, *Nat. Commun.*, 2020, **11**, 2558.
148. O. Vázquez and O. Seitz, *J. Pept. Sci.*, 2014, **20**, 78-86.
149. M. T. Jacobsen, P. W. Erickson and M. S. Kay, *Biorg. Med. Chem.*, 2017, **25**, 4946-4952.
150. C. L. Lee, H. Liu, C. T. T. Wong, H. Y. Chow and X. Li, *J. Am. Chem. Soc.*, 2016, **138**, 10477-10484.
151. M. Bacchi, M. Jullian, S. Sirigu, B. Fould, T. Huet, L. Bruyand, M. Antoine, L. Vuillard, L. Ronga, L. M. G. Chavas, O. Nosjean, G. Ferry, K. Puget and J. A. Boutin, *Protein Sci.*, 2016, **25**, 2225-2242.

152. H. P. Sørensen, H. U. Sperling-Petersen and K. K. Mortensen, *Protein Expr. Purif.*, 2003, **31**, 149-154.
153. S. Laps, F. Atamleh, G. Kamnesky, H. Sun and A. Brik, *Journal*, 2021, **12**, 870.
154. S. Yao, A. Moyer, Y. Zheng, Y. Shen, X. Meng, C. Yuan, Y. Zhao, H. Yao, D. Baker and C. Wu, *Nat. Commun.*, 2022, **13**, 1539.
155. Y. Wu, S. Fan, M. Dong, J. Li, C. Kong, J. Zhuang, X. Meng, S. Lu, Y. Zhao and C. Wu, *Chem. Sci.*, 2022, **13**, 7780-7789.
156. G. J. Pagano and R. J. Arsenault, *Expert Rev. Proteom.*, 2019, **16**, 431-441.
157. S. Bondalapati, W. Mansour, M. A. Nakasone, S. K. Maity, M. H. Glickman and A. Brik, *Chem. Eur. J.*, 2015, **21**, 7360-7364.
158. L. J. Getz, C. S. Runte, J. K. Rainey and N. A. Thomas, *J. Bacteriol.*, 2019, **201**, e00205-00219.
159. T. M. T. Jensen, C. R. O. Bartling, O. A. Karlsson, E. Åberg, L. M. Haugaard-Kedström, K. Strømgaard and P. Jemth, *ACS Chem. Biol.*, 2021, **16**, 1191-1200.
160. Y. S. Y. Hsieh, L. C. Wijeyewickrema, B. L. Wilkinson, R. N. Pike and R. J. Payne, *Angew. Chem. Int. Ed.*, 2014, **53**, 3947-3951.
161. E. J. Kim, M. R. Bond, D. C. Love and J. A. Hanover, *Crit. Rev. Biochem. Mol. Biol.*, 2014, **49**, 327-342.
162. N. Guidotti, C. C. Lechner, A. L. Bachmann and B. Fierz, *Chembiochem*, 2019, **20**, 1124-1128.
163. J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier and J. J. Ottesen, *J. Mol. Biol.*, 2011, **408**, 187-204.
164. W. J. Gui, G. A. Davidson and Z. H. Zhuang, *RSC Chem. Biol.*, 2021, **2**, 450-467.
165. Y. Zhou, Q. Xie, H. Wang and H. Sun, *J. Pept. Sci.*, 2022, **28**, e3367.
166. M. Msallam, H. Sun, R. Meledin, P. Franz and A. Brik, *Chem. Sci.*, 2020, **11**, 5526-5531.
167. H. Sun and A. Brik, *Acc. Chem. Res.*, 2019, **52**, 3361-3371.
168. D.-L. Huang, C. Montigny, Y. Zheng, V. Beswick, Y. Li, X.-X. Cao, T. Barbot, C. Jaxel, J. Liang, M. Xue, C.-L. Tian, N. Jamin and J.-S. Zheng, *Angew. Chem. Int. Ed.*, 2020, **59**, 5178-5184.

169. T. S. Chisholm, S. S. Kulkarni, K. R. Hossain, F. Cornelius, R. J. Clarke and R. J. Payne, *J. Am. Chem. Soc.*, 2020, **142**, 1090-1100.
170. E. E. Mulvihill, *Curr. Opin. Lipidol.*, 2018, **29**, 95-103.
171. L. B. Knudsen and J. Lau, *Front. Endocrinol.*, 2019, **10**, 155.
172. M. S. Hashemzadeh, M. Mohammadi, H. E. G. Ghaleh, M. Sharti, A. Choopani and A. K. Panda, *Protein Pept. Lett.*, 2021, **28**, 122-130.
173. S. Kobayashi, *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.*, 2007, **83**, 215-247.
174. H. Liu and X. Li, *Acc. Chem. Res.*, 2018, **51**, 1643-1655.
175. F. Rohrbacher, T. G. Wucherpfennig and J. W. Bode, *Top. Curr. Chem.*, 2015, **363**, 1-31.
176. J. B. Blanco-Canosa and P. E. Dawson, *Angew. Chem. Int. Ed. Engl.*, 2008, **47**, 6851-6855.
177. T. Noguchi, H. Ishiba, K. Honda, Y. Kondoh, H. Osada, H. Ohno, N. Fujii and S. Oishi, *Bioconjugate Chem.*, 2017, **28**, 609-619.
178. L. Raibaut, H. Adihou, R. Desmet, A. F. Delmas, V. Aucagne and O. Melnyk, *Chem. Sci.*, 2013, **4**, 4061-4066.
179. S. Tang, Y.-Y. Si, Z.-P. Wang, K.-R. Mei, X. Chen, J.-Y. Cheng, J.-S. Zheng and L. Liu, *Angew. Chem. Int. Ed.*, 2015, **54**, 5713-5717.
180. M. Pan, Y. He, M. Wen, F. Wu, D. Sun, S. Li, L. Zhang, Y. Li and C. Tian, *Chem. Commun.*, 2014, **50**, 5837-5839.
181. J. Li, Y. Li, Q. He, Y. Li, H. Li and L. Liu, *Org. Biomol. Chem.*, 2014, **12**, 5435-5441.
182. Q. Wan and S. J. Danishefsky, *Angew. Chem.-Int. Edit.*, 2007, **46**, 9248-9252.
183. T. Moyal, H. P. Hemantha, P. Siman, M. Refua and A. Brik, *Chem. Sci.*, 2013, **4**, 2496-2501.
184. R. E. Thompson, X. Liu, N. Alonso-García, P. J. B. Pereira, K. A. Jolliffe and R. J. Payne, *J. Am. Chem. Soc.*, 2014, **136**, 8161-8164.
185. Q.-Y. Guo, L.-H. Zhang, C. Zuo, D.-L. Huang, Z. A. Wang, J.-S. Zheng and C.-L. Tian, *Protein Cell*, 2019, **10**, 211-216.
186. M. Murakami, T. Kiuchi, M. Nishihara, K. Tezuka, R. Okamoto, M. Izumi and Y. Kajihara, *Sci. Adv.*, 2016, **2**, e1500678.
187. C. G. F. Graf, C. Schulz, M. Schmälzlein, C. Heinlein, M. Mönnich, L. Perkams, M. Püttner, I. Boos, M. Hessefort, J. N. Lombana Sanchez, M. Weyand, C.

- Steegborn, B. Breiden, K. Ross, G. Schwarzmann, K. Sandhoff and C. Unverzagt, *Angew. Chem. Int. Ed.*, 2017, **56**, 5252-5257.
188. M. Murakami, R. Okamoto, M. Izumi and Y. Kajihara, *Angew. Chem. Int. Ed.*, 2012, **51**, 3567-3572.
 189. M. Haj-Yahya, P. Gopinath, K. Rajasekhar, H. Mirbaha, M. I. Diamond and H. A. Lashuel, *Angew. Chem. Int. Ed.*, 2020, **59**, 4059-4067.
 190. D. Garenne, M. C. Haines, E. F. Romantseva, P. Freemont, E. A. Strychalski and V. Noireaux, *Nat. Rev. Methods Primers*, 2021, **1**, 49.
 191. Y. Goto, T. Katoh and H. Suga, *Nat. Protoc.*, 2011, **6**, 779-790.
 192. D. N. Woolfson, *J. Mol. Biol.*, 2021, **433**, 167160.
 193. V. Kaškonienė, M. Stankevičius, K. Bimbraitė-Survilienė, G. Naujokaitytė, L. Šernienė, K. Mulkytė, M. Malakauskas and A. Maruška, *Appl. Microbiol. Biotechnol.*, 2017, **101**, 1323-1335.
 194. D. K. Verma, M. Thakur, S. Singh, S. Tripathy, A. K. Gupta, D. Baranwal, A. R. Patel, N. Shah, G. L. Utama, A. K. Niamah, M. L. Chávez-González, C. F. Gallegos, C. N. Aguilar and P. P. Srivastav, *Food Biosci.*, 2022, **46**, 101594.
 195. P. C. Fagundes, F. M. Farias, O. C. Santos, N. E. de Oliveira, J. A. da Paz, H. Ceotto-Vigoder, D. S. Alviano, M. T. Romanos and M. C. F. Bastos, *J. Appl. Microbiol.*, 2016, **121**, 435-444.
 196. R. H. Perez, T. Zendo and K. Sonomoto, *Front. Microbiol.*, 2018, **9**, 2085.
 197. D. J. A. Netz, R. Pohl, A. G. Beck-Sickinger, T. Selmer, A. J. Pierik, M. C. F. Bastos and H.-G. Sahl, *J. Mol. Biol.*, 2002, **319**, 745-756.
 198. K. Fujita, S. Ichimasa, T. Zendo, S. Koga, F. Yoneyama, J. Nakayama and K. Sonomoto, *Appl. Environ. Microbiol.*, 2007, **73**, 2871-2877.
 199. D. J. A. Netz, M. C. F. Bastos and H.-G. Sahl, *Appl. Environ. Microbiol.*, 2002, **68**, 5274-5280.
 200. J. Z. Acedo, M. J. van Belkum, C. T. Lohans, K. M. Towle, M. Miskolzie and J. C. Vederas, *Biochemistry*, 2016, **55**, 733-742.
 201. K. Hammond, H. Lewis, S. Halliwell, F. Desriac, B. Nardone, J. Ravi, B. W. Hoogenboom, M. Upton, J. P. Derrick and M. G. Ryadnov, *iScience*, 2020, **23**, 101423.
 202. G. G. Zhanel, H. J. Adam, M. R. Baxter, J. Fuller, K. A. Nichol, A. J. Denisuik, A. R. Golden, R. Hink, P. R. S. Lagacé-Wiens, A. Walkty, M. R. Mulvey, F.

- Schweizer, D. Bay, D. J. Hoban and J. A. Karlowsky, *J. Antimicrob. Chemother.*, 2019, **74**, iv5-iv21.
203. S. T. Henriques, H. Peacock, A. H. Benfield, C. K. Wang and D. J. Craik, *J. Am. Chem. Soc.*, 2019, **141**, 20460-20469.
 204. K. Tam, V. J. Torres, V. A. Fischetti, R. P. Novick, J. J. Ferretti, D. A. Portnoy, M. Braunstein and J. I. Rood, *Microbiol. Spectr.*, 2019, **7**, 7.2.16.
 205. M. Sieprawska-Lupa, P. Mydel, K. Krawczyk, K. Wójcik, M. Puklo, B. Lupa, P. Suder, J. Silberring, M. Reed, J. Pohl, W. Shafer, F. McAleese, T. Foster, J. Travis and J. Potempa, *Antimicrob. Agents Chemother.*, 2004, **48**, 4673-4679.
 206. L. Lo Conte, C. Chothia and J. Janin, *J. Mol. Biol.*, 1999, **285**, 2177-2198.
 207. E. Krissinel and K. Henrick, *J. Mol. Biol.*, 2007, **372**, 774-797.
 208. N. Shagaghi, E. A. Palombo, A. H. A. Clayton and M. Bhawe, *World J. Microbiol. Biotechnol.*, 2016, **32**, 31.
 209. A. Chaudhuri, S. Haldar, H. Sun, R. E. Koeppe, 2nd and A. Chattopadhyay, *Biochim. Biophys. Acta*, 2014, **1838**, 419-428.
 210. H. N. Hunter, W. Jing, D. J. Schibli, T. Trinh, I. Y. Park, S. C. Kim and H. J. Vogel, *Biochim. Biophys. Acta Biomembr.*, 2005, **1668**, 175-189.
 211. X. Bi, C. Wang, L. Ma, Y. Sun and D. Shang, *J. Appl. Microbiol.*, 2013, **115**, 663-672.
 212. W. M. Yau, W. C. Wimley, K. Gawrisch and S. H. White, *Biochemistry*, 1998, **37**, 14713-14718.
 213. S. Iwatani, T. Zendo, F. Yoneyama, J. Nakayama and K. Sonomoto, *Biosci. Biotechnol. Biochem.*, 2007, **71**, 1984-1992.
 214. S. Sandiford and M. Upton, *Antimicrob. Agents Chemother.*, 2012, **56**, 1539-1547.
 215. E. F. Haney, A. P. Petersen, C. K. Lau, W. Jing, D. G. Storey and H. J. Vogel, *Biochim. Biophys. Acta Biomembr.*, 2013, **1828**, 1802-1813.
 216. M. Arias, E. R. Hoffarth, H. Ishida, J. M. Aramini and H. J. Vogel, *Biochim. Biophys. Acta*, 2016, **1858**, 1012-1023.
 217. G. P. Doss, G. Agoramoorthy and C. Chakraborty, *Front. Biosci. (Landmark Ed.)*, 2014, **19**, 1028-1040.
 218. R. Fischer, R. E. Kontermann and K. Pfizenmaier, *Front. Cell. Dev. Biol.*, 2020, **8**, 401-401.

219. Y. Palacios, A. Ruiz, L. A. Ramón-Luing, R. Ocaña-Guzman, O. Barreto-Rodriguez, A. Sánchez-Monciváis, B. Tecuatzi-Cadena, A. G. Regalado-García, R. D. Pineda-Gudiño, A. García-Martínez, F. Juárez-Hernández, J. P. Farias-Contreras, I. Fricke-Galindo, G. Pérez-Rubio, R. Falfán-Valencia, I. Buendia-Roldan, K. Medina-Quero and L. Chavez-Galan, *Int. J. Mol. Sci.*, 2021, **22**, 8423.
220. E. R. Bowman, C. M. A. Cameron, A. Avery, J. Gabriel, A. Kettelhut, M. Hecker, C. U. Sontich, B. Tamilselvan, C. N. Nichols, B. Richardson, M. Cartwright, N. T. Funderburg and M. J. Cameron, *J. Infect. Dis.*, 2021, **223**, 805-810.
221. S. Steeland, L. Puimège, R. E. Vandenbroucke, F. Van Hauwermeiren, J. Haustraete, N. Devoogdt, P. Hulpiau, G. Leroux-Roels, D. Laukens, P. Meuleman, M. De Vos and C. Libert, *J. Biol. Chem.*, 2015, **290**, 4022-4037.
222. D. W. Banner, A. D'Arcy, W. Janes, R. Gentz, H. J. Schoenfeld, C. Broger, H. Loetscher and W. Lesslauer, *Cell*, 1993, **73**, 431-445.
223. D. McMillan, C. Martinez-Fleites, J. Porter, D. Fox, 3rd, R. Davis, P. Mori, T. Ceska, B. Carrington, A. Lawson, T. Bourne and J. O'Connell, *Nat Commun*, 2021, **12**, 582.
224. A. Norman, C. Franck, M. Christie, P. M. E. Hawkins, K. Patel, A. S. Ashhurst, A. Aggarwal, J. K. K. Low, R. Siddiquee, C. L. Ashley, M. Steain, J. A. Triccas, S. Turville, J. P. Mackay, T. Passioura and R. J. Payne, *ACS Cent. Sci.*, 2021, **7**, 1001-1008.
225. H. Nakamoto and J. C. Bardwell, *Biochim. Biophys. Acta Mol. Cell Res.*, 2004, **1694**, 111-119.
226. M. S. Zambrano-Mila, K. E. S. Blacio and N. S. Vispo, *Ther. Innov. Regul. Sci.*, 2020, **54**, 308-317.
227. B. Kamber, A. Hartmann, K. Eisler, B. Riniker, H. Rink, P. Sieber and W. Rittel, *Helv. Chim. Acta*, 1980, **63**, 899-915.
228. X. Pan and T. Kortemme, *J. Biol. Chem.*, 2021, **296**, 100558.
229. D. N. Woolfson, *J. Mol. Biol.*, 2021, **433**, 167160.
230. A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. W. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y.-E. A. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popović, J. J. Havranek, J.

- Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, in *Methods Enzymol.*, eds. M. L. Johnson and L. Brand, Academic Press, 2011, vol. 487, pp. 545-574.
231. G. J. Rocklin, T. M. Chidyausiku, I. Goreshnik, A. Ford, S. Houliston, A. Lemak, L. Carter, R. Ravichandran, V. K. Mulligan, A. Chevalier, C. H. Arrowsmith and D. Baker, *Science*, 2017, **357**, 168-175.
 232. L. Cao, I. Goreshnik, B. Coventry, J. B. Case, L. Miller, L. Kozodoy, R. E. Chen, L. Carter, A. C. Walls, Y.-J. Park, E.-M. Strauch, L. Stewart, M. S. Diamond, D. Veessler and D. Baker, *Science*, 2020, **370**, 426-431.
 233. L. Cao, B. Coventry, I. Goreshnik, B. Huang, W. Sheffler, J. S. Park, K. M. Jude, I. Marković, R. U. Kadam, K. H. G. Verschueren, K. Verstraete, S. T. R. Walsh, N. Bennett, A. Phal, A. Yang, L. Kozodoy, M. DeWitt, L. Picton, L. Miller, E.-M. Strauch, N. D. DeBouver, A. Pires, A. K. Bera, S. Halabiya, B. Hammerson, W. Yang, S. Bernard, L. Stewart, I. A. Wilson, H. Ruohola-Baker, J. Schlessinger, S. Lee, S. N. Savvides, K. C. Garcia and D. Baker, *Nature*, 2022, **605**, 551-560.
 234. J. K. Leman, B. D. Weitzner, S. M. Lewis, J. Adolf-Bryfogle, N. Alam, R. F. Alford, M. Aprahamian, D. Baker, K. A. Barlow, P. Barth, B. Basanta, B. J. Bender, K. Blacklock, J. Bonet, S. E. Boyken, P. Bradley, C. Bystroff, P. Conway, S. Cooper, B. E. Correia, B. Coventry, R. Das, R. M. De Jong, F. DiMaio, L. Dsilva, R. Dunbrack, A. S. Ford, B. Frenz, D. Y. Fu, C. Geniesse, L. Goldschmidt, R. Gowthaman, J. J. Gray, D. Gront, S. Guffy, S. Horowitz, P.-S. Huang, T. Huber, T. M. Jacobs, J. R. Jeliazkov, D. K. Johnson, K. Kappel, J. Karanicolas, H. Khakzad, K. R. Khar, S. D. Khare, F. Khatib, A. Khramushin, I. C. King, R. Kleffner, B. Koepnick, T. Kortemme, G. Kuenze, B. Kuhlman, D. Kuroda, J. W. Labonte, J. K. Lai, G. Lapidoth, A. Leaver-Fay, S. Lindert, T. Linsky, N. London, J. H. Lubin, S. Lyskov, J. Maguire, L. Malmström, E. Marcos, O. Marcu, N. A. Marze, J. Meiler, R. Moretti, V. K. Mulligan, S. Nerli, C. Norn, S. Ó'Conchúir, N. Ollikainen, S. Ovchinnikov, M. S. Pacella, X. Pan, H. Park, R. E. Pavlovicz, M. Pethe, B. G. Pierce, K. B. Pilla, B. Raveh, P. D. Renfrew, S. S. R. Burman, A. Rubenstein, M. F. Sauer, A. Scheck, W. Schief, O. Schueler-Furman, Y. Sedan, A. M. Sevy, N. G. Sgourakis, L. Shi, J. B. Siegel, D.-A. Silva, S. Smith, Y. Song, A. Stein, M. Szegedy, F. D. Teets, S. B. Thyme,

- R. Y.-R. Wang, A. Watkins, L. Zimmerman and R. Bonneau, *Nat. Methods*, 2020, **17**, 665-680.
235. E. Oancea, M. N. Teruel, A. F. G. Quest and T. Meyer, *J. Cell Biol.*, 1998, **140**, 485-498.
236. H. Zhou, S. Ekmekcioglu, J. W. Marks, K. A. Mohamedali, K. Asrani, K. K. Phillips, S. A. Brown, E. Cheng, M. B. Weiss, W. N. Hittelman, N. L. Tran, H. Yagita, J. A. Winkles and M. G. Rosenblum, *J. Invest. Dermatol.*, 2013, **133**, 1052-1062.
237. M. Remer, A. White, M. Glennie, A. Al-Shamkhani and P. Johnson, *Curr. Top. Microbiol. Immunol.*, 2017, **405**, 165-207.
238. J. S. Park, S. M. Kim, K. A. Jung, J. Lee, S. K. Kwok, M. L. Cho and S. H. Park, *Histol. Histopathol.*, 2017, **32**, 481-490.
239. B. Zhang, T. Wu, M. Chen, Y. Zhou, D. Yi and R. Guo, *Immunol. Lett.*, 2013, **153**, 58-61.
240. G. Gradisteanu Pircalabioru, L. I. Popa, L. Marutescu, I. Gheorghe, M. Popa, I. Czobor Barbu, R. Cristescu and M.-C. Chifiriuc, *Pharmaceutics*, 2021, **13**, 196.
241. E. Watkins-Dulaney, S. Straathof and F. Arnold, *Chembiochem*, 2021, **22**, 5-16.
242. E. N. Marsh and Y. Suzuki, *ACS Chem. Biol.*, 2014, **9**, 1242-1250.
243. L. Mäler, *Mol. Membr. Biol.*, 2012, **29**, 155-176.
244. B. Bechinger and E. S. Salnikov, *Chem. Phys. Lipids*, 2012, **165**, 282-301.
245. H. Wajant and P. Scheurich, *FEBS J.*, 2011, **278**, 862-876.
246. J. Dou, A. A. Vorobieva, W. Sheffler, L. A. Doyle, H. Park, M. J. Bick, B. Mao, G. W. Foight, M. Y. Lee and L. A. Gagnon, *Nature*, 2018, **561**, 485-491.
247. A. Chevalier, D.-A. Silva, G. J. Rocklin, D. R. Hicks, R. Vergara, P. Murapa, S. M. Bernard, L. Zhang, K.-H. Lam and G. Yao, *Nature*, 2017, **550**, 74-79.
248. N. J. Anthis and G. M. Clore, *Protein Sci.*, 2013, **22**, 851-858.
249. P. R. Evans and G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr*, 2013, **69**, 1204-1214.
250. A. A. Lebedev and M. N. Isupov, *Acta crystallographica. Section D, Biological crystallography*, 2014, **70**, 2430-2443.
251. A. A. Lebedev, A. A. Vagin and G. N. Murshudov, *Acta Crystallogr D Biol Crystallogr*, 2008, **64**, 33-39.

- 252. A. Perrakis, M. Harkiolaki, K. S. Wilson and V. S. Lamzin, *Acta Crystallogr D Biol Crystallogr*, 2001, **57**, 1445-1450.
- 253. P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta crystallographica. Section D, Biological crystallography*, 2010, **66**, 486-501.
- 254. G. N. Murshudov, P. Skubák, A. A. Lebedev, N. S. Pannu, R. A. Steiner, R. A. Nicholls, M. D. Winn, F. Long and A. A. Vagin, *Acta Crystallogr D Biol Crystallogr*, 2011, **67**, 355-367.
- 255. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235-242.
- 256. H. T. Jenkins, *Acta Crystallogr D Struct Biol*, 2018, **74**, 205-214.
- 257. K. Cowtan, *Acta Crystallographica Section D*, 2006, **62**, 1002-1011.
- 258. J. H. Naismith, T. Q. Devine, T. Kohno and S. R. Sprang, *Structure*, 1996, **4**, 1251-1262.
- 259. I. Wiegand, K. Hilpert and R. E. Hancock, *Nat. Protoc.*, 2008, **3**, 163-175.

Chapter 7 : Appendices

7.1. Supplementary figures – Chapter 2

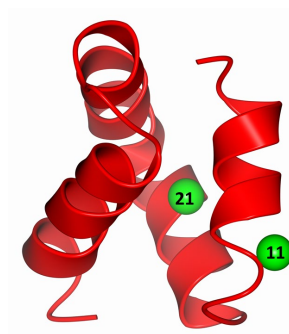


Figure S7.1: Positions of AucA Ala11 and Ala21 in the folded state, shown as green van der Waals (VDW) spheres.

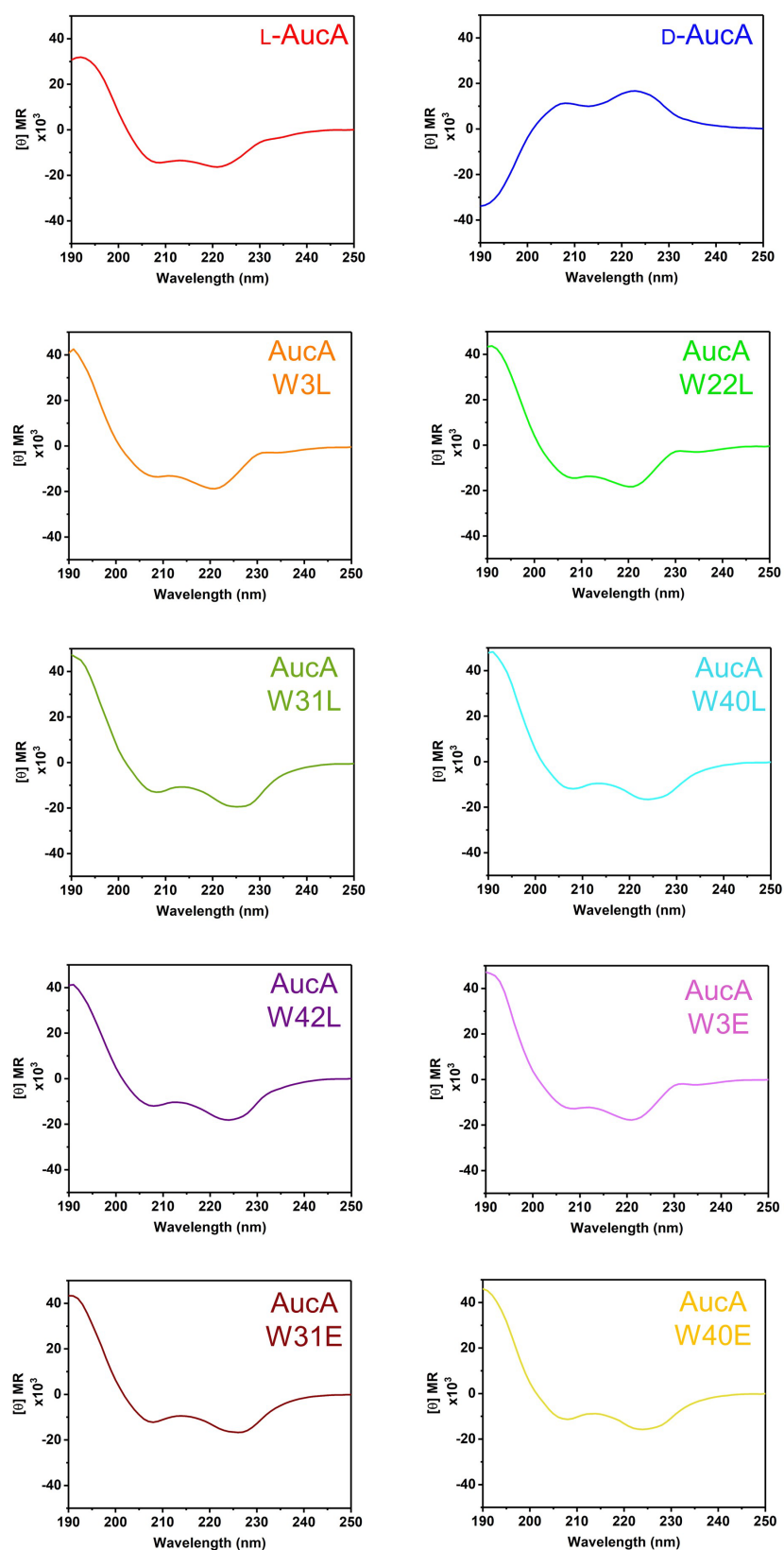


Figure S7.2: Circular dichroism spectra of AucA and variants.

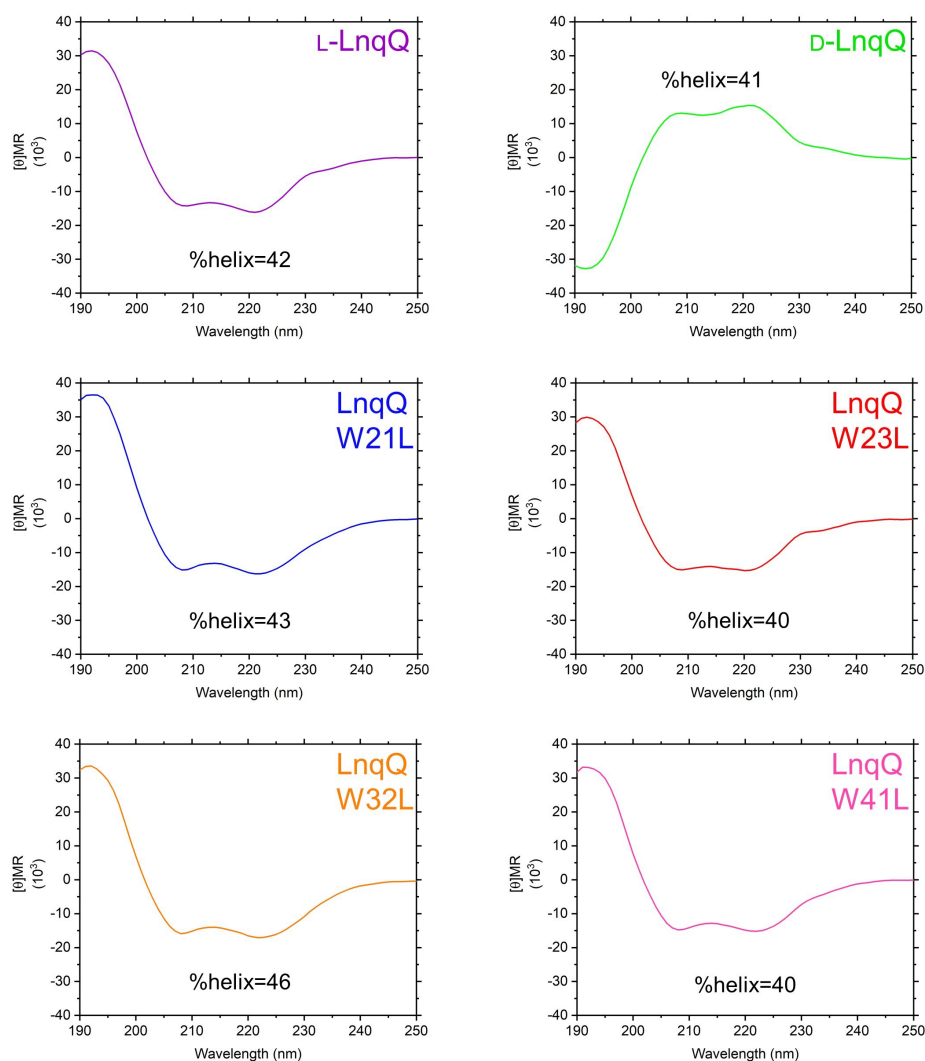


Figure S7.3: Circular dichroism spectra of LnqQ and variants.

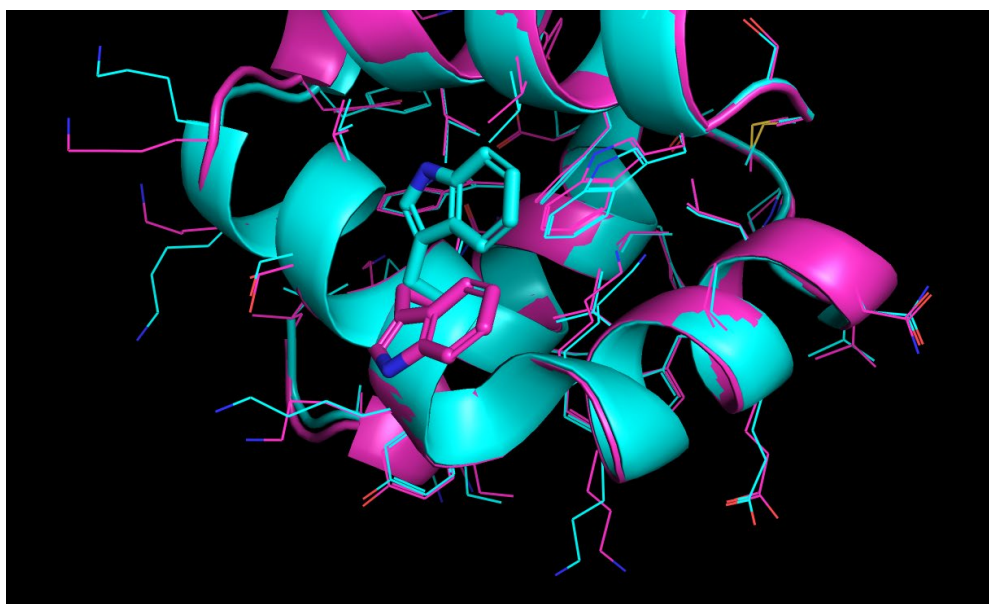


Figure S7.4: Reorientation of AucA W22 between dimeric and monomeric forms. Overlay of AucA (SO₄²⁻) (PDB: 8AVR, Cyan) and AucA (dimer) (PDB: 8AVU, Magenta) chain A showing reorientation of Trp22 (Bold).

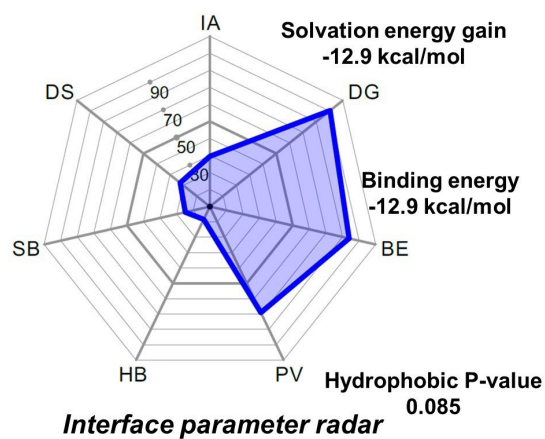


Figure S7.5: Analysis of AucA dimeric interface (PDB: 8AVU) using Protein Interfaces, Surfaces and Assemblies (PISA) software²⁰⁷ in the CCP4 software suite.

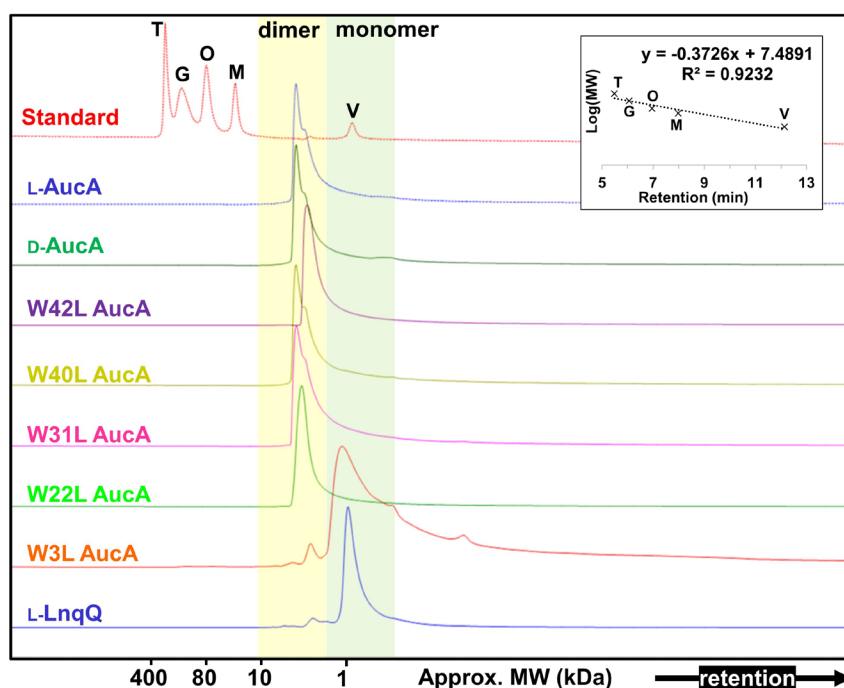


Figure S7.6: Analytical HP-SEC of AucA and Trp→Leu variants. Gel filtration standard shown (top) was used to create calibration plot (top right), with approximate calculated MW shown along the X-axis. LnqQ is used as monomeric bacteriocin reference.

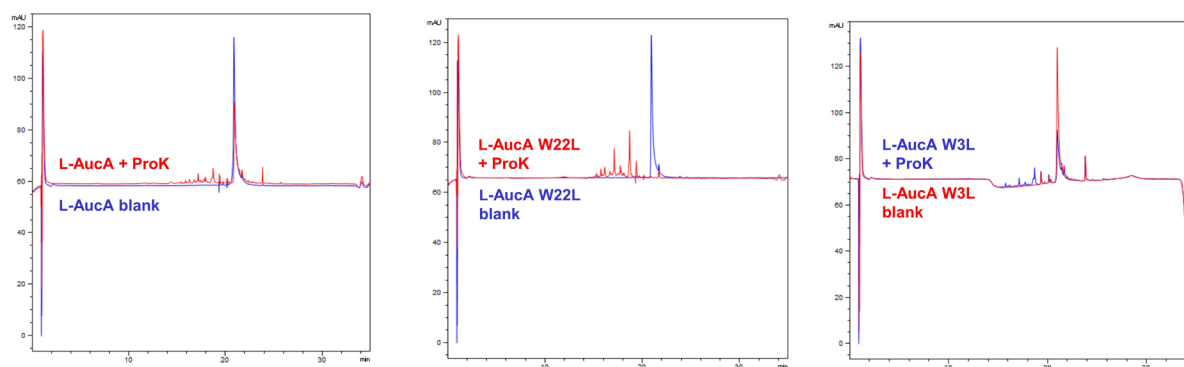



Figure S7.7: AucA proteolytic stability assay. LC traces at UV210nm of protease reaction quenched after 24 hours incubation with proteinase K (+ProK) at 37 °C, and samples prepared and incubated in the same manner without addition of protease (blank). HPLC conducted using a 0-99% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm).

Table S7.1: Multiple sequence alignment (MSA) of homologous bacteriocins ($\geq 34\%$ homology to AucA) aligned using Clustal Omega and constructed in Jalview, showing conservation of tryptophan (red) and lysine (blue) residues, along with other potential H-bond donor residues arginine (green) and glutamine (yellow).

Name	Species	#res.	AucA identity (%)	Isolated	10	20	30	40	50	Accession		
Aureocin A53	<i>Staphylococcus aureus</i>	51	100.0	Y	-MSWLNFL	KYIAKYGKKAVSAAWKY	KGKVL	EWNLVGPTLEWV	WQKLKKIAGL	WP 032072954.1		
Epidermicin N101	<i>Staphylococcus epidermidis</i>	51	38.3	Y	MAAFMKLI	QFLATKGQKYVSLAWKH	KGTL	IKWINAGQSFEW	IYKQIKKLWA	6SIF A		
Lacticin Z	<i>Lactococcus lactis</i>	53	46.0	Y	MAGFLKVVQ	LAKYGSKAVQWAWANK	GGIL	LDWINAGQAI	DWVVEKIKQILGIR	BAF75975.1		
Lacticin Q	<i>Lactococcus lactis</i>	53	48.0	Y	MAGFLKVVQ	LAKYGSKAVQWAWAN	KGKIL	DWLNAGQAI	DWVVSQIKQILGIR	WP 058206662.1		
	<i>Staphylococcus argenteus</i>	37	97.3	N	-----	-----	KKAI	SAAWKY	KGKVL	EWNLVGPTLEWV	WQKLKKIAGL	MCG9844006.1
	<i>Bacillales bacterium</i>	48	54.7	N	---MATFL	KIVALQGT	KAAKAWAN	KGKVL	EWRDGLAIDWI	DKINDIVG	MBE3569898.1	
	<i>Lactiplantibacillus plantarum</i>	50	47.8	N	MGTFLKLV	KWAATYGR	KAVSAVWKH	KGQIL	KWINGFQSLDWI	KNKIKKWF	WP 157113116.1	
	<i>Corynebacterium jeikeium</i>	52	46.9	N	MAGFLKVVQ	KAVAKYGS	KAVKWCWDN	KGKIL	EWLNIGMAVDWI	VEQVRKIVGA	WP 010976360.1	
	<i>Bacillus cereus</i>	48	46.7	N	---MLAFL	KLVAKLGP	KAAKAWAN	KGKVL	GWIRDGLAIDWI	INKINDIVN	WP 063539056.1	
	<i>Bacillus anthracis</i>	48	44.4	N	---MLAFL	KLVAKLGP	KAAKAWAN	KSRL	GWIRDGLAIDWI	INKINDIVN	WP 078984773.1	
	<i>Oceanobacillus halophilus</i>	48	42.2	N	---MLAFL	RLVGQLGS	KAAKAWDN	KGRV	LEWLRDGMFSFWI	VDKIEDIVN	WP 121205640.1	
	<i>Bacillus mycoides</i>	48	41.9	N	---MVAFL	RLVGQLGA	KAAKAWAN	KGKVL	GWIRDGMAIEWI	INKINDMVS	WP 1215571576.1	
	<i>Fronthabibans sp.</i>	48	41.7	N	---MRIL	GLLGKYGK	KAVDWAN	KGRIL	NWLNAGQAI	DWVVAQVRKAVGV	WP 228513925.1	
	<i>Clostridium cibarium</i>	49	41.3	N	---MGAVI	KAVAKYGS	KAIKVVWAN	KATVL	KWLDRGMTVAWI	ANEIRKALGL	WP 191770381.1	
	<i>Fronthabibans sp.</i>	52	41.2	N	MVAFLRLI	GLLGKYGK	KAVDWAN	KGRIL	NWLNAGQAI	DWVVAQVRKAVGV	MBF4574738.1	
	<i>Arcanobacterium phocae</i>	52	39.6	N	MAVFFRLI	QLIVARYGR	SASVSWVAH	KGQIL	DWINAGQAVEWI	VQKVRSAVGT	WP 216389689.1	
	<i>Bacillus mycoides</i>	48	39.5	N	---MVAFL	RLVGQLGA	KAAKAWAN	KGKVL	DWIRDGMAIEWI	INKINDMVS	WP 1215585204.1	
	<i>Bacillus</i>	48	39.5	N	---MVAFL	RLVGQLGA	KAAKAWAN	KGKVL	DWIRDGMAIEWI	INKINDMVN	WP 063539054.1	
	<i>Bacillus cereus</i>	48	39.5	N	---MVAFL	RLVGQLGA	KAAKAWAN	KGKVL	DWIRDGMAIEWI	INKINDMVN	WP 063539056.1	
	<i>Curtobacterium</i>	53	39.1	N	MAVFARIL	QLLAKYGAR	AVNNAKAN	IQRVL	NWLNAGQAI	DWIVSKIKQILGIR	WP 123311236.1	
	<i>Staphylococcus sp. TE8</i>	51	39.1	N	MAGFMKLI	QFLATKGQKYVSLAWKH	KGTL	IKWINAGQSFEW	IYKQIKKLWS	WP 037551471.1		
	<i>Propionibacterium sp.</i>	53	38.8	N	MTIFLRIL	QLIAKYGKRA	IDWCWAN	KDRIL	NWLRNGMAIDWI	INKIKEILGIR	NMD46367.1	
	<i>Bacillus paramycoides</i>	48	38.3	N	---MMAFL	KLVGKLGP	KAAKAWAN	KGKVM	DWIAQGMADWI	IDQINRIVG	WP 144572616.1	
	<i>Bacillus paramycoides</i>	48	38.3	N	---MMAFL	KLVGKLGP	KAAKAWAN	KGKVM	DWITQGMADWI	IDQINRIVG	WP 144572613.1	
	<i>Bacillus toyonensis</i>	48	38.3	N	---MLSFA	KLVARLSASK	AKAWANN	KGKV	VEWIKNGATFEWI	SNKIDQMIG	WP 220553589.1	
	<i>Bacillus wiedmannii</i>	48	38.3	N	---MLSFA	KLVARLSASK	AKAWANN	KGKV	VEWIKNGATFEWI	SNKIDQMMG	WP 098048139.1	
	<i>Curtobacterium Herbarum</i>	53	36.2	N	MAVFARIL	QLLAKYGAR	AVAWAKAHV	QQVL	NWINIGQAI	DWIVSKIKQILGIR	WP 121851038.1	
	<i>Curtobacterium sp. Csp2</i>	53	36.2	N	MAVFPRIL	QLLAKYGAR	AVEWAKAHV	QQIL	NWINAGQAI	DWIVAKIKQILGIR	WP 174779547.1	
	<i>Oceanobacillus halophilus</i>	48	36.2	N	---MVTFL	RLVAQLGS	KAAKAWDN	KGRV	LDWIRNGMAFDWI	IDKINSIVN	WP 121205641.1	
	<i>Scardocia wiggisiae</i>	53	36.0	N	MGAFFRL	LLSILARYGAR	AVQWAWSH	RGTVL	RWLGAQAI	DWIVIKQIKRLLGIR	MBF1666445.1	
	<i>Clostridium sp.</i>	49	34.8	N	---MGAFL	KAVAKYGS	KAVYVWAH	KSTIM	KWIDRGMSAADI	QKIRQLILGM	WP 251862045.1	
	<i>Scardocia wiggisiae</i>	53	34.0	N	MGAFFRL	LLSILARYGAR	AVQWAWAH	RGTVL	RWIGAGQAI	DWIVIKQIKRLLGIR	WP 017147590.1	
	<i>Curtobacterium sp. WW7</i>	53	34.0	N	MAVFARIL	QLLAKYGAR	AVAWAKAHV	QQVL	NWINIGQAI	DWIVSKIKQVGLIR	WP 166781659.1	
												
					MAV+LAFLKLVAKYGA							

7.2. Supplementary figures – Chapter 3

Table S7.2: Conditions used during screening of TNRC2-2 refolding. First screening oxidation method (GSSG/GSH best), then buffer pH and temperature (0.1 M NaPi, pH 6.5, 25 °C best), then finally reagent concentrations (all worse than previous). The conditions with the highest yield of folded TNRC2 are highlighted.

Component	Oxidation method				Buffer pH and temperature				Concentrations				
Denaturation	i	ii	iii	iv	i	ii	iii	iv	i	ii	iii	iv	v
GnHCl	6M	6M	6M	6M	6M	6M	6M	6M	6M	6M	6M	6M	6M
Oxidant	6mM GSSG	10% DMSO	atmos. O ₂	6mM GSSG	6mM GSSG	6mM GSSG	6mM GSSG	6mM GSSG	6mM GSSG	25mM GSSG	6mM GSSG	25mM GSSG	6mM GSSG
Reductant	60mM GSH	-	-	60mM GSH	60mM GSH	60mM GSH	60mM GSH	60mM GSH	120mM GSH	50mM GSH	120mM GSH	50mM GSH	120mM GSH
Buffer	-	-	-	0.1M Tris	0.1M Tris	0.1M NaPi	0.1M Tris	0.1M NaPi	0.1M NaPi	0.1M NaPi	0.1M NaPi	0.1M NaPi	0.1M NaPi
Dilute x5 with	0.25M NaHCO ₃	0.25M NaHCO ₃	0.25M NaHCO ₃	0.1M Tris	0.1M Tris	0.1M NaPi	0.1M Tris	0.1M NaPi	0.1M NaPi	0.1M NaPi	0.1M NaPi	0.1M NaPi	0.1M NaPi
pH	8.5	8.5	8.5	8.5	8.5	6.5	8.5	6.5	6.5	6.5	6.5	6.5	6.5
TNRC2 final conc. (mg/mL)	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.4	0.4
T (°C)	25	25	25	25	25	25	4	4	25	25	25	25	25
refolding % 96 hrs	7%	<2%	0	9%	17%	33%	0%	3%	15%	21%	7%	5%	13%

Table S7.3: Results of bacteriophage biopanning experiment 1 with the sequence format **CX₉C**. Enrichment factor for each round is defined as the amount of phage eluted from the target well containing D-TNRCD2 divided by the amount of phage eluted from the target well without immobilized target (streptavidin/neutravidin only).

Screening round number	Amount of immobilized D-TNRCD2 (μg)	Amount of phage input (pfu)	Amount of phage in elution from target well (pfu)		Enrichment factor
1	3	4.96 x 10 ¹²	Test	4.18 x 10 ⁶	58
			Control	7.20 x 10 ⁴	
2	3	2.85 x 10 ¹³	Test	1.08 x 10 ⁷	5.36
			Control	2.02 x 10 ⁶	
3	3	7.00 x 10 ¹²	Test	9.43 x 10 ⁷	152.32
			Test	6.19 x 10 ⁵	

Table S7.4: Next generation sequencing of enriched **CX₉C** bacteriophage library following screening experiment 1. A total of 20282 sequences were identified, with sequences in abundance >40 listed below, in decreasing order.

Amino acid sequence											Abundance
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	
C	F	H	C	V	W	L	G	M	E	C	7590
C	W	H	V	A	W	L	G	E	G	C	630
C	I	T	V	L	P	G	I	V	V	C	556
C	Q	E	K	R	G	T	P	E	E	C	366
C	W	W	R	E	D	Q	Y	Q	Q	C	166
C	E	C	V	D	V	E	P	F	F	C	130
C	V	E	S	L	L	P	P	W	W	C	130
C	L	R	A	V	A	W	W	D	D	C	114
C	M	L	P	A	P	L	E	L	L	C	100
C	W	E	N	W	E	D	T	W	W	C	94
C	Y	A	D	Y	R	E	L	K	K	C	80
C	L	E	V	R	S	R	R	D	D	C	72
C	R	L	E	Y	L	P	F	V	V	C	64
C	V	V	V	V	P	G	R	I	I	C	54
C	R	L	K	L	E	L	A	S	S	C	52
C	N	V	L	R	E	G	R	S	S	C	52
C	W	A	R	E	S	P	R	K	K	C	50
C	I	L	P	G	P	L	E	L	L	C	50
C	E	L	F	G	P	L	W	S	S	C	48
C	P	P	A	P	E	R	T	P	P	C	46
C	V	E	P	W	E	G	K	I	I	C	42
C	F	H	V	V	W	L	G	G	E	C	40
C	W	E	V	I	W	R	E	G	W	C	40

Table S7.5: Results of bacteriophage biopanning experiment 2 with the sequence format **CX₄WLGX₂C**. Enrichment factor for each round is defined as the amount of phage eluted from the target well containing D-TNRCD2 divided by the amount of phage eluted from the target well without immobilized target (streptavidin/neutravidin only).

Screening round number	Amount of immobilized D-TNRCD2 (μg)	Amount of phage input (pfu)	Amount of phage in elution from target well (pfu)		Enrichment factor
1	3	2.00×10^{13}	Test	7.20×10^6	10
			Control	7.20×10^5	
2	3	7.50×10^{12}	Test	1.08×10^7	5.36
			Control	2.02×10^6	
3	3	2.75×10^{13}	Test	2.45×10^{10}	8292.68
			Test	2.95×10^6	

Table S7.6: Sequencing of enriched **CX₄WLGX₂C** bacteriophage library following screening experiment 2. A total of 22 monoclones were randomly selected, with sequences listed below, in decreasing order of abundance.

Amino acid sequence											Abundance
P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	
C	F	H	C	I	W	L	G	D	E	C	4
C	Y	H	C	V	W	L	G	H	E	C	3
C	Y	H	I	I	W	L	G	D	E	C	2
C	F	H	C	I	W	L	G	P	E	C	2
C	Y	H	I	V	W	L	G	N	E	C	1
C	Y	H	C	V	W	L	G	T	E	C	1
C	Y	H	C	I	W	L	G	F	E	C	1
C	F	H	C	I	W	L	G	L	E	C	1
C	F	H	C	I	W	L	G	T	E	C	1
C	Y	E	V	I	W	L	G	H	E	C	1
C	F	H	C	I	W	L	G	N	E	C	1
C	F	H	C	I	W	L	G	S	E	C	1
C	F	H	I	I	W	L	G	N	E	C	1
C	F	H	C	I	W	L	G	F	E	C	1
C	M	W	C	E	W	L	G	E	D	C	1

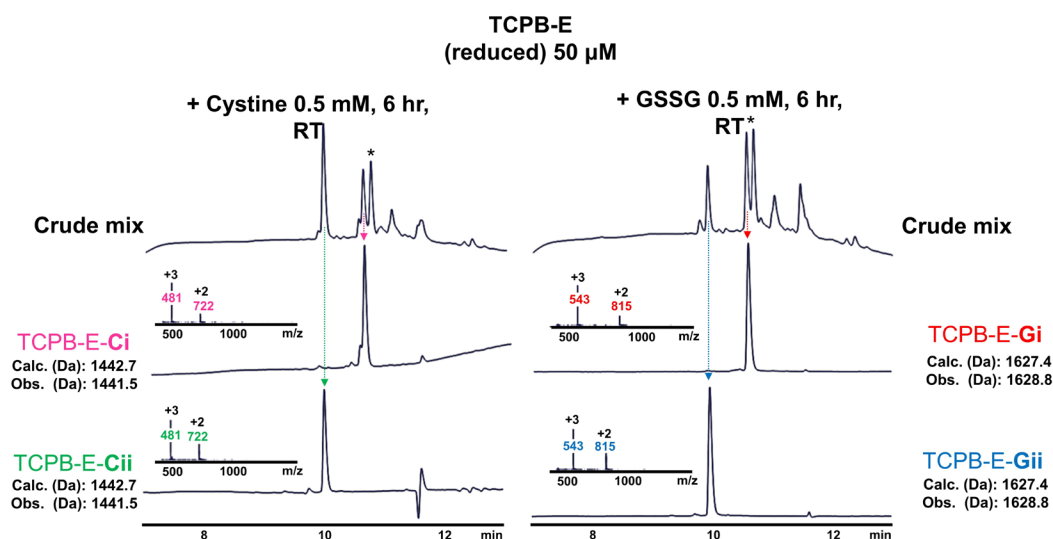


Figure S7.8: Solution-phase oxidation of TCPB-E with cystine (left) or glutathione disulfide (right) for deduction of active bacteriophage conformation. See below for identification of peaks by disulfide bond mapping experiments. Peak corresponding to dimerized product with three disulfide bonds indicate by asterisk (*).

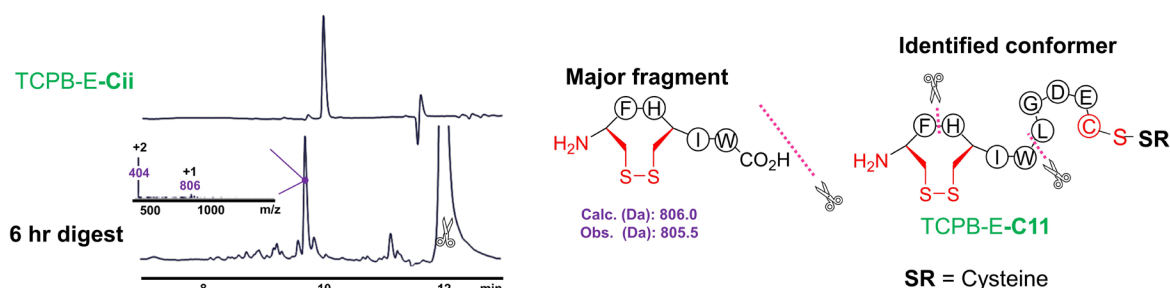


Figure S7.9: LCMS disulfide bond mapping of isolated TCPB-E peak ii, oxidized with cystine. Lyophilized peptide was digested with chymotrypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 6 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peak is shown in bottom left. Identity of the major digested fragment and the deduced, undigested conformer are shown to the right. No digestion was observed between Phe2 and His3 after 6 hours.

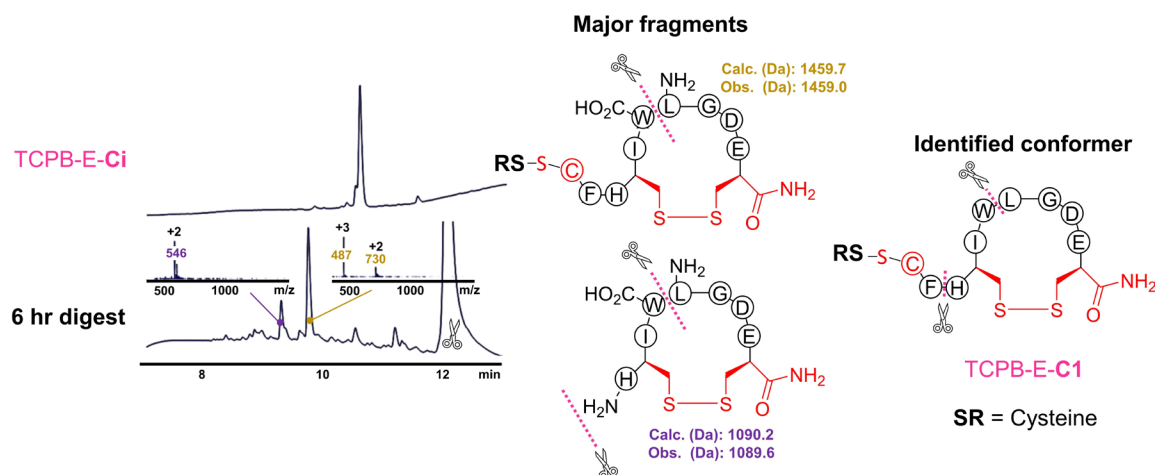


Figure S7.10: LCMS disulfide bond mapping of isolated TCPB-E peak i, oxidized with cystine. Lyophilized peptide was digested with chymotrypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 6 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peaks are shown in bottom left. Identity of the major digested fragments and the deduced, undigested conformer are shown to the right.

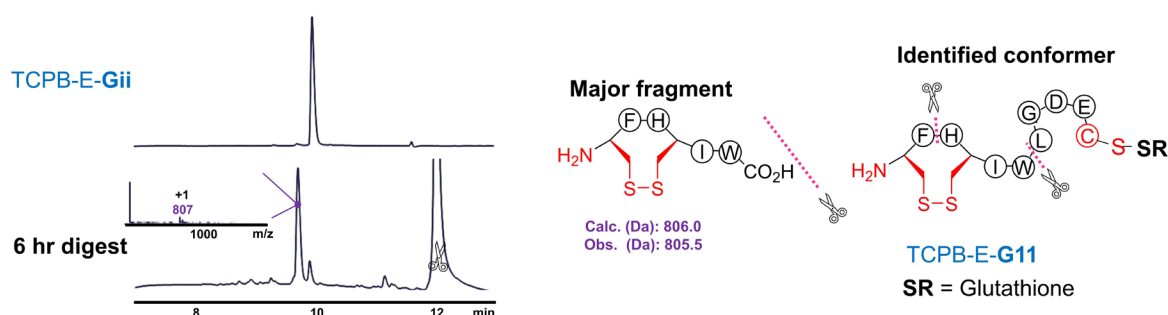


Figure S7.11: LCMS disulfide bond mapping of isolated TCPB-E peak ii, oxidized with glutathione disulfide. Lyophilized peptide was digested with chymotrypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 6 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peak is shown in bottom left. Identity of the major digested fragment and the deduced, undigested conformer are shown to the right. No digestion was observed between Phe2 and His3 after 6 hours.

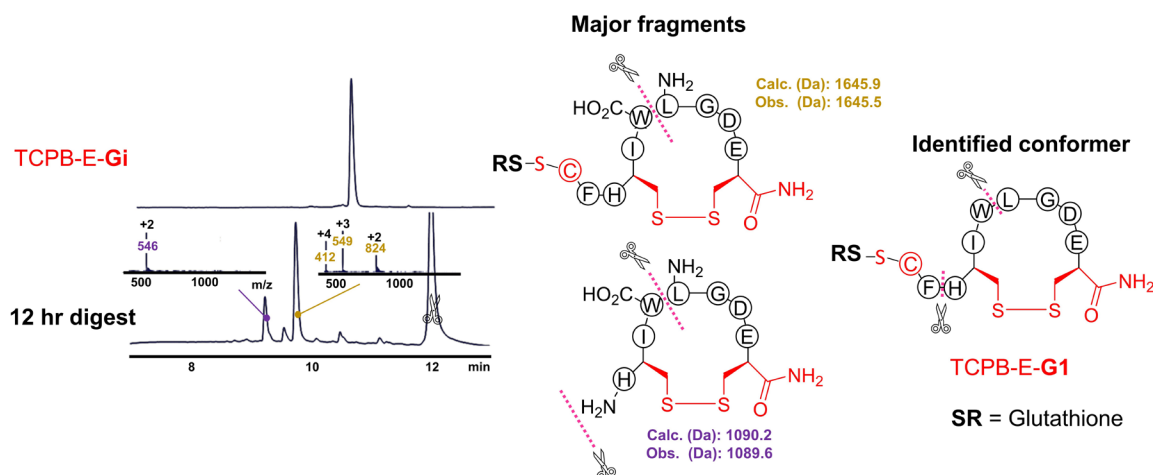


Figure S7.12: LCMS disulfide bond mapping of isolated TCPB-E peak i, oxidized with glutathione disulfide. Lyophilized peptide was digested with chymotrypsin (0.1 mg/mL) in phosphate buffer (0.1 M, pH 6) at room temperature for 12 hours. HPLC trace of undigested peptide is shown in top left, and HPLC trace and corresponding ESI+ MS of major digested peaks are shown in bottom left. Identity of the major digested fragments and the deduced, undigested conformer are shown to the right.

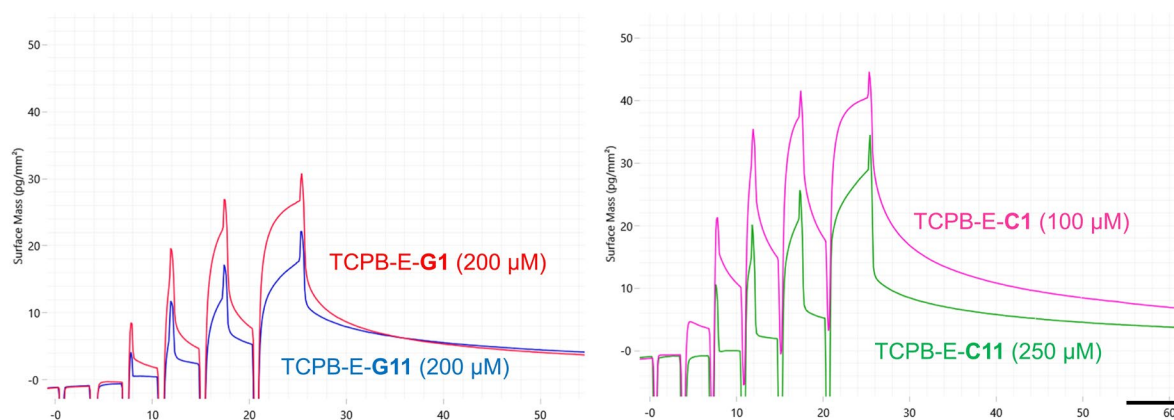


Figure S7.13: Grating-coupled interferometry (GCI) analysis of TCPB-E isomers binding to D-TNRCD2. Repeated analyte pulses of increasing duration (RAPID) were passed over the sensor surface, with fixed concentrations of each peptide conformer shown above. Sensograms shown are blank subtracted (target flow cell - flow cell with no target). The highest sensor response with respect to concentration was observed for TCPB-E-C1.

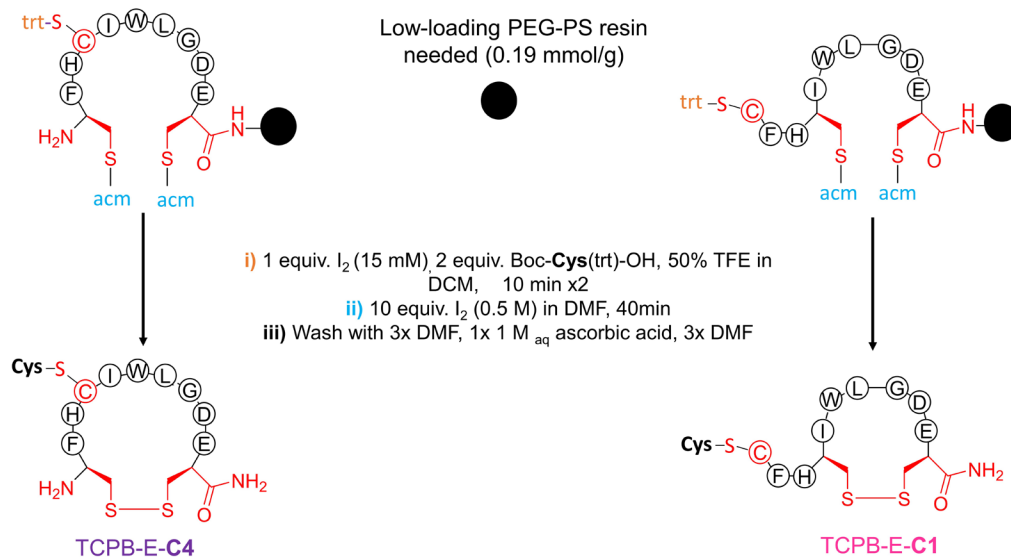


Figure S7.14: Illustrative representation of a solid-phase synthetic route to TCPB-E peptide conformers. An orthogonal protection scheme was implemented using cysteine acetamidomethyl (acm) or trityl (trt). Disulfide bond formation was controlled by the relative reactivity of Iodine towards protected cysteine residues in non-polar (trt-trt) and polar (acm-acm) solvents, based on a modified procedure.²²⁷

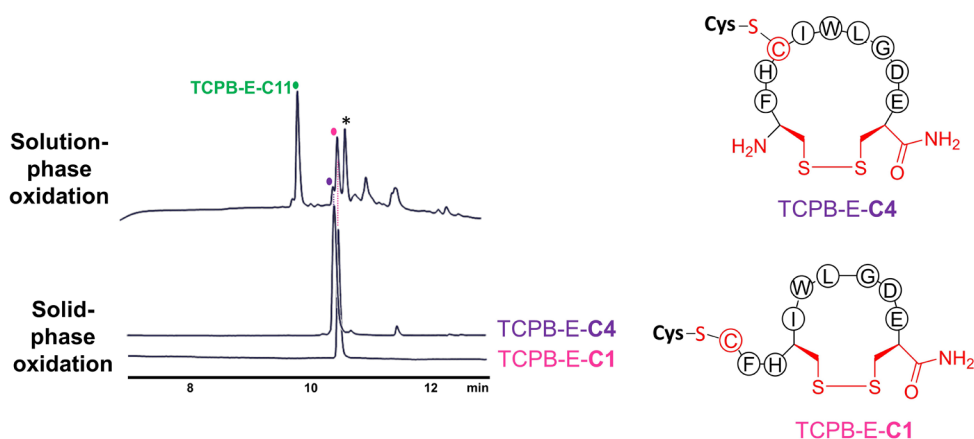


Figure S7.15: Comparison of HPLC retention times of TCPB-E-C1 and -C4 prepared on solid-phase with the solution-phase oxidation of TCPB-E with cystine, reinforcing the results of LCMS disulfide bond mapping. Small quantities of TCPB-E-C4 appear to have been formed in solution-phase but was insufficient for isolation, likely suggesting that this conformer is thermodynamically unfavourable.

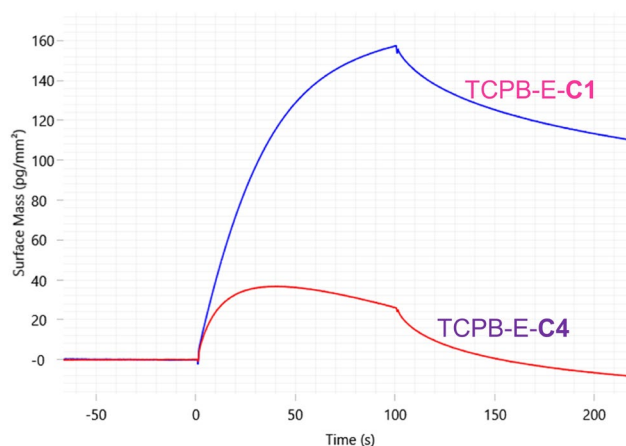


Figure S7.16: Grating-coupled interferometry (GCI) analysis of TCPB-E-**C1** and -**C4** isomers binding to D-TNRCD2. A single cycle of association and dissociation were passed over the sensor surface, with 50 μ M of each peptide conformer. Sensograms shown are blank subtracted (target flow cell - flow cell with no target). The highest sensor response with was observed for TCPB-E-**C1**, with little binding observed for TCPB-E-**C4**.

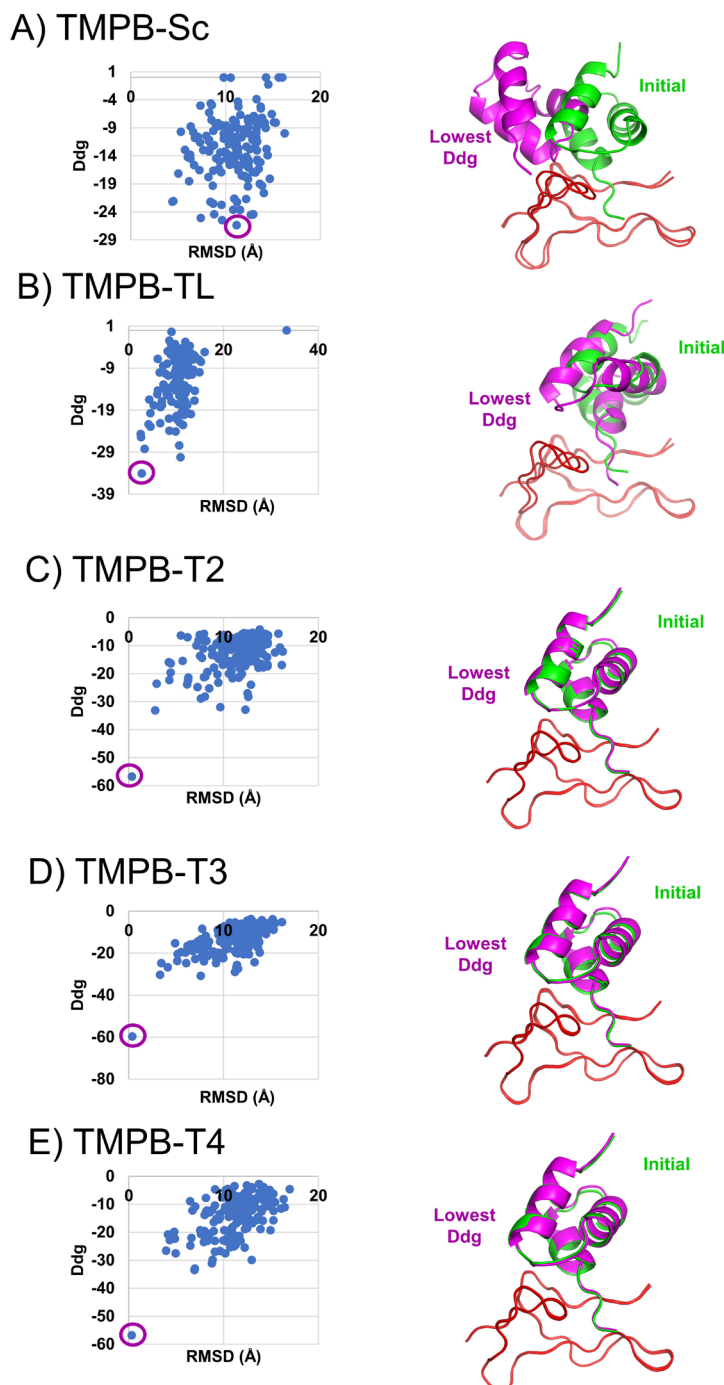


Figure S7.17: RosettaDock results of *de novo* mini protein docking to the D-TNRCD2, showing: A) undesigned scaffold (TMPB-Sc), B) low scoring protein design (TMPB-TL), and C-E) binder candidates (TMPB-T2, -T3 and -T4). Plots of the estimated binding energy of the docked pose (Ddg) versus the deviation (RMSD) of the docked structure to the initial design pose are shown on the left. The lowest Ddg docked structure overlaid with the initial design pose is shown on the right, as a visualization of the single lowest energy point on the plots (highlighted by purple circle). Proteins with low Ddg docked poses and high RMSD to the initial design pose are characterized as poor binders *in silico*.

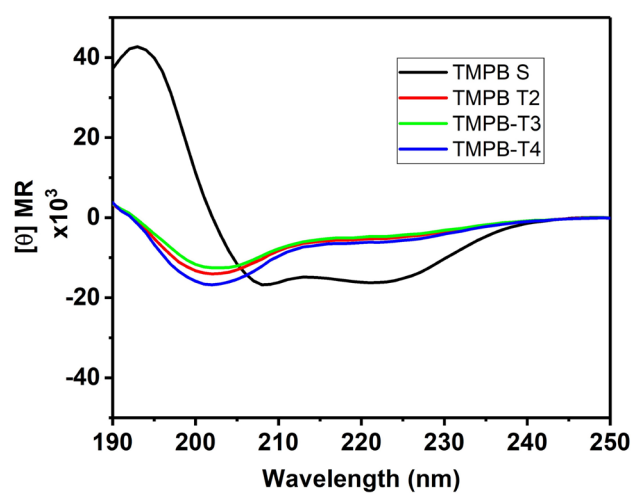


Figure S7.18: Circular dichroism (CD) spectra of designed proteins TMPB-T2, -T3 and -T4 compared to the undesigned scaffold protein TMPB-S, showed a lack of alpha-helical fold following *in silico* protein design.

7.3. Peptide and protein LCMS data

Columns and gradients used for LCMS are as per method 5.3.1, unless otherwise stated in the caption. ESI+ mass spectra are reported as the integrated spectra for the duration of the major peak in each UV210nm chromatogram.

7.3.1. L-aureocin A53 (Met¹-Ile¹⁰) peptide hydrazide

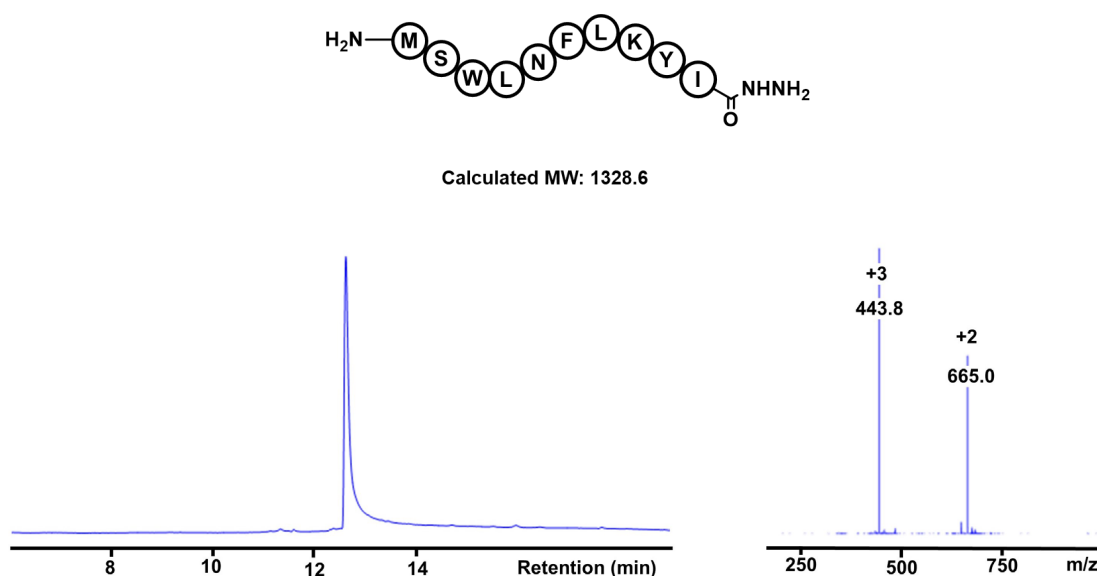
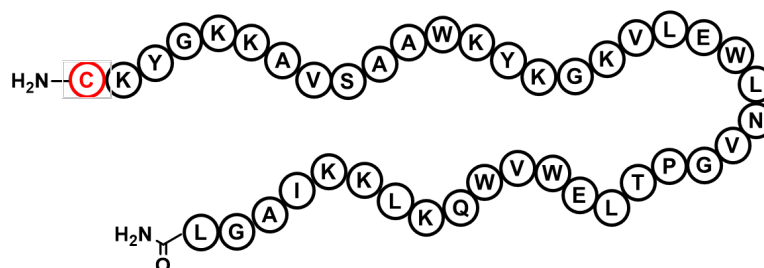


Figure S7.19: LC trace at UV210nm (left) of isolated L-aureocin A53 (Met¹-Ile¹⁰) peptide hydrazide using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 665.3 [M+2H]²⁺, 443.9 [M+3H]³⁺, observed 665.0 [M+2H]²⁺, 443.8 [M+3H]³⁺).

7.3.2. L-aureocin A53 (Cys¹¹-Leu⁵¹) N-cysteine peptide



Calculated MW: 4718.7

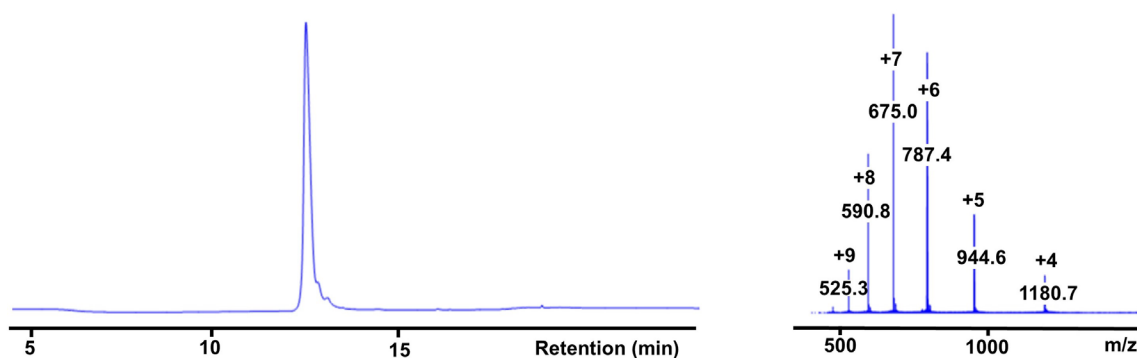


Figure S7.20: LC trace at UV210nm (left) of isolated L-aureocin A53 (Cys¹¹-Leu⁵¹) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfinityLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 1180.7 [M+4H]⁴⁺, 944.7 [M+5H]⁵⁺, 787.4 [M+6H]⁶⁺, 675.1 [M+7H]⁷⁺, 590.8 [M+8H]⁸⁺, 525.3 [M+9H]⁹⁺, observed 1180.7 [M+4H]⁴⁺, 944.6 [M+5H]⁵⁺, 787.4 [M+6H]⁶⁺, 675.0 [M+7H]⁷⁺, 590.8 [M+8H]⁸⁺, 525.3 [M+9H]⁹⁺).

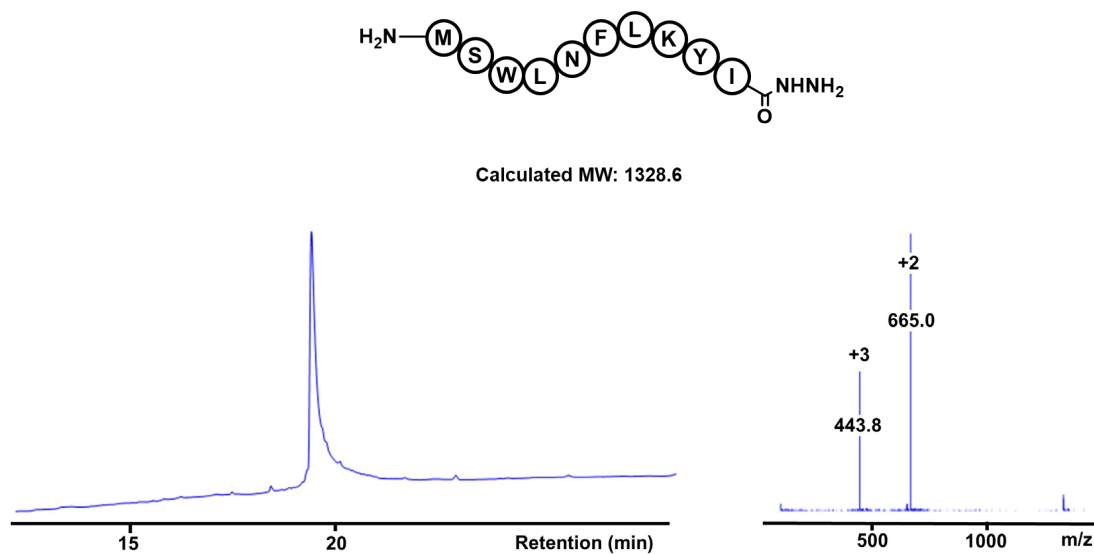
7.3.3. D-aureocin A53 (Met¹-Ile¹⁰) peptide hydrazide

Figure S7.21: LC trace at UV210nm (left) of isolated D-aureocin A53 (Met¹-Ile¹⁰) peptide hydrazide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfinityLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 665.3 [M+2H]²⁺, 443.9 [M+3H]³⁺, observed 665.0 [M+2H]²⁺, 443.8 [M+3H]³⁺).

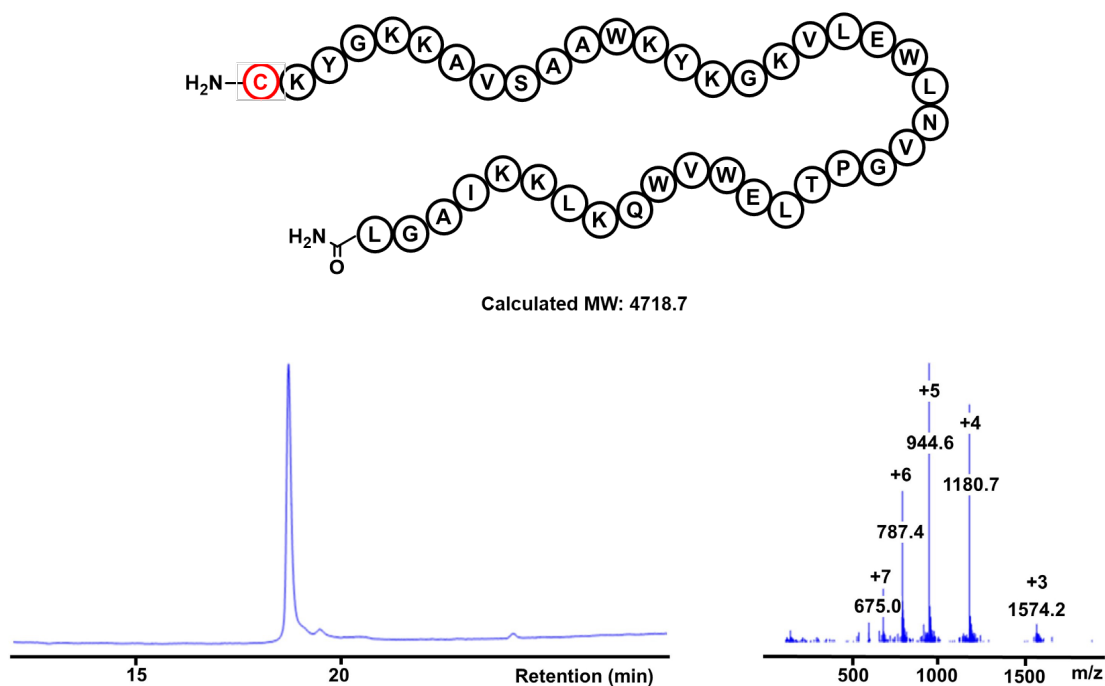
7.3.4. D-aureocin A53 (Cys¹¹-Leu⁵¹) N-cysteine peptide

Figure S7.22: LC trace at UV210nm (left) of isolated D-aureocin A53 (Cys¹¹-Leu⁵¹) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfinityLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 1572.9 [M+3H]³⁺, 1180.7 [M+4H]⁴⁺, 944.7 [M+5H]⁵⁺, 787.4 [M+6H]⁶⁺, 675.1 [M+7H]⁷⁺, observed 1574.2 [M+3H]³⁺, 1180.7 [M+4H]⁴⁺, 944.6 [M+5H]⁵⁺, 787.4 [M+6H]⁶⁺, 675.0 [M+7H]⁷⁺).

7.3.5. Example LCMS ligation monitoring of lactacin Q

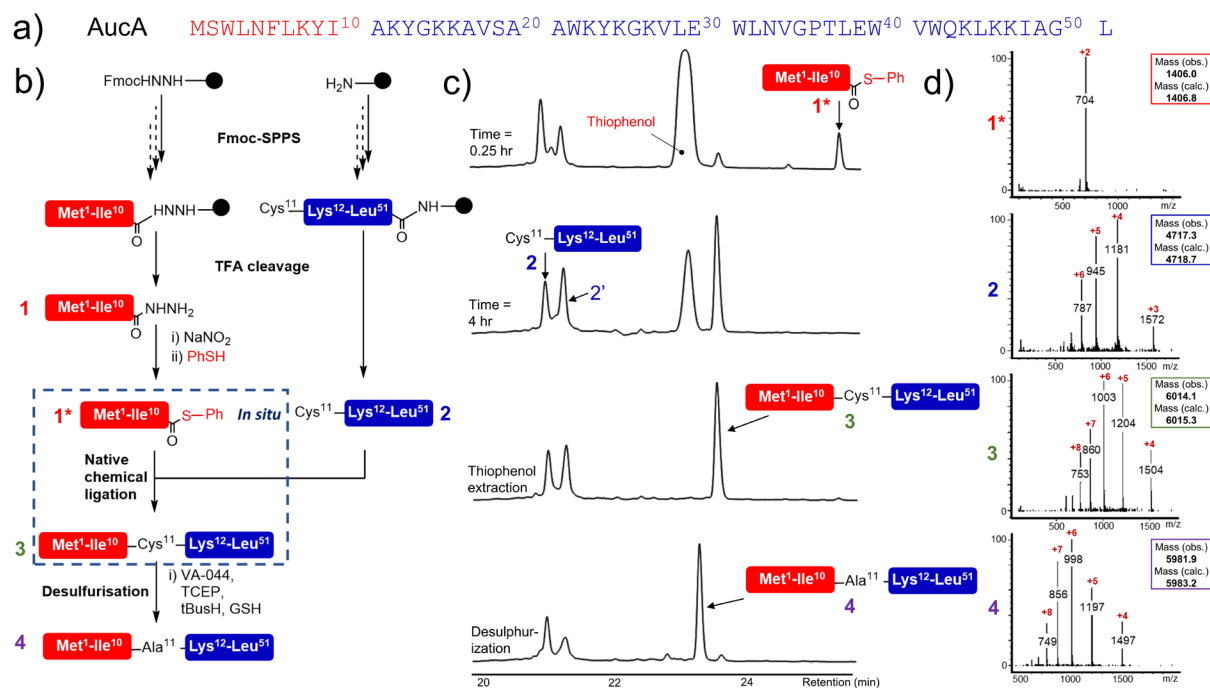


Figure S7.23: Example LCMS ligation scheme for AucA. a) AucA amino acid sequence, b) one-pot ligation-desulfurization scheme, c) HPLC reaction monitoring, and d) ESI+ MS for AucA product, reactants and intermediates.

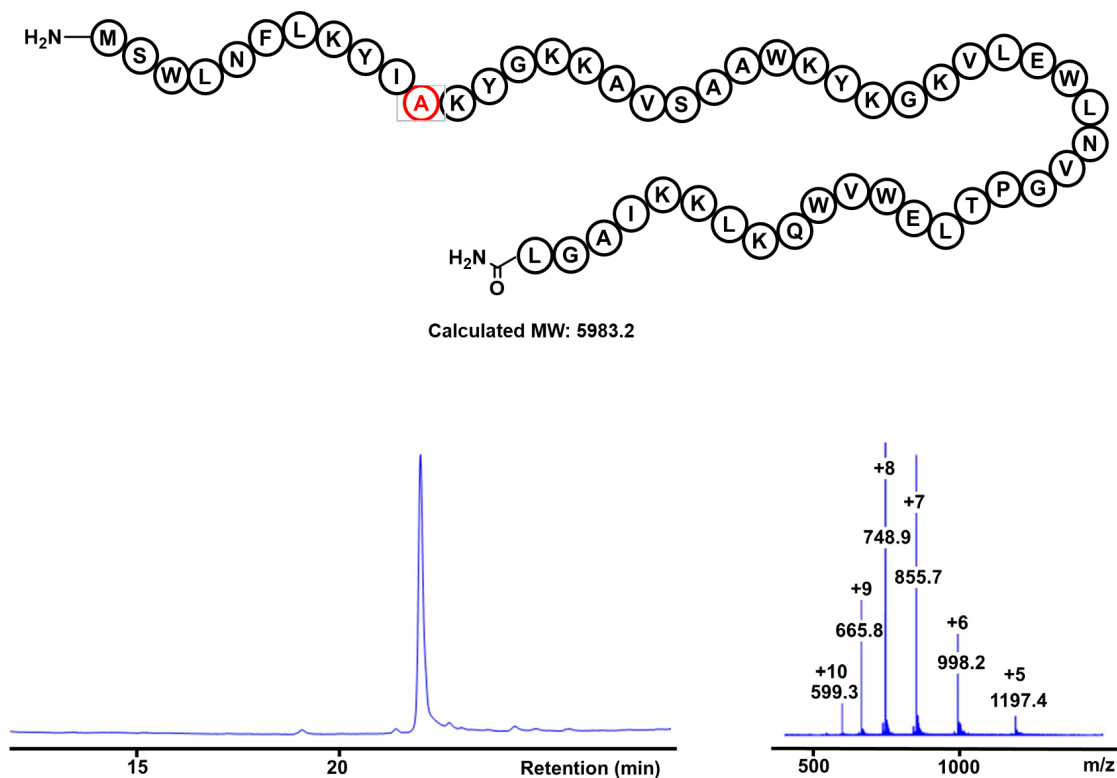
7.3.6. L-aureocin A53 (Met¹-Leu⁵¹)

Figure S7.24: LC trace at UV210nm (left) of isolated L-aureocin A53 (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1197.6 [M+5H]⁵⁺, 998.2 [M+6H]⁶⁺, 855.7 [M+7H]⁷⁺, 748.9 [M+8H]⁸⁺, 665.8 [M+9H]⁹⁺, 599.3 [M+10H]¹⁰⁺, observed 1197.4 [M+5H]⁵⁺, 998.2 [M+6H]⁶⁺, 855.7 [M+7H]⁷⁺, 748.9 [M+8H]⁸⁺, 665.8 [M+9H]⁹⁺, 599.3 [M+10H]¹⁰⁺).

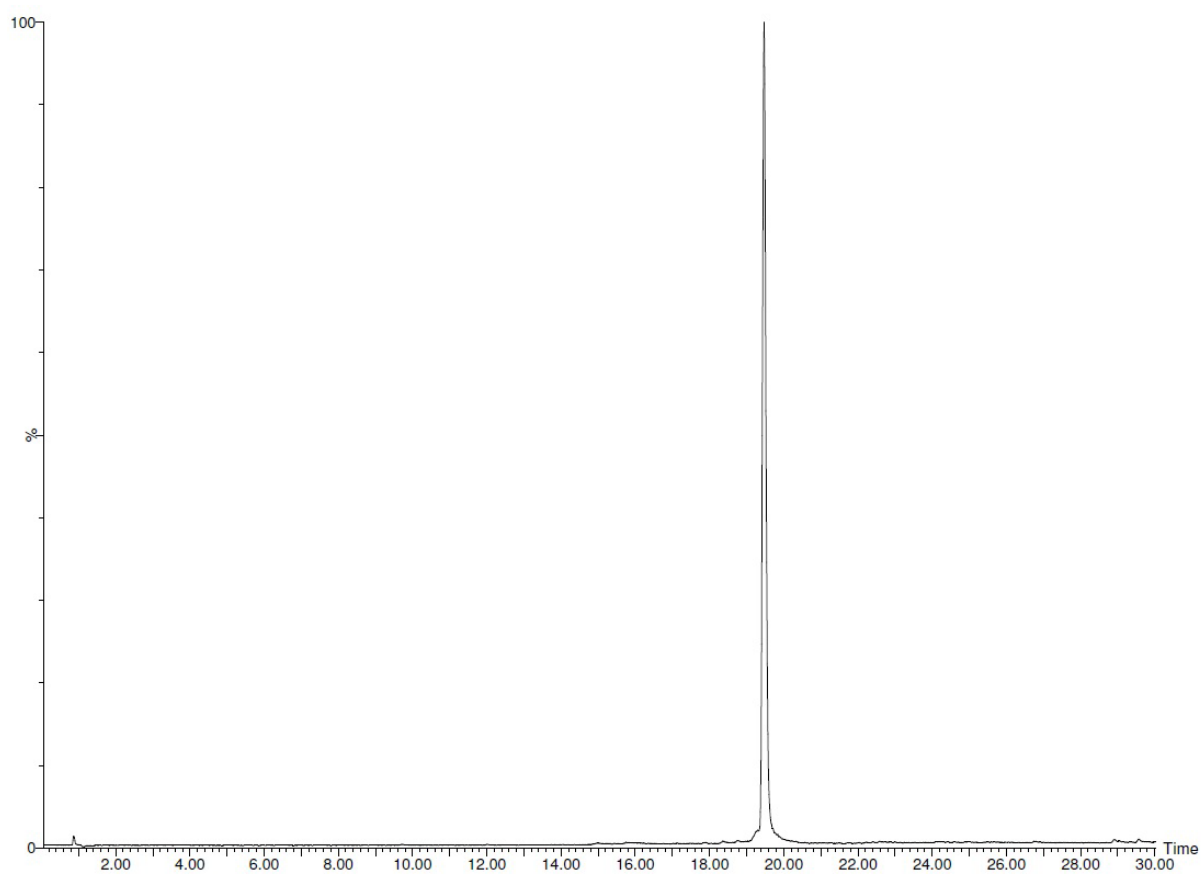


Figure S7.25: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 (Met¹-Leu⁵¹) [19.47 min].

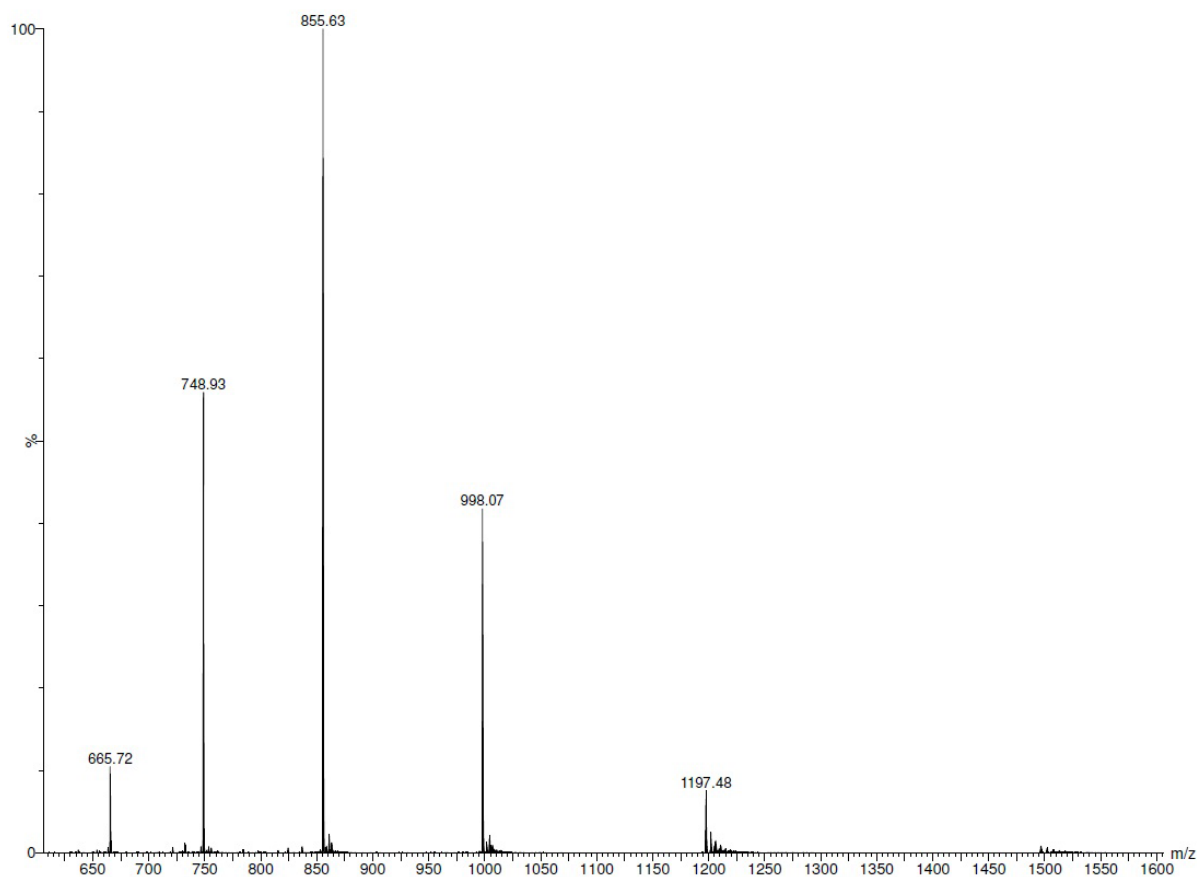


Figure S7.26: High-definition ESI+ mass spec of isolated L-aureocin A53 (Met¹-Leu⁵¹), (ESI-MS (m/z): calculated 1197.6 [M+5H]⁵⁺, 998.2 [M+6H]⁶⁺, 855.7 [M+7H]⁷⁺, 748.9 [M+8H]⁸⁺, 665.8 [M+9H]⁹⁺, observed 1197.5 [M+5H]⁵⁺, 998.0 [M+6H]⁶⁺, 855.6 [M+7H]⁷⁺, 748.9 [M+8H]⁸⁺, 665.7 [M+9H]⁹⁺, deconvoluted: calculated 5983.2, observed 5983.5).

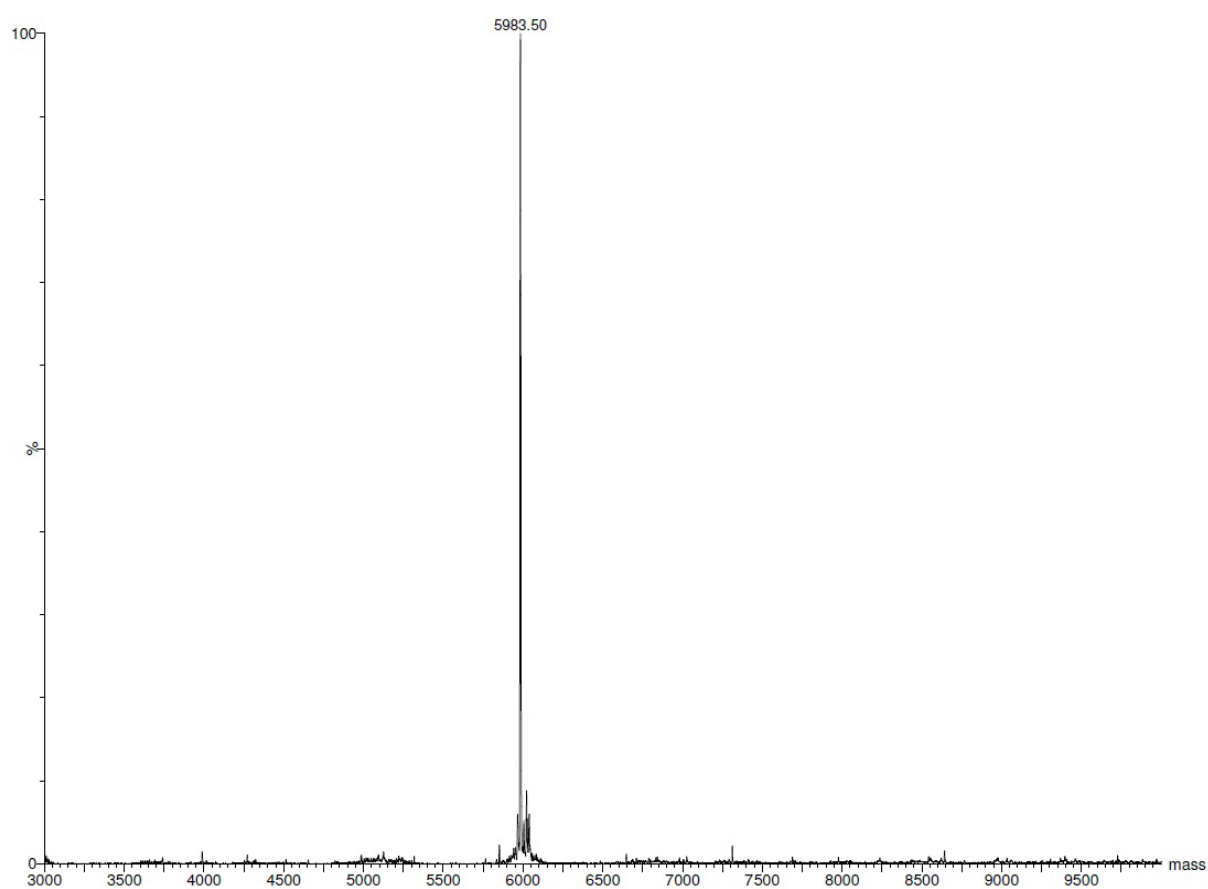


Figure S7.27: High-definition ESI+ mass spec of isolated L-aureocin A53 (Met¹-Leu⁵¹), (ESI-MS (m/z) deconvoluted: calculated 5983.2, observed 5983.5).

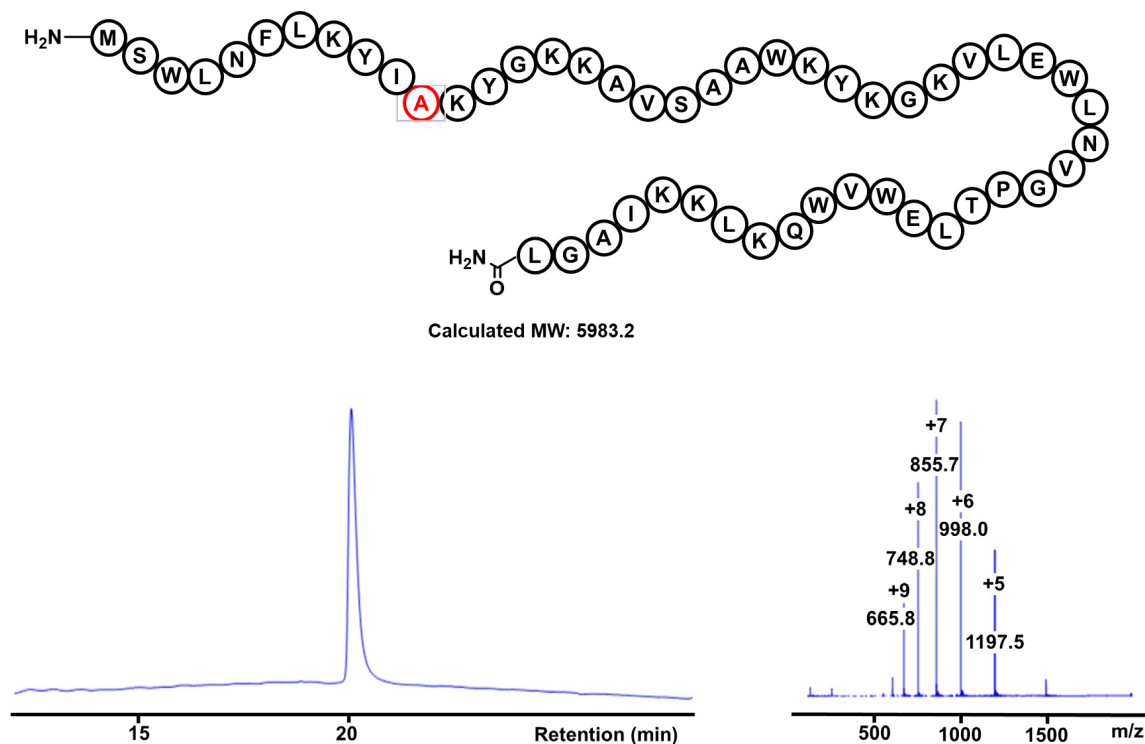
7.3.7. D-aureocin A53 (Met¹-Leu⁵¹)

Figure S7.28: LC trace at UV210nm (left) of isolated D-aureocin A53 (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfintiyLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 1197.6 [M+5H]⁵⁺, 998.2 [M+6H]⁶⁺, 855.7 [M+7H]⁷⁺, 748.9 [M+8H]⁸⁺, 665.8 [M+9H]⁹⁺, observed 1197.5 [M+5H]⁵⁺, 998.0 [M+6H]⁶⁺, 855.7 [M+7H]⁷⁺, 748.8 [M+8H]⁸⁺, 665.8 [M+9H]⁹⁺).

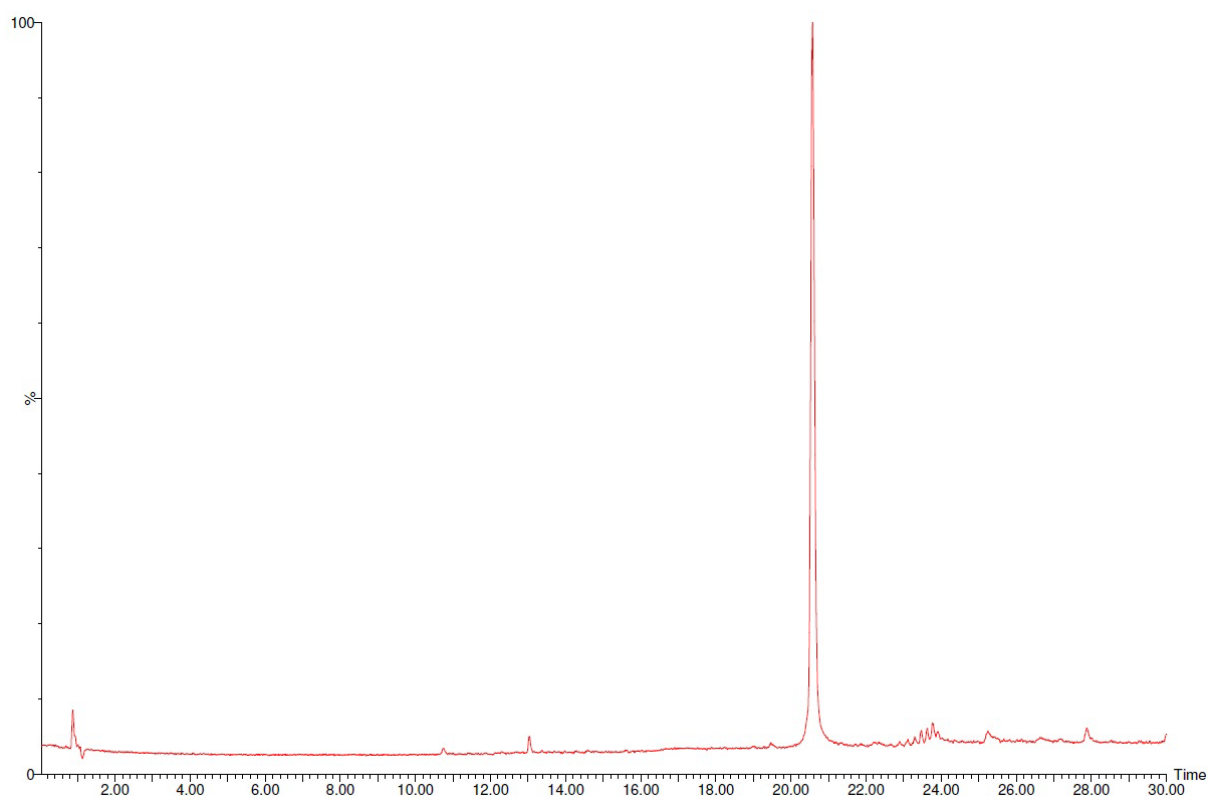


Figure S7.29: UPLC total ion chromatogram in ESI+ mode of isolated D-aureocin A53 (Met¹-Leu⁵¹) [20.57 min].

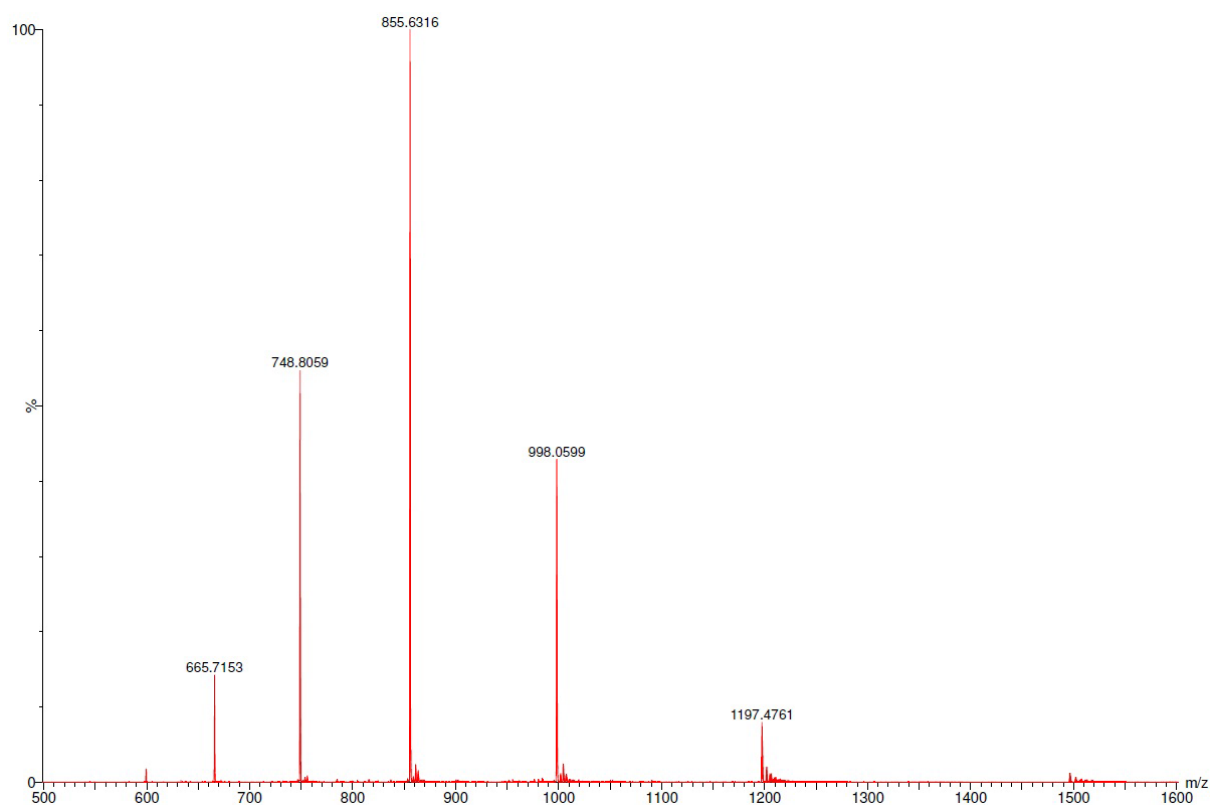


Figure S7.30: High-definition ESI+ mass spec of isolated D-aureocin A53 (Met¹-Leu⁵¹), (ESI-MS (m/z): calculated 1197.6 [M+5H]⁵⁺, 998.2 [M+6H]⁶⁺, 855.7 [M+7H]⁷⁺, 748.9 [M+8H]⁸⁺, 665.8 [M+9H]⁹⁺, observed 1197.5 [M+5H]⁵⁺, 998.1 [M+6H]⁶⁺, 855.6 [M+7H]⁷⁺, 748.8 [M+8H]⁸⁺, 665.7 [M+9H]⁹⁺, deconvoluted: calculated 5983.2, observed 5983.0).

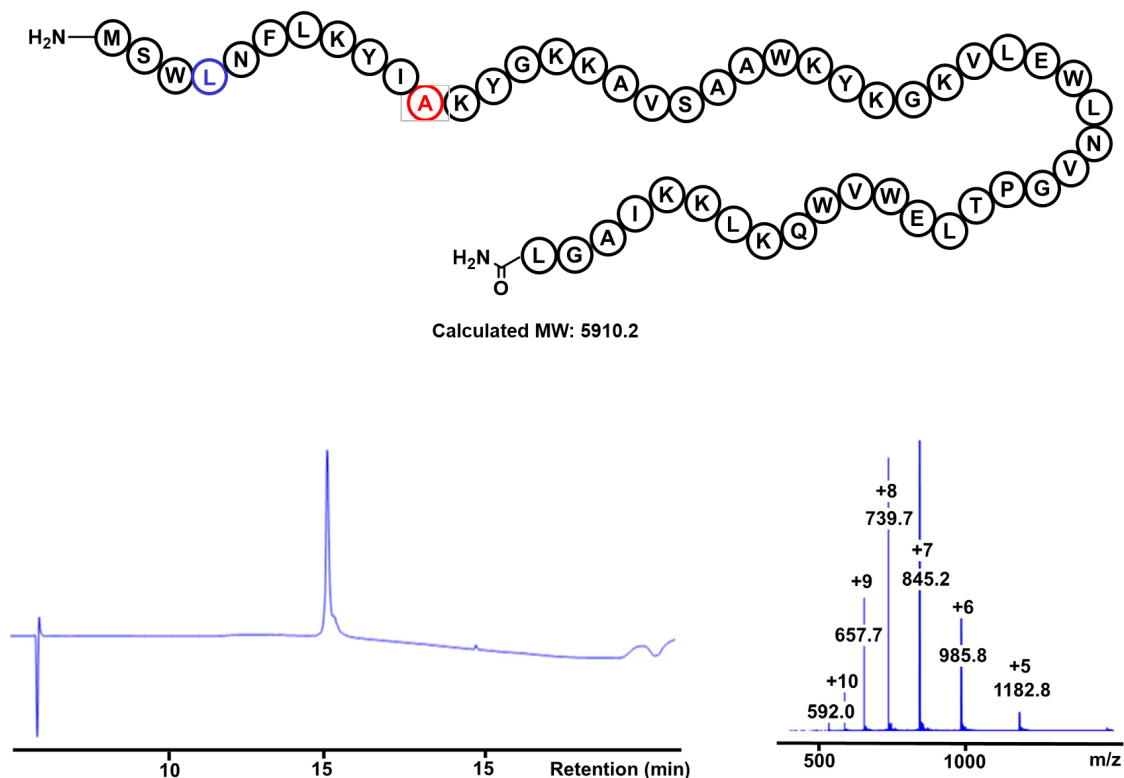
7.3.8. L-aureocin A53 W3L (Met¹-Leu⁵¹)

Figure 7.31: LC trace at UV210nm (left) of isolated L-aureocin A53 W3L (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.8 [M+5H]⁵⁺, 985.8 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺).

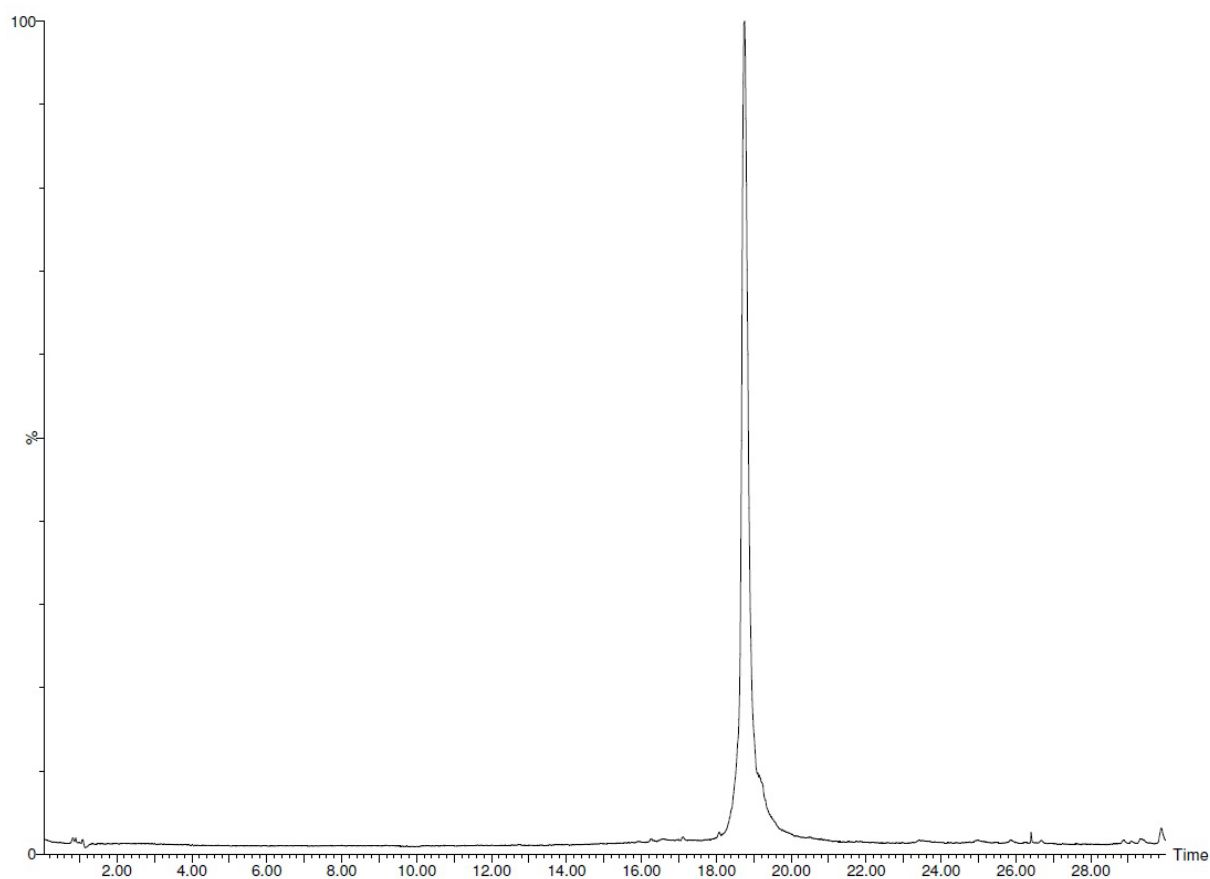


Figure S7.32: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 W3L (Met¹-Leu⁵¹) [18.78 min].

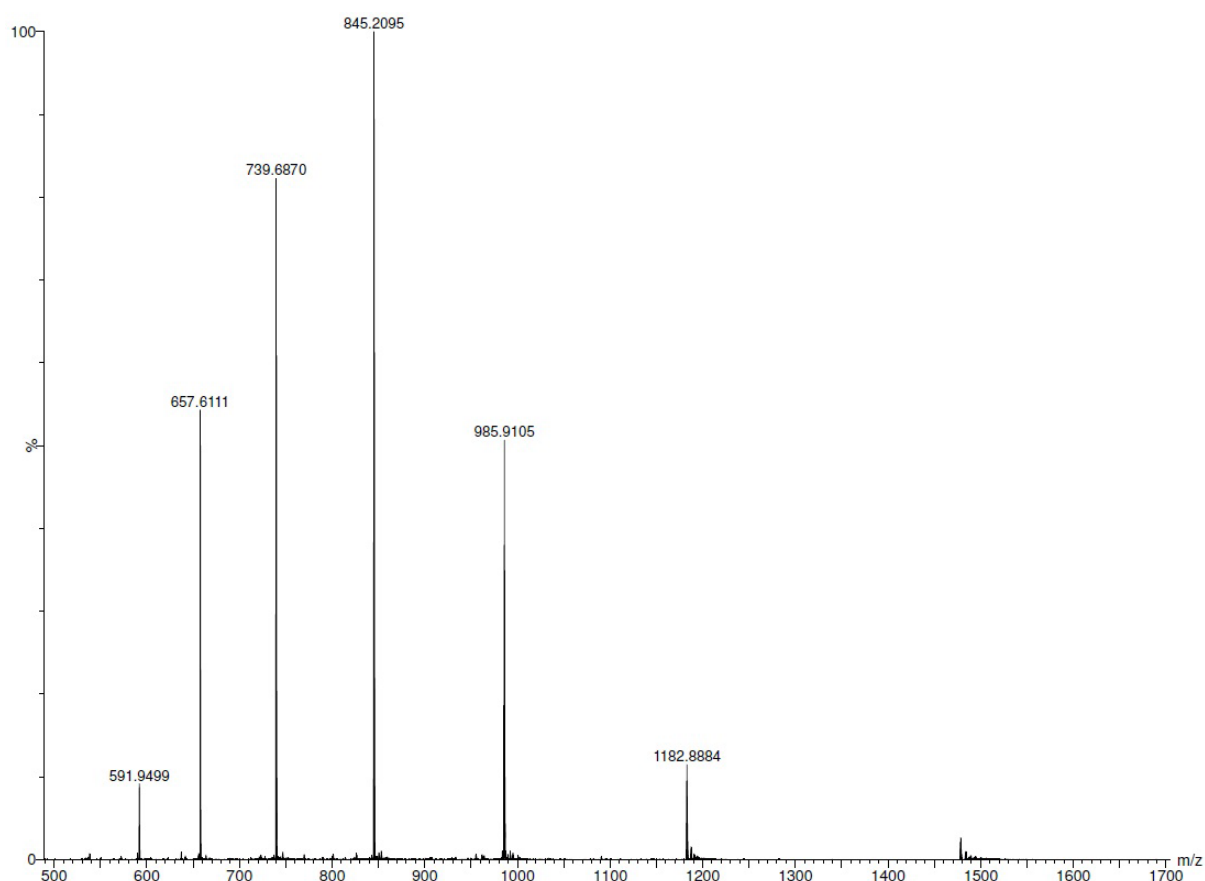


Figure S7.33: High-definition ESI+ mass spec of isolated L-aureocin A53 W3L (Met¹-Leu⁵¹), (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.9 [M+5H]⁵⁺, 985.9 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.6 [M+9H]⁹⁺, 591.9 [M+10H]¹⁰⁺, deconvoluted: calculated 5910.2, observed 5909.5).

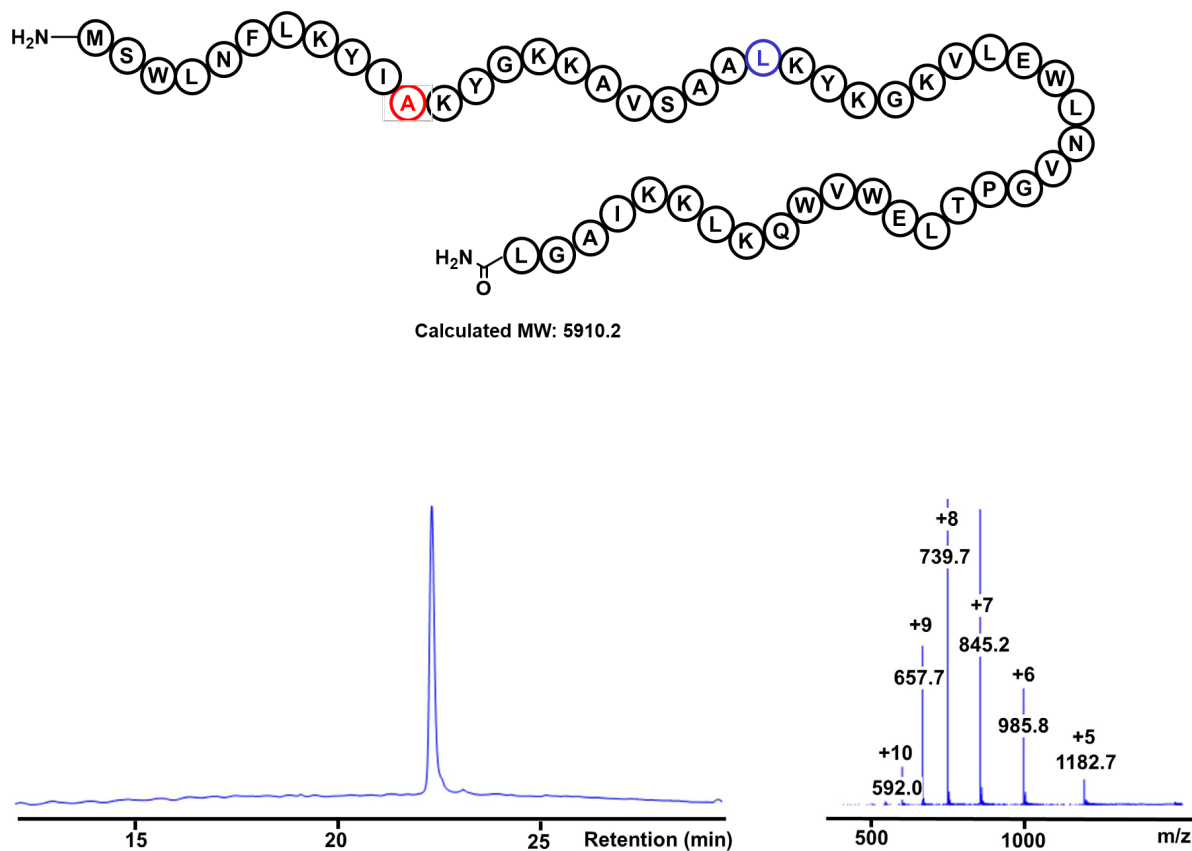
7.3.9. L-aureocin A53 W22L (Met¹-Leu⁵¹)

Figure S7.34: LC trace at UV210nm (left) of isolated L-aureocin A53 W22L (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.7 [M+5H]⁵⁺, 985.8 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺).

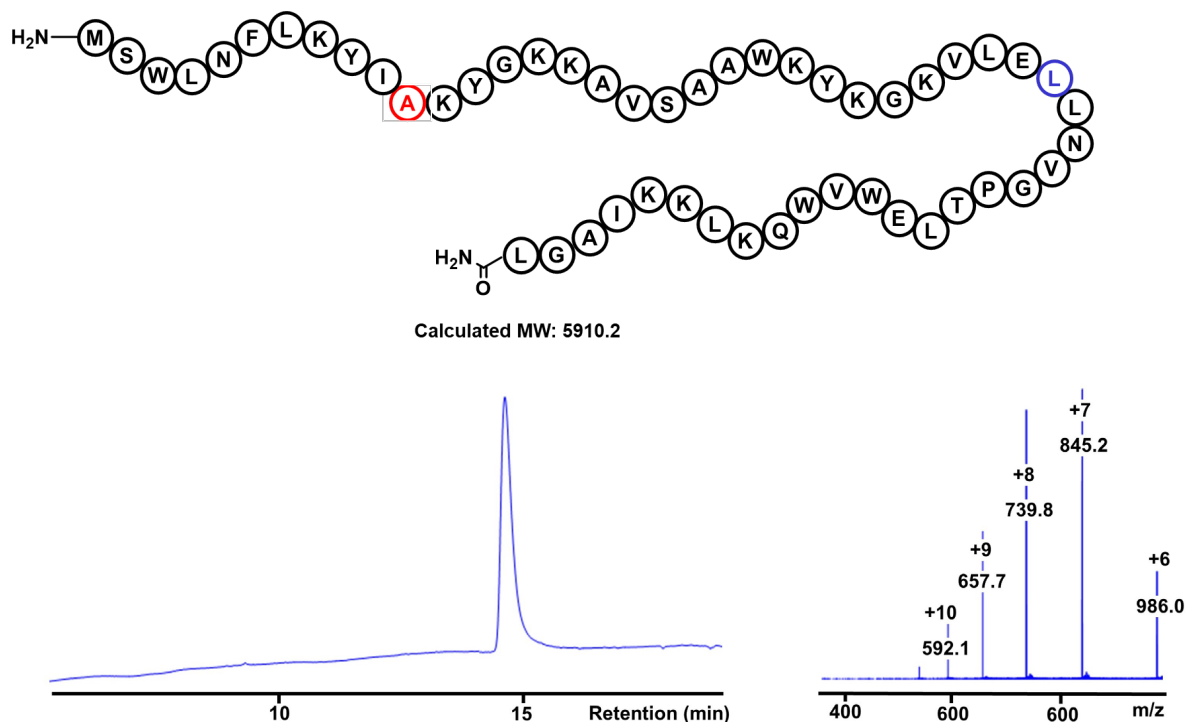
7.3.10. L-aureocin A53 W31L (Met¹-Leu⁵¹)

Figure S7.35: LC trace at UV210nm (left) of isolated L-aureocin A53 W31L (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 986.0 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.1 [M+10H]¹⁰⁺).

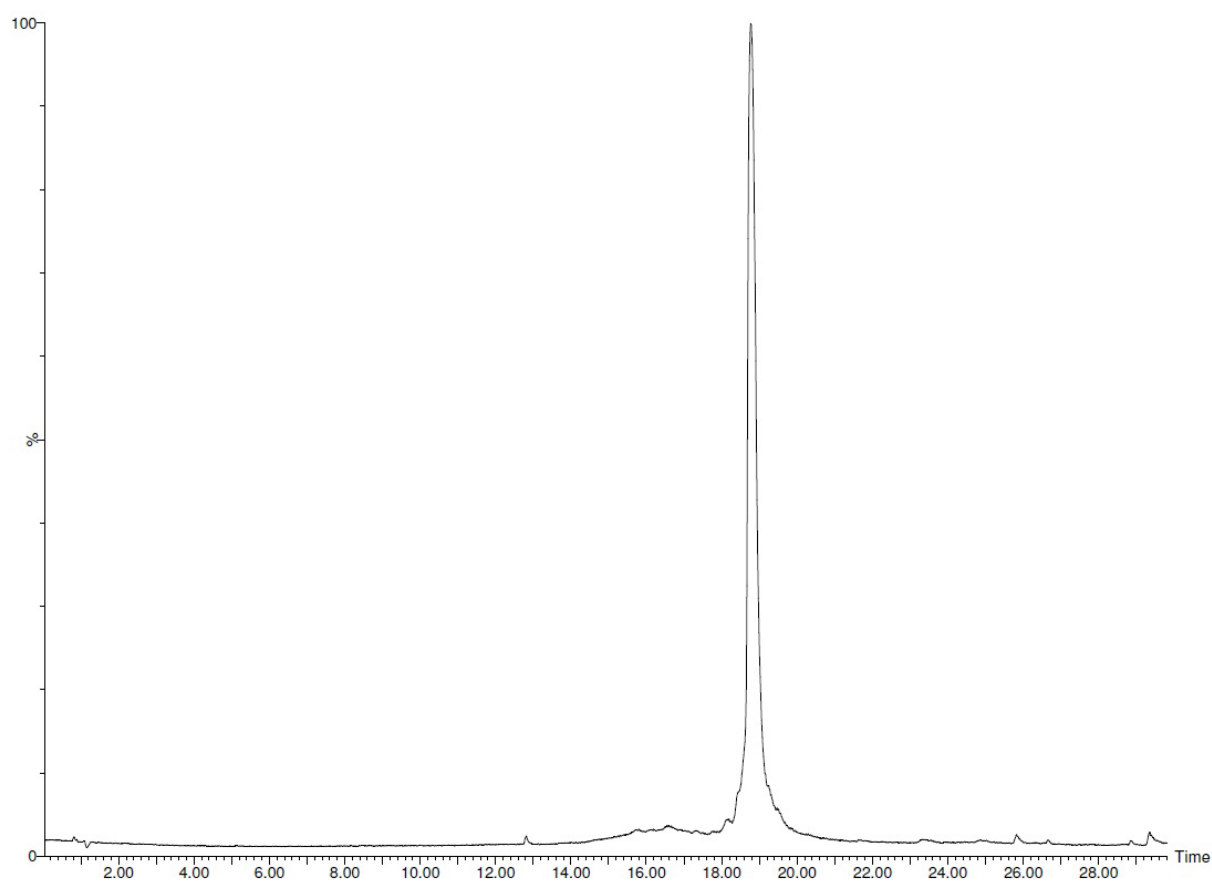


Figure S7.36: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 W31L (Met¹-Leu⁵¹) [18.77 min].

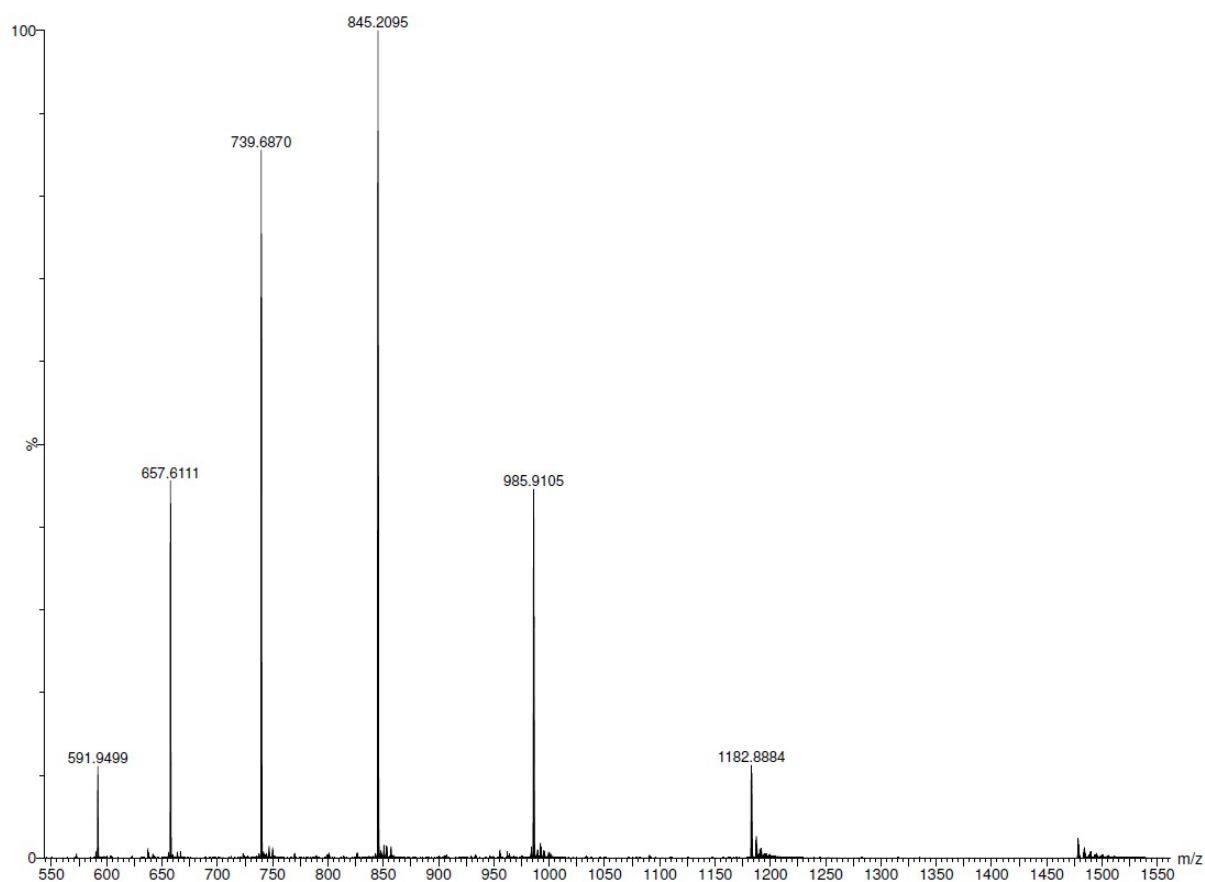


Figure S7.37: High-definition ESI+ mass spec of isolated L-aureocin A53 W31L (Met¹-Leu⁵¹), (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.9 [M+5H]⁵⁺, 985.9 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.6 [M+9H]⁹⁺, 591.9 [M+10H]¹⁰⁺, deconvoluted: calculated 5910.2, observed 5909.5).

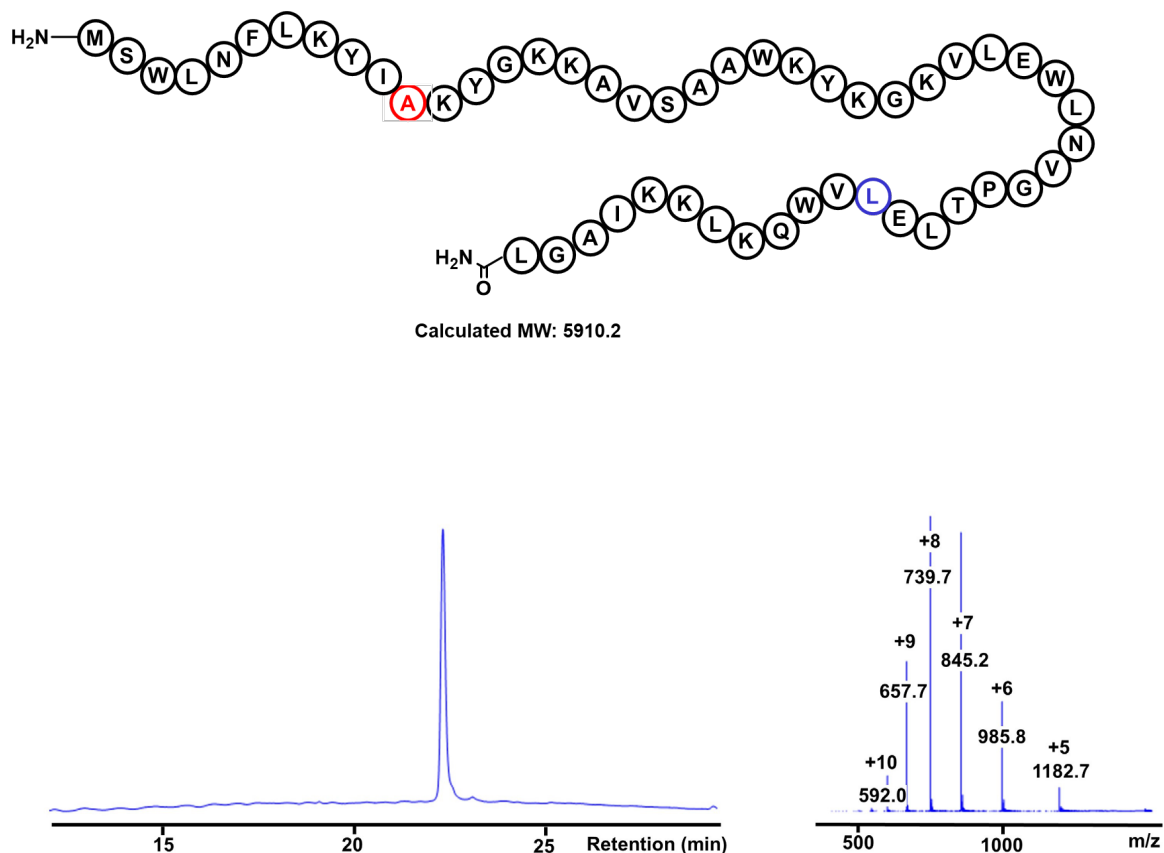
7.3.11. L-aureocin A53 W40L (Met¹-Leu⁵¹)

Figure S7.38: LC trace at UV210nm (left) of isolated L-aureocin A53 W40L (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.7 [M+5H]⁵⁺, 985.8 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺).

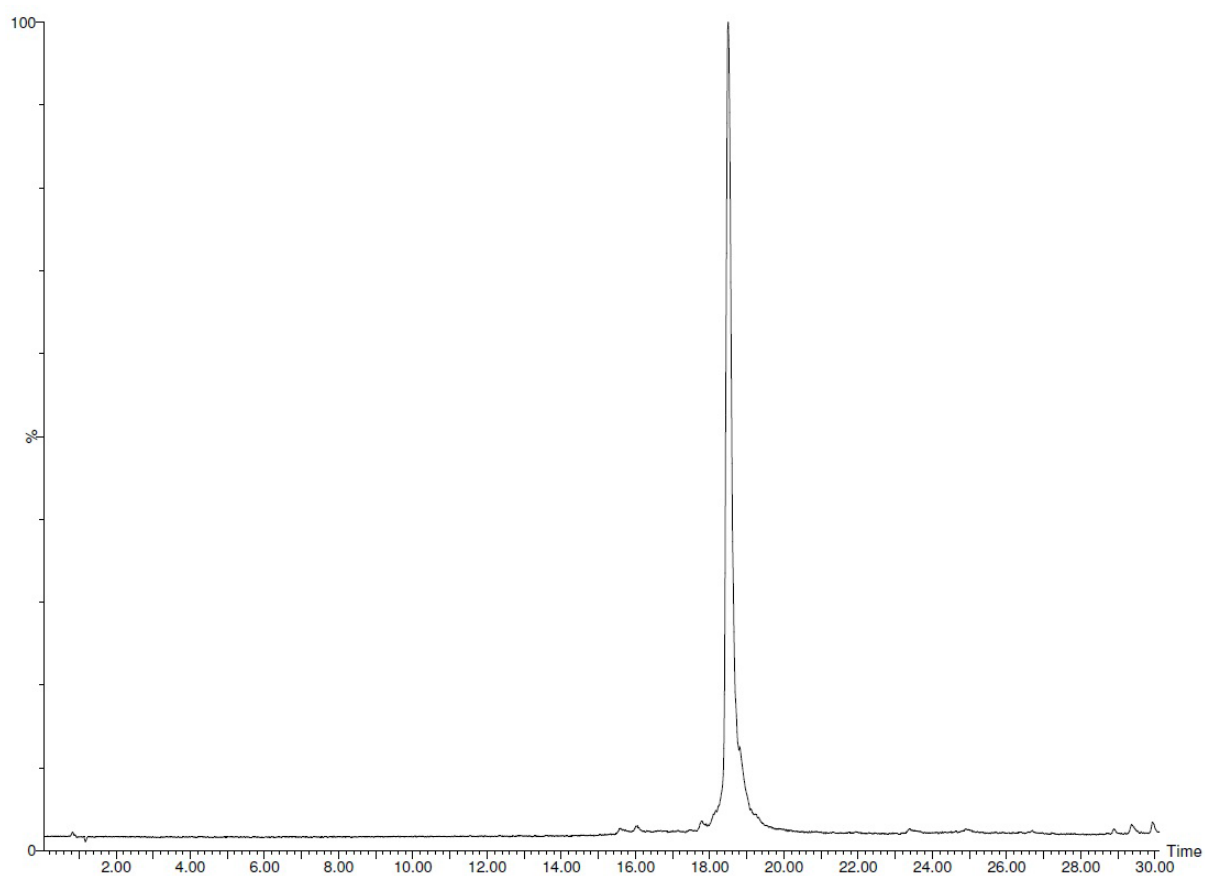


Figure S7.39: UPLC total ion chromatogram in ESI+ mode of isolated L-aureocin A53 W40L (Met¹-Leu⁵¹) [18.50 min].

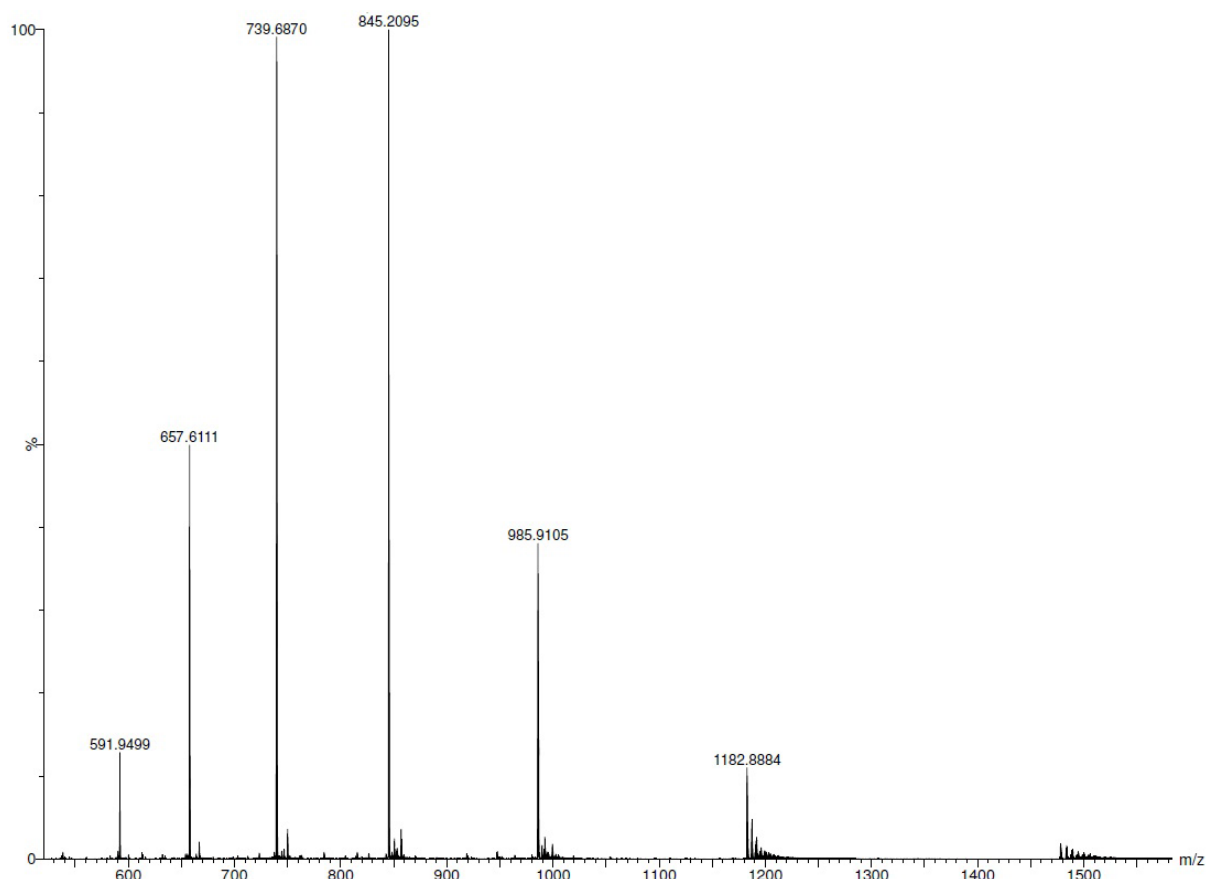


Figure S7.40: High-definition ESI+ mass spec of isolated L-aureocin A53 W40L (Met¹-Leu⁵¹), (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.9 [M+5H]⁵⁺, 985.9 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.6 [M+9H]⁹⁺, 591.9 [M+10H]¹⁰⁺, deconvoluted: calculated 5910.2, observed 5909.5).

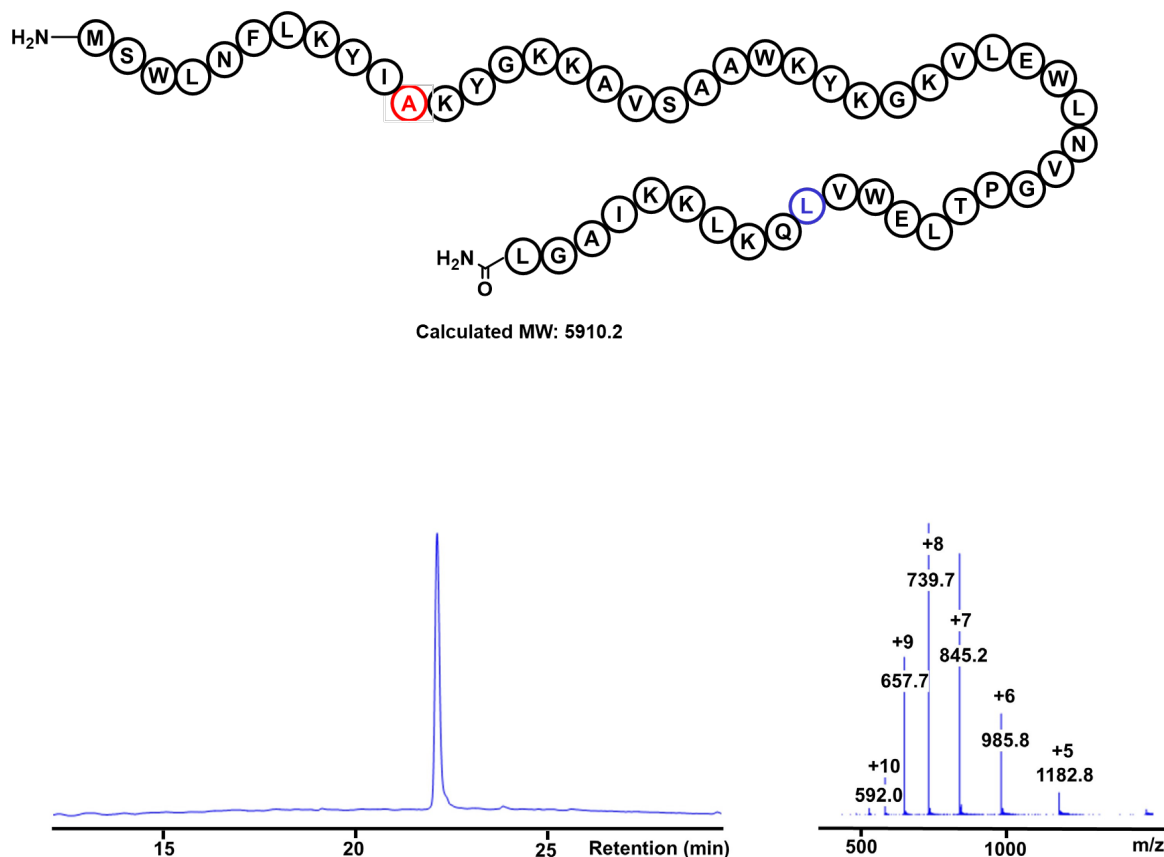
7.3.12. L-aureocin A53 W42L (Met¹-Leu⁵¹)

Figure S7.41: LC trace at UV210nm (left) of isolated L-aureocin A53 W42L (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1183.0 [M+5H]⁵⁺, 986.0 [M+6H]⁶⁺, 845.3 [M+7H]⁷⁺, 739.8 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺, observed 1182.8 [M+5H]⁵⁺, 985.8 [M+6H]⁶⁺, 845.2 [M+7H]⁷⁺, 739.7 [M+8H]⁸⁺, 657.7 [M+9H]⁹⁺, 592.0 [M+10H]¹⁰⁺).

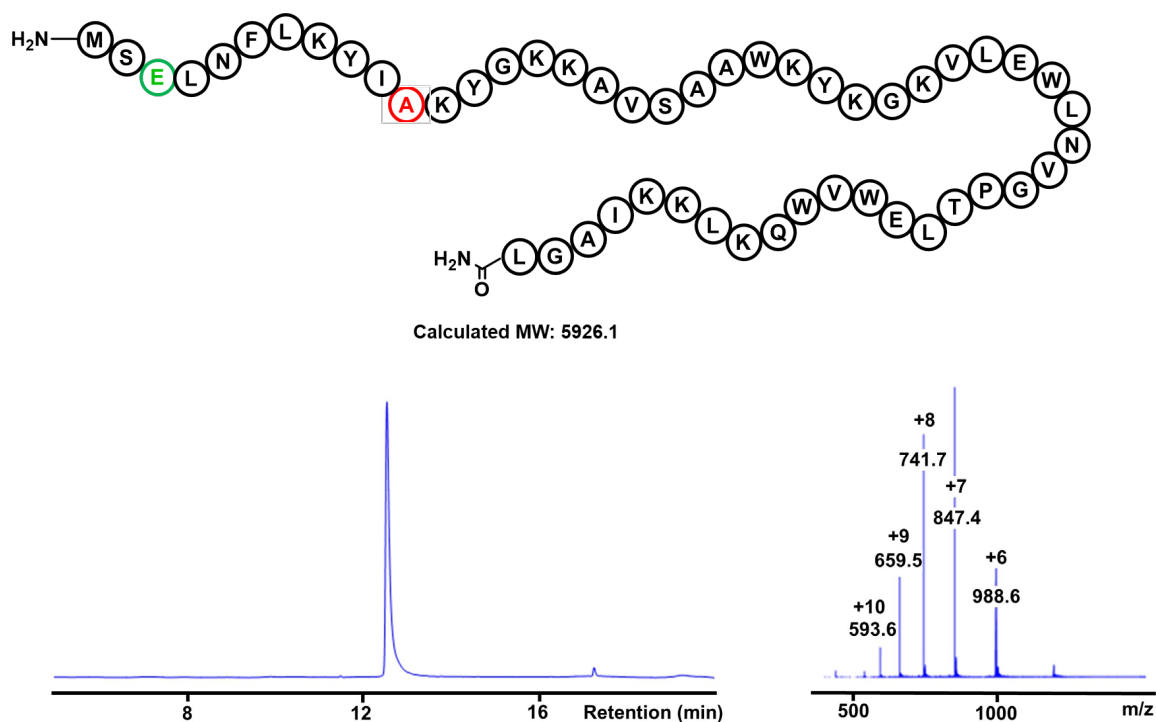
7.3.13. L-aureocin A53 W3E (Met¹-Leu⁵¹)

Figure S7.42: LC trace at UV210nm (left) of isolated L-aureocin A53 W3E (Met¹-Leu⁵¹) using a 20-90% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 988.7 [M+6H]⁶⁺, 847.6 [M+7H]⁷⁺, 741.8 [M+8H]⁸⁺, 659.5 [M+9H]⁹⁺, 593.6 [M+10H]¹⁰⁺, observed 988.6 [M+6H]⁶⁺, 847.4 [M+7H]⁷⁺, 741.7 [M+8H]⁸⁺, 659.5 [M+9H]⁹⁺, 593.6 [M+10H]¹⁰⁺).

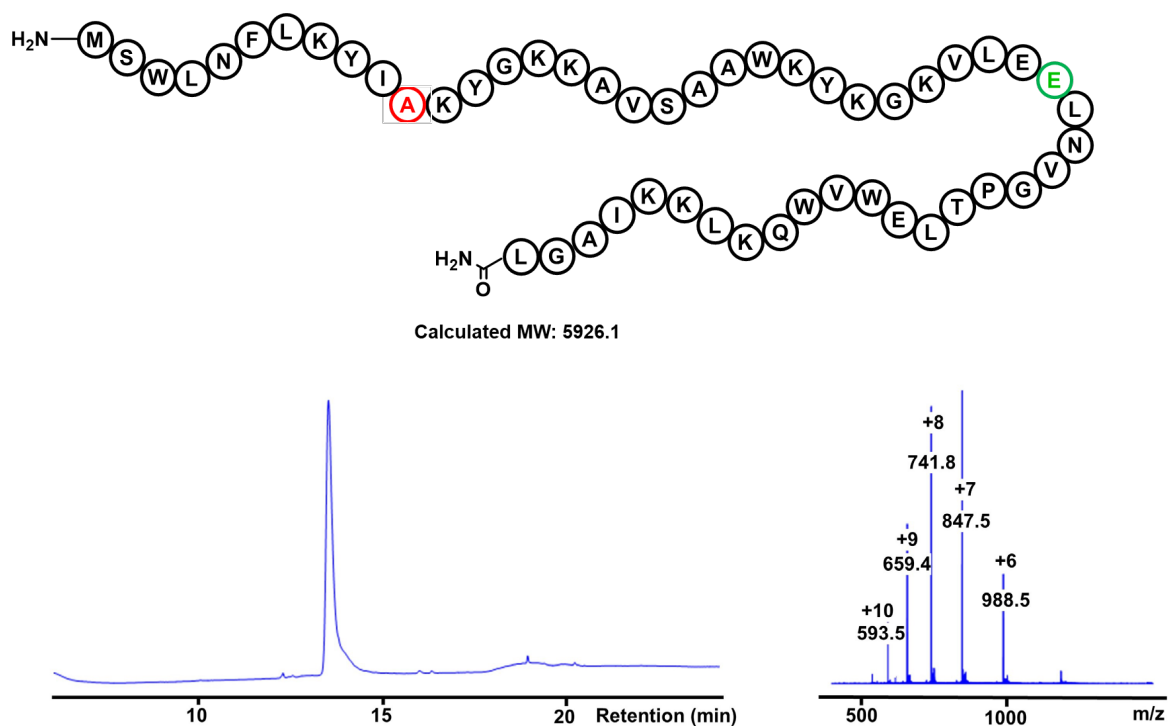
7.3.14. L-aureocin A53 W31E (Met¹-Leu⁵¹)

Figure S7.43: LC trace at UV210nm (left) of isolated L-aureocin A53 W31E (Met¹-Leu⁵¹) using a 5-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 988.7 [M+6H]⁶⁺, 847.6 [M+7H]⁷⁺, 741.8 [M+8H]⁸⁺, 659.5 [M+9H]⁹⁺, 593.6 [M+10H]¹⁰⁺, observed 988.5 [M+6H]⁶⁺, 847.5 [M+7H]⁷⁺, 741.8 [M+8H]⁸⁺, 659.4 [M+9H]⁹⁺, 593.5 [M+10H]¹⁰⁺).

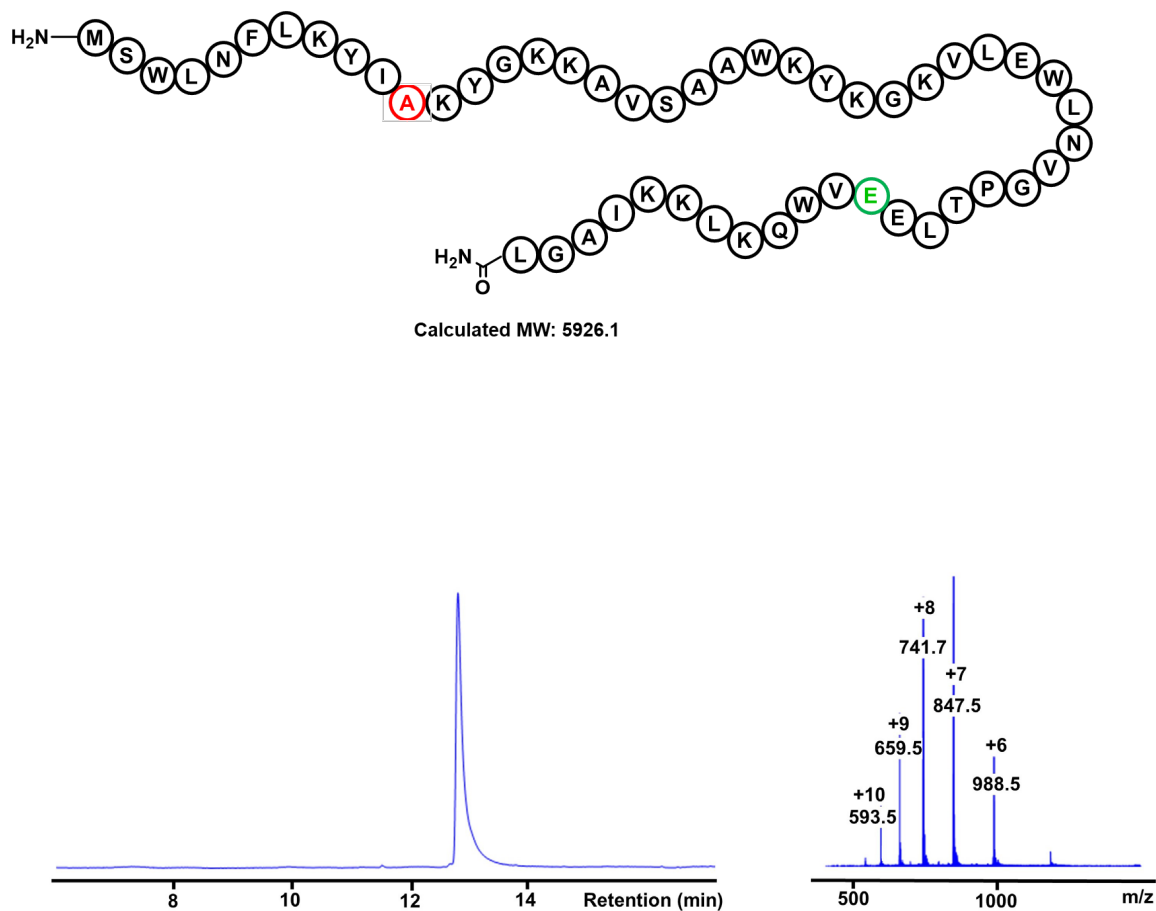
7.3.15. L-aureocin A53 W40E (Met¹-Leu⁵¹)

Figure S7.44: LC trace at UV210nm (left) of isolated L-aureocin A53 W40E (Met¹-Leu⁵¹) using a 20-90% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 988.7 [M+6H]⁶⁺, 847.6 [M+7H]⁷⁺, 741.8 [M+8H]⁸⁺, 659.5 [M+9H]⁹⁺, 593.6 [M+10H]¹⁰⁺, observed 988.5 [M+6H]⁶⁺, 847.5 [M+7H]⁷⁺, 741.7 [M+8H]⁸⁺, 659.5 [M+9H]⁹⁺, 593.5 [M+10H]¹⁰⁺).

7.3.16. L-lacticin Q (Met¹-Trp²³) peptide hydrazide

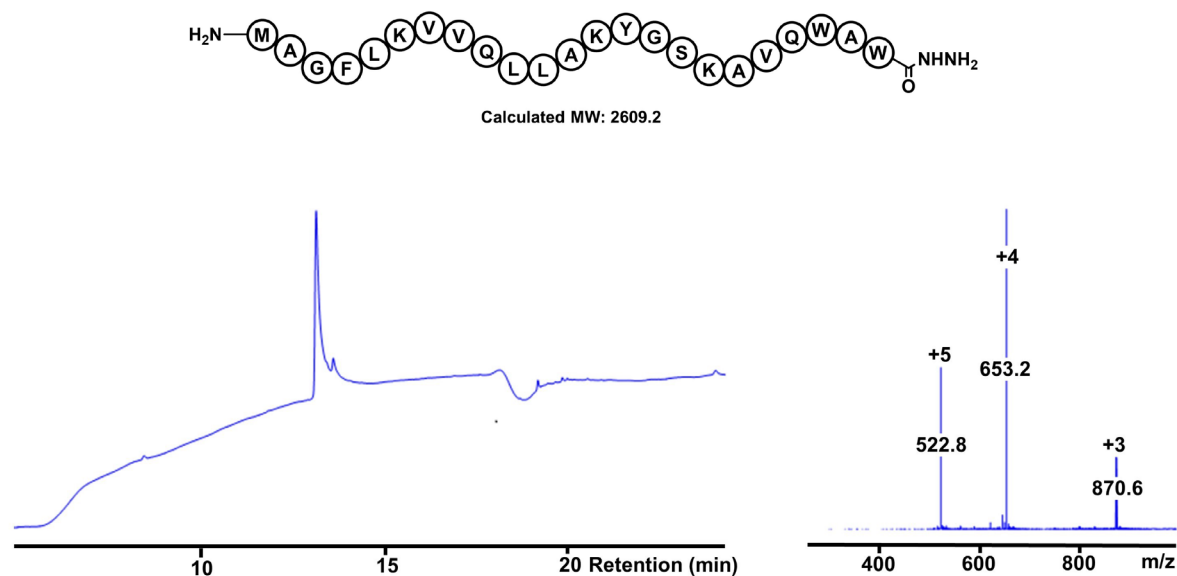


Figure S7.45: LC trace at UV210nm (left) of isolated L-lacticin Q (Met¹-Trp²³) peptide hydrazide using a 10-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 870.7 [M+3H]³⁺, 653.3 [M+4H]⁴⁺, 522.8 [M+5H]⁵⁺, observed 870.6 [M+3H]³⁺, 653.2 [M+4H]⁴⁺, 522.8 [M+5H]⁵⁺).

7.3.17. L-lactacin Q (Cys²⁴-Lys⁵³) N-cysteine peptide

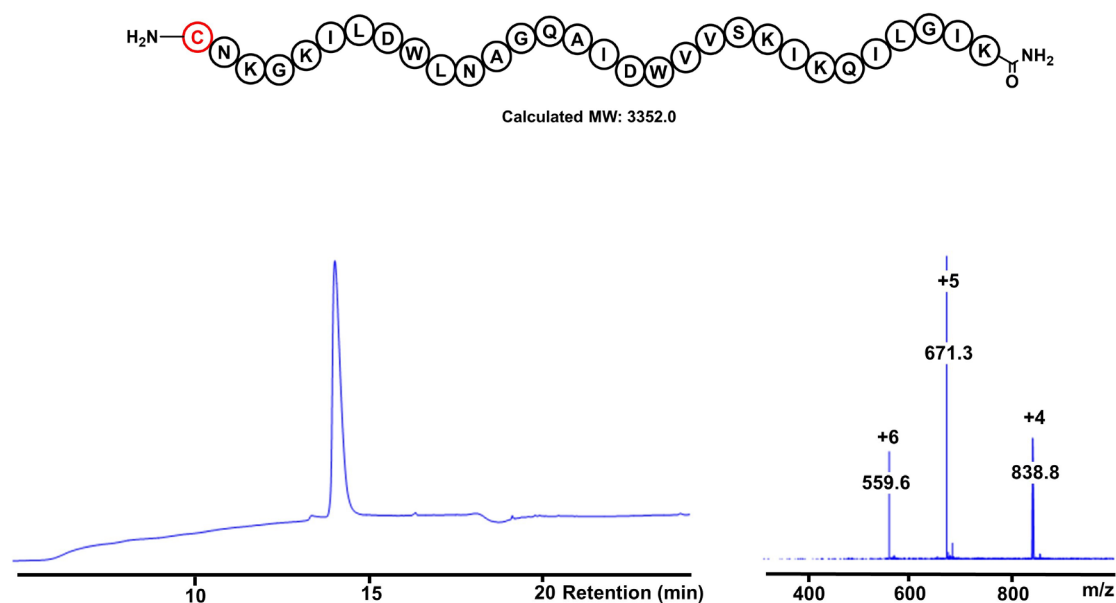


Figure S7.46: LC trace at UV210nm (left) of isolated L-lactacin Q (Cys²⁴-Lys⁵³) N-cysteine peptide using a 10-70% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 839.0 [M+4H]⁴⁺, 671.4 [M+5H]⁵⁺, 559.7 [M+6H]⁶⁺, observed 838.8 [M+4H]⁴⁺, 671.3 [M+5H]⁵⁺, 559.6 [M+6H]⁶⁺).

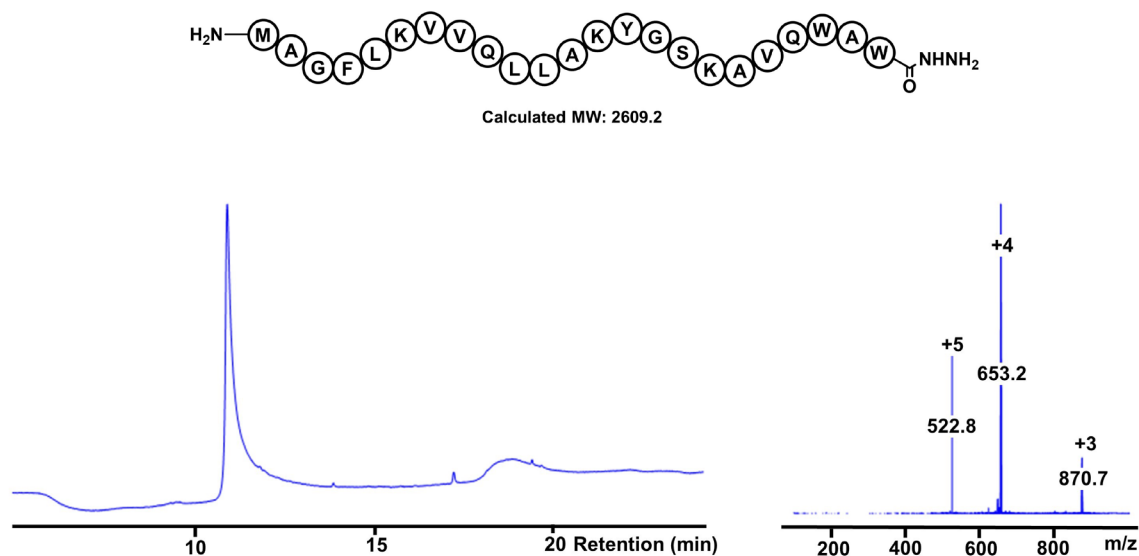
7.3.18. D-lactacin Q (Met¹-Trp²³) peptide hydrazide

Figure S7.47: LC trace at UV210nm (left) of isolated D-lactacin Q (Met¹-Trp²³) peptide hydrazide using a 20-80% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 870.7 [M+3H]³⁺, 653.3 [M+4H]⁴⁺, 522.8 [M+5H]⁵⁺, observed 870.7 [M+3H]³⁺, 653.2 [M+4H]⁴⁺, 522.8 [M+5H]⁵⁺).

7.3.19. D-lactacin Q (Cys²⁴-Lys⁵³) N-cysteine peptide

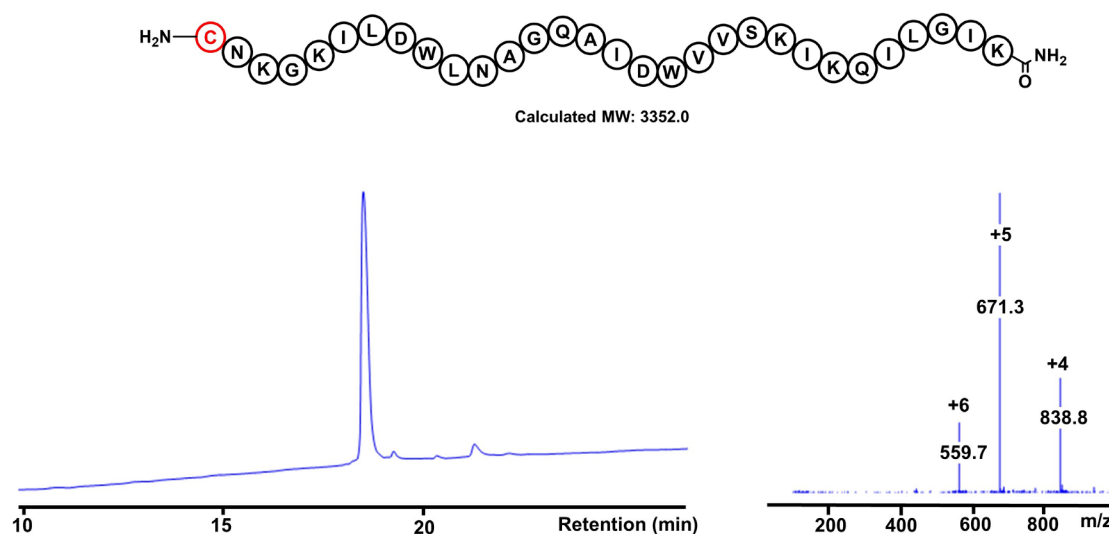


Figure S7.48: LC trace at UV210nm (left) of isolated D-lactacin Q (Cys²⁴-Lys⁵³) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (InfintiyLab poroshell, 4.6 mm x 100 mm, 120 Å, 4 µm). Right - (ESI-MS (m/z): calculated 839.0 [M+4H]⁴⁺, 671.4 [M+5H]⁵⁺, 559.7 [M+6H]⁶⁺, observed 838.8 [M+4H]⁴⁺, 671.3 [M+5H]⁵⁺, 559.7 [M+6H]⁶⁺).

7.3.20. Example LCMS ligation monitoring of lactacin Q

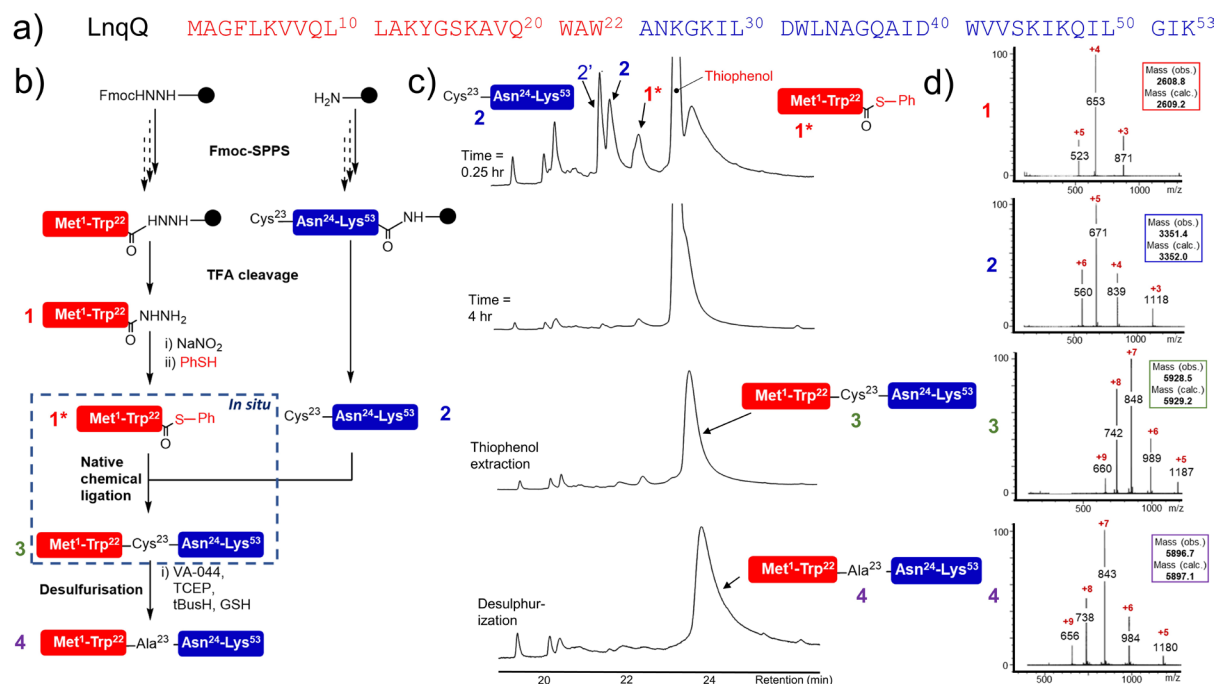


Figure S7.49: Example LCMS ligation scheme for L_{nk}Q. a) L_{nk}Q amino acid sequence, b) one-pot ligation-desulfurization scheme, c) HPLC reaction monitoring, and d) ESI+ MS for L_{nk}Q product, reactants and intermediates.

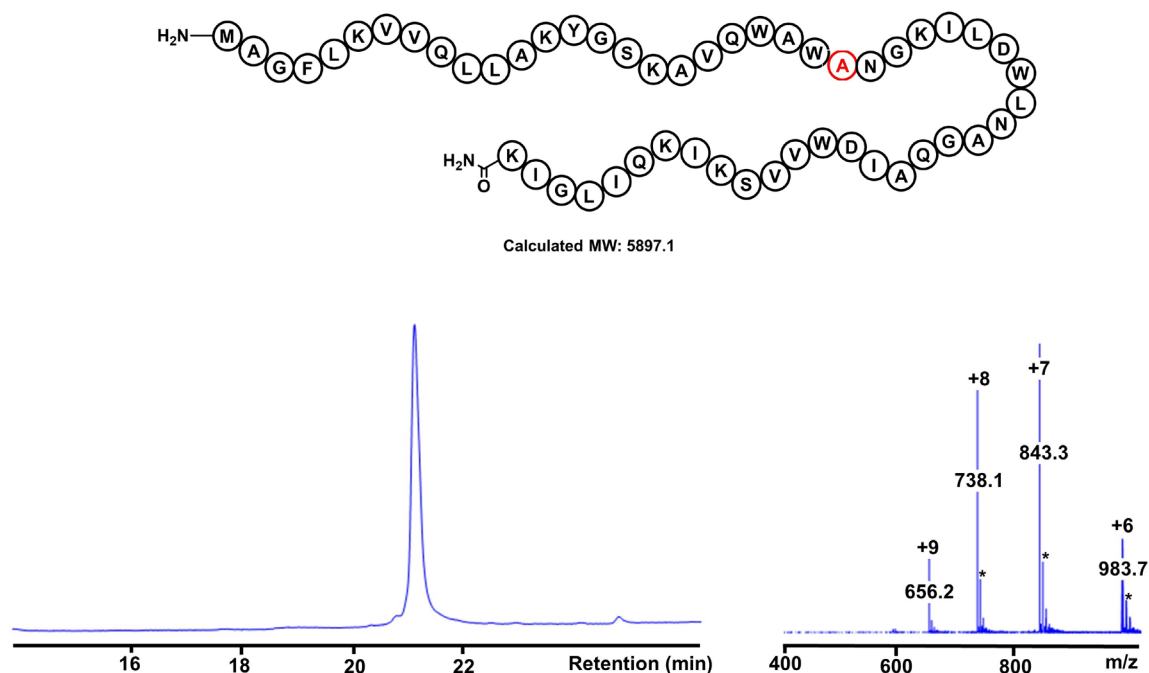
7.3.21. L-lactacin Q (Met¹-Lys⁵³)

Figure S7.50: LC trace at UV210nm (left) of isolated L-lactacin Q (Met¹-Lys⁵³) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated *990.0 [M+5H+K]⁶⁺, 983.8 [M+6H]⁶⁺, *848.7 [M+6H+K]⁷⁺, 843.4 [M+7H]⁷⁺, *742.8 [M+7H+K]⁸⁺, 738.1 [M+8H]⁸⁺, 656.2 [M+9H]⁹⁺, observed *990.2 [M+5H+K]⁶⁺, 983.7 [M+6H]⁶⁺, *848.7 [M+6H+K]⁷⁺, 843.3 [M+7H]⁷⁺, *742.8 [M+5H+K]⁶⁺, 738.1 [M+8H]⁸⁺, 656.2 [M+9H]⁹⁺).

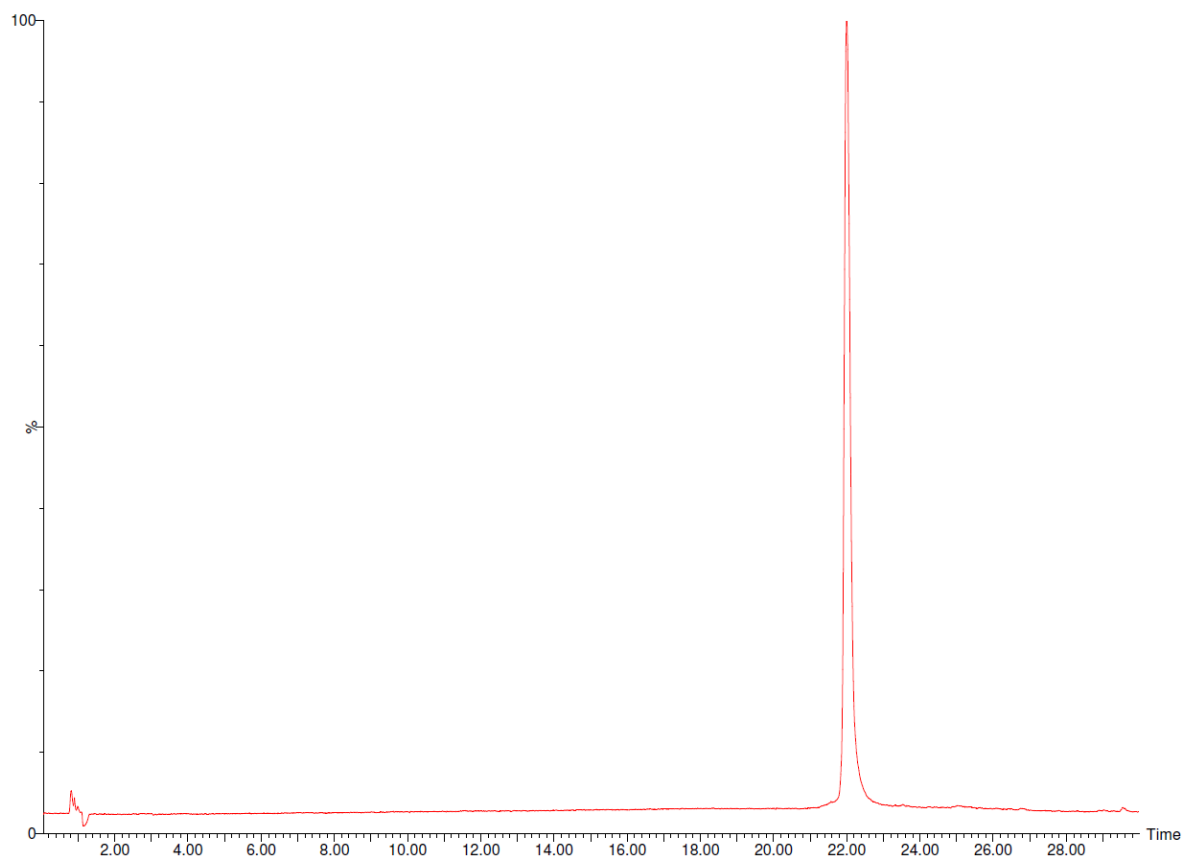


Figure S7.51: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q (Met¹-Lys⁵³) [22.30 min].

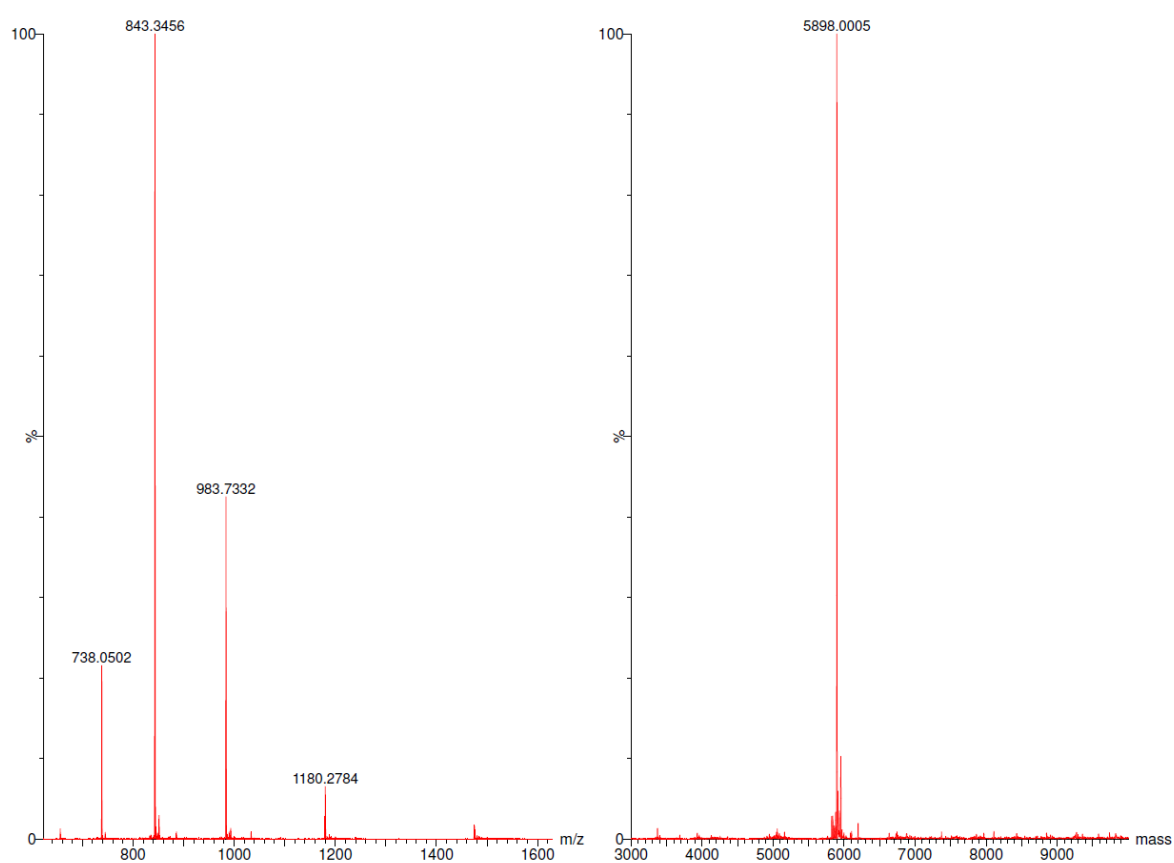


Figure S7.52: High-definition ESI+ mass spec of isolated L-lactacin Q (Met¹-Lys⁵³) (left), (ESI-MS (m/z): calculated 1180.4 [M+5H]⁵⁺, 983.8 [M+6H]⁶⁺, 843.4 [M+7H]⁷⁺, 738.1 [M+8H]⁸⁺, observed 1180.3 [M+5H]⁵⁺, 983.7 [M+6H]⁶⁺, 843.3 [M+7H]⁷⁺, 738.1 [M+8H]⁸⁺). (Right – deconvoluted mass), calculated 5897.1, observed 5898.0.

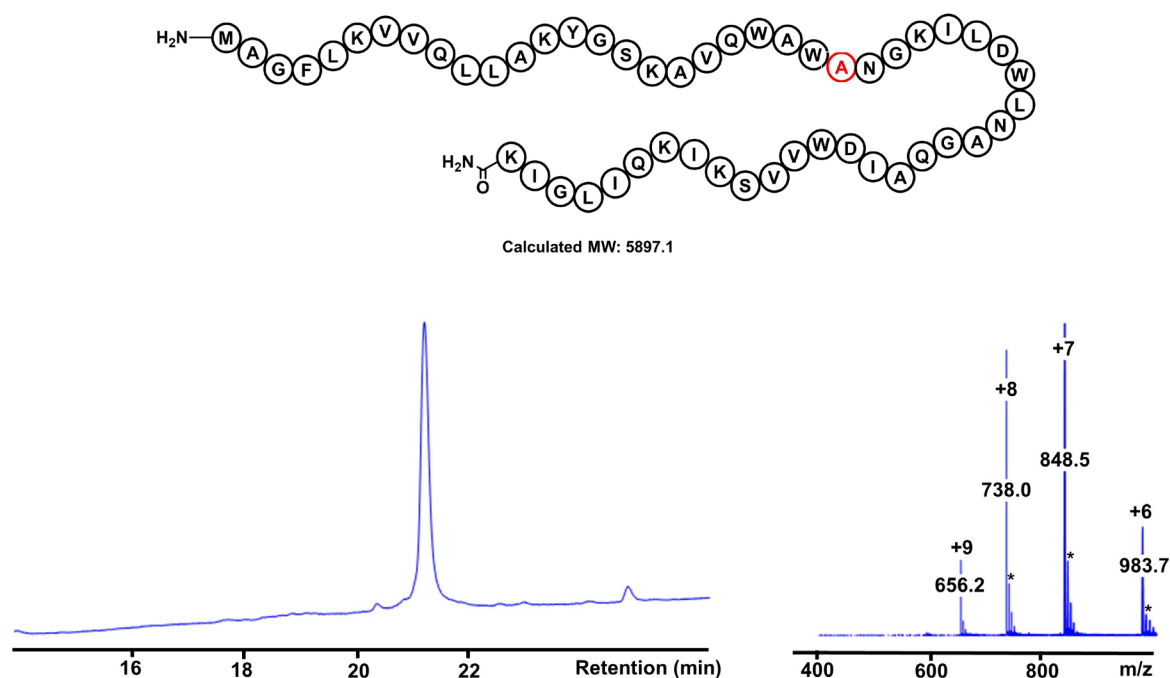
7.3.22. D-lactacin Q (Met¹-Lys⁵³)

Figure S7.53: LC trace at UV210nm (left) of isolated D-lactacin Q (Met¹-Lys⁵³) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated *990.0 [M+5H+K]⁶⁺, 983.8 [M+6H]⁶⁺, *848.7 [M+6H+K]⁷⁺, 843.4 [M+7H]⁷⁺, *742.8 [M+7H+K]⁸⁺, 738.1 [M+8H]⁸⁺, 656.2 [M+9H]⁹⁺, observed *990.0 [M+5H+K]⁶⁺, 983.7 [M+6H]⁶⁺, *848.8 [M+6H+K]⁷⁺, 843.5 [M+7H]⁷⁺, *742.6 [M+5H+K]⁶⁺, 738.0 [M+8H]⁸⁺, 656.2 [M+9H]⁹⁺).

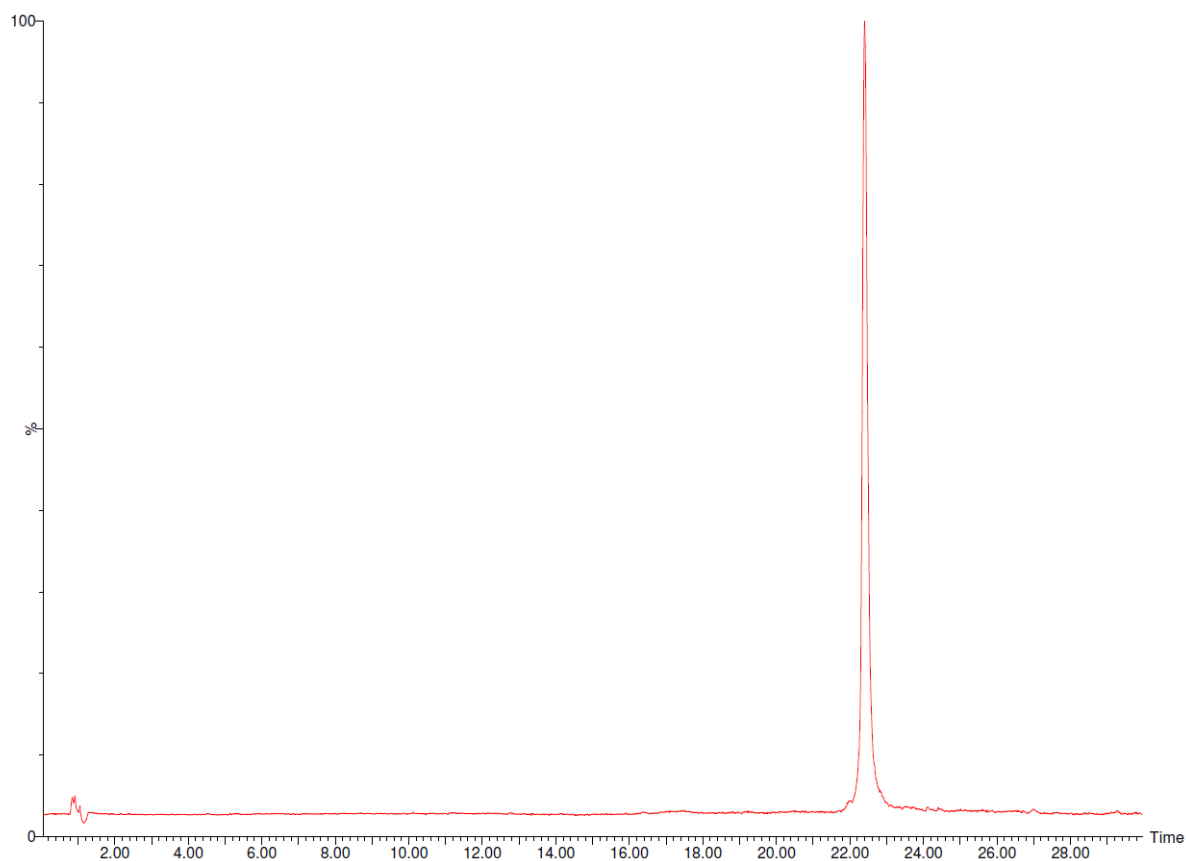


Figure S7.54: UPLC total ion chromatogram in ESI+ mode of isolated D-lactacin Q (Met¹-Lys⁵³) [22.40 min].

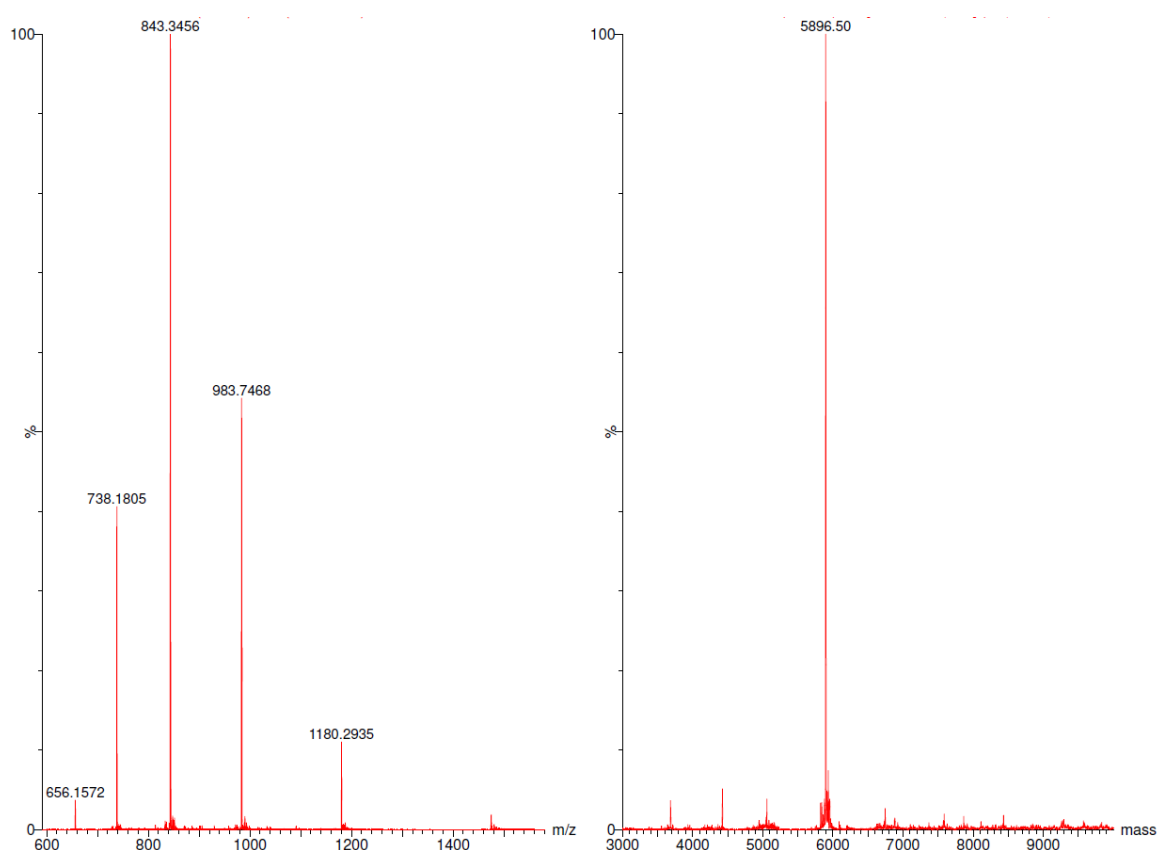


Figure S7.55: High-definition ESI+ mass spec of isolated D-lactacin Q (Met¹-Lys⁵³) (left), (ESI-MS (m/z): calculated 1180.4 [M+5H]⁵⁺, 983.8 [M+6H]⁶⁺, 843.4 [M+7H]⁷⁺, 738.1 [M+8H]⁸⁺, , 656.2 [M+9H]⁹⁺, observed 1180.3 [M+5H]⁵⁺, 983.7 [M+6H]⁶⁺, 843.3 [M+7H]⁷⁺, 738.2 [M+8H]⁸⁺, 656.2 [M+9H]⁹⁺). (Right – deconvoluted mass), calculated 5897.1, observed 5896.5.

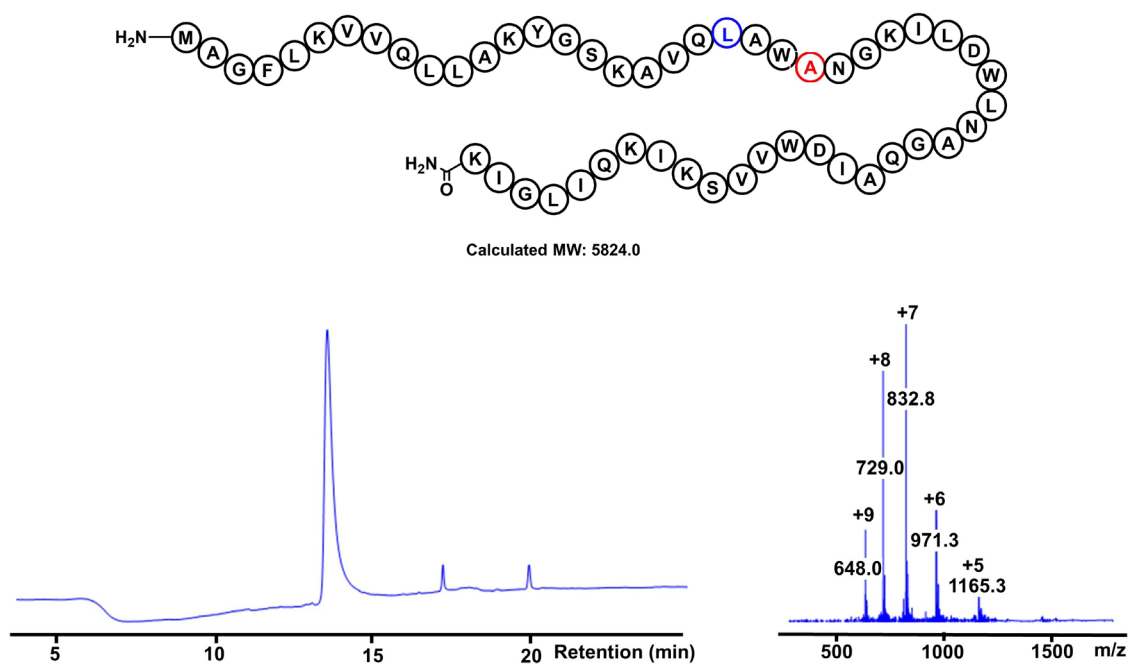
7.3.23. L-lactacin Q W21L (Met¹-Lys⁵³)

Figure S7.56: LC trace at UV210nm (left) of isolated L-lactacin Q W21L (Met¹-Lys⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.3 [M+5H]⁵⁺, 971.3 [M+6H]⁶⁺, 832.8 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.0 [M+9H]⁹⁺).

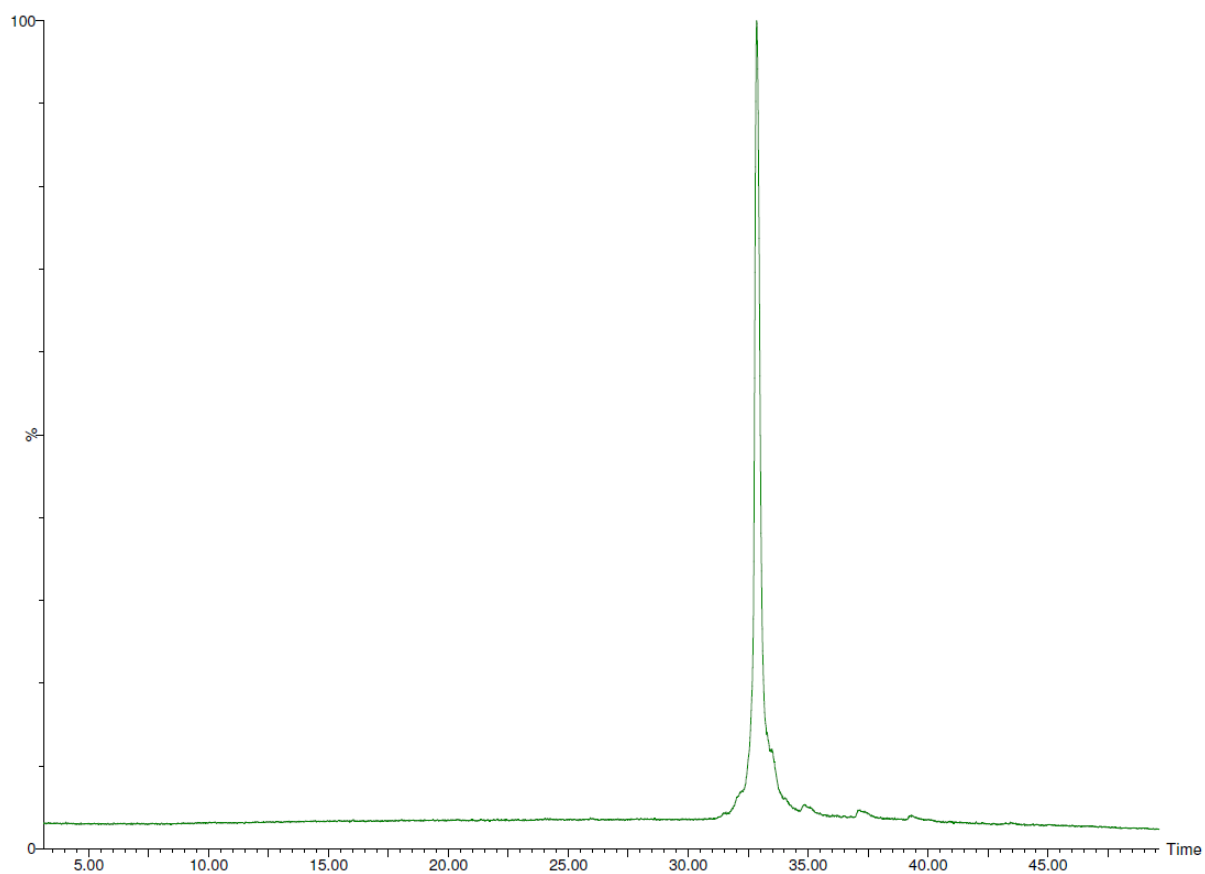


Figure S7.57: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W21L (Met¹-Lys⁵³) [32.85 min].

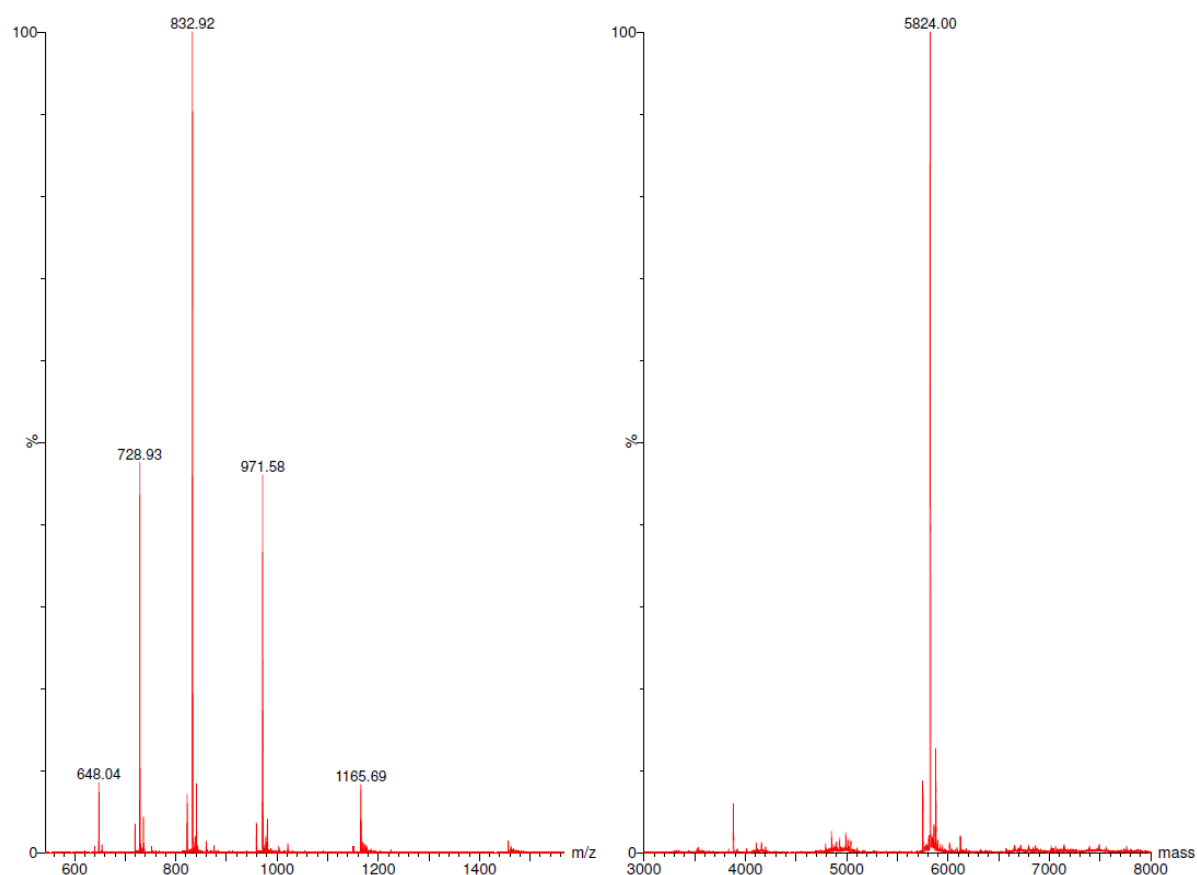


Figure S7.58: High-definition ESI+ mass spec of isolated L-lactacin Q W21L (Met¹-Lys⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.7 [M+5H]⁵⁺, 971.6 [M+6H]⁶⁺, 832.9 [M+7H]⁷⁺, 728.9 [M+8H]⁸⁺, 648.0 [M+9H]⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5824.0.

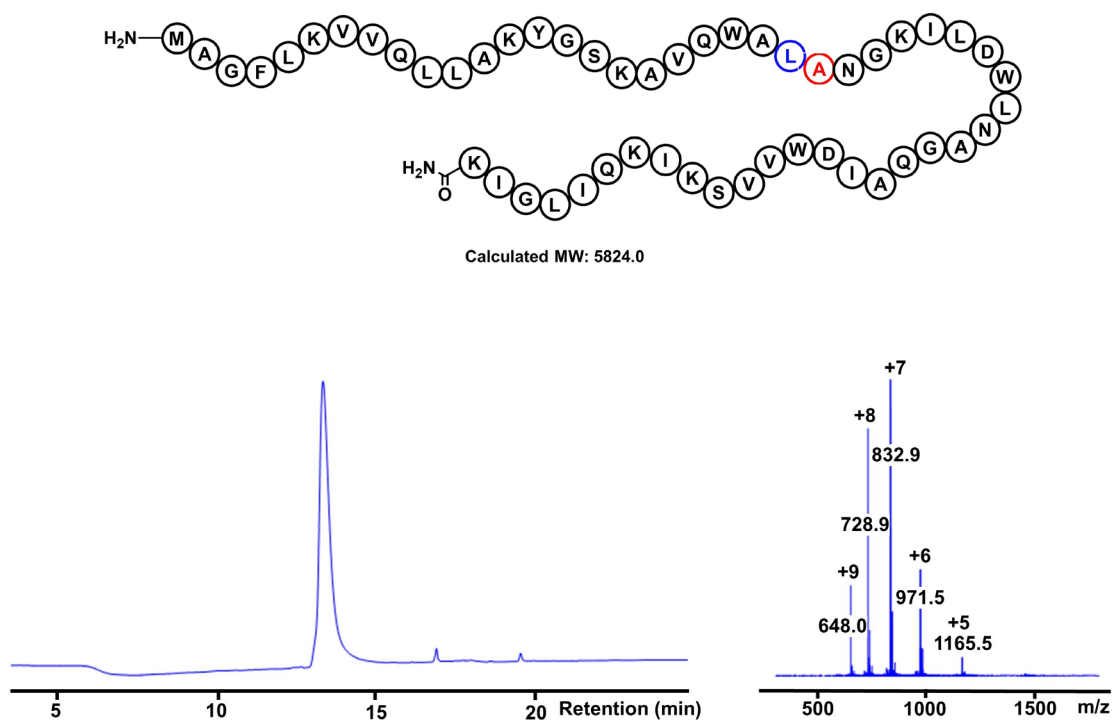
7.3.24. L-lactacin Q W23L (Met¹-Lys⁵³)

Figure S7.59: LC trace at UV210nm (left) of isolated L-lactacin Q W23L (Met¹-Lys⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.5 [M+5H]⁵⁺, 971.5 [M+6H]⁶⁺, 832.9 [M+7H]⁷⁺, 728.9 [M+8H]⁸⁺, 648.0 [M+9H]⁹⁺).

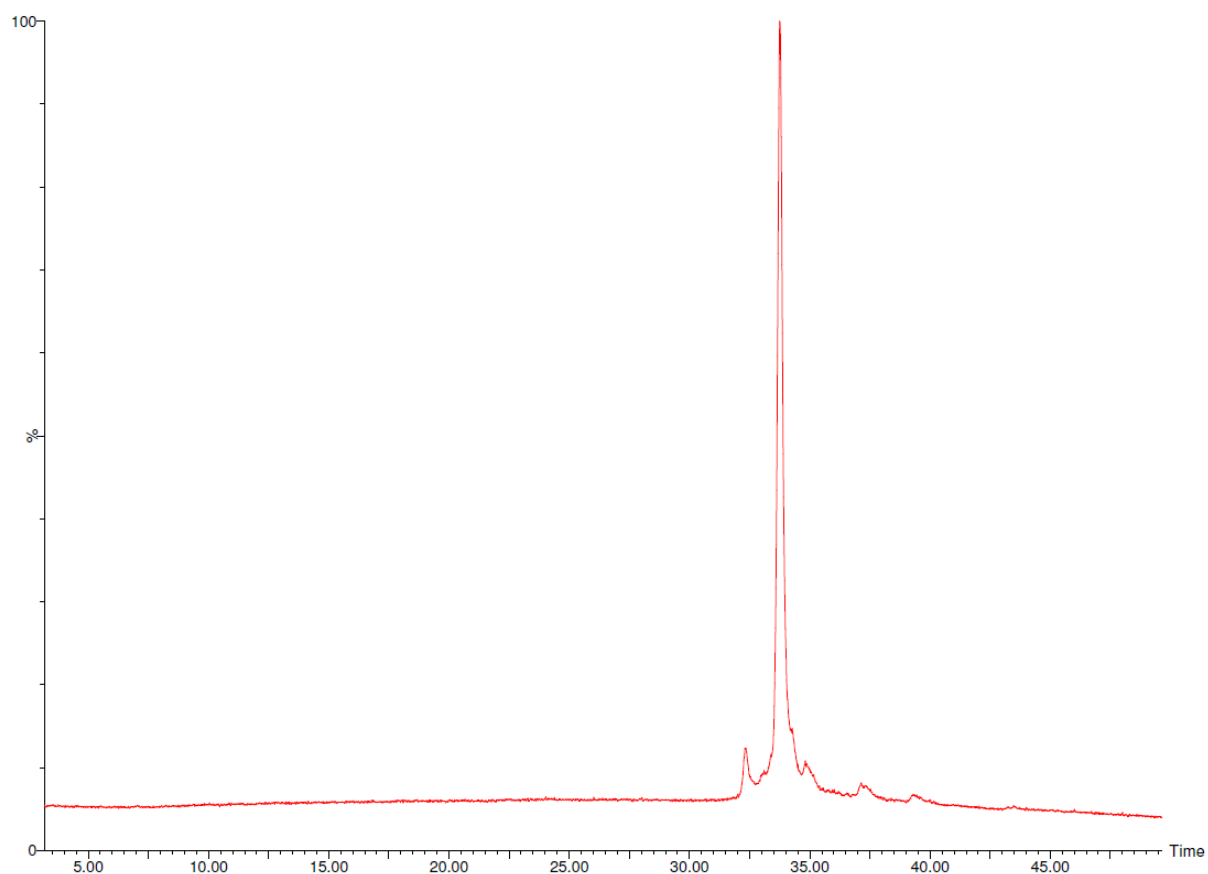


Figure S7.60: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W23L (Met¹-Lys⁵³) [32.75 min].

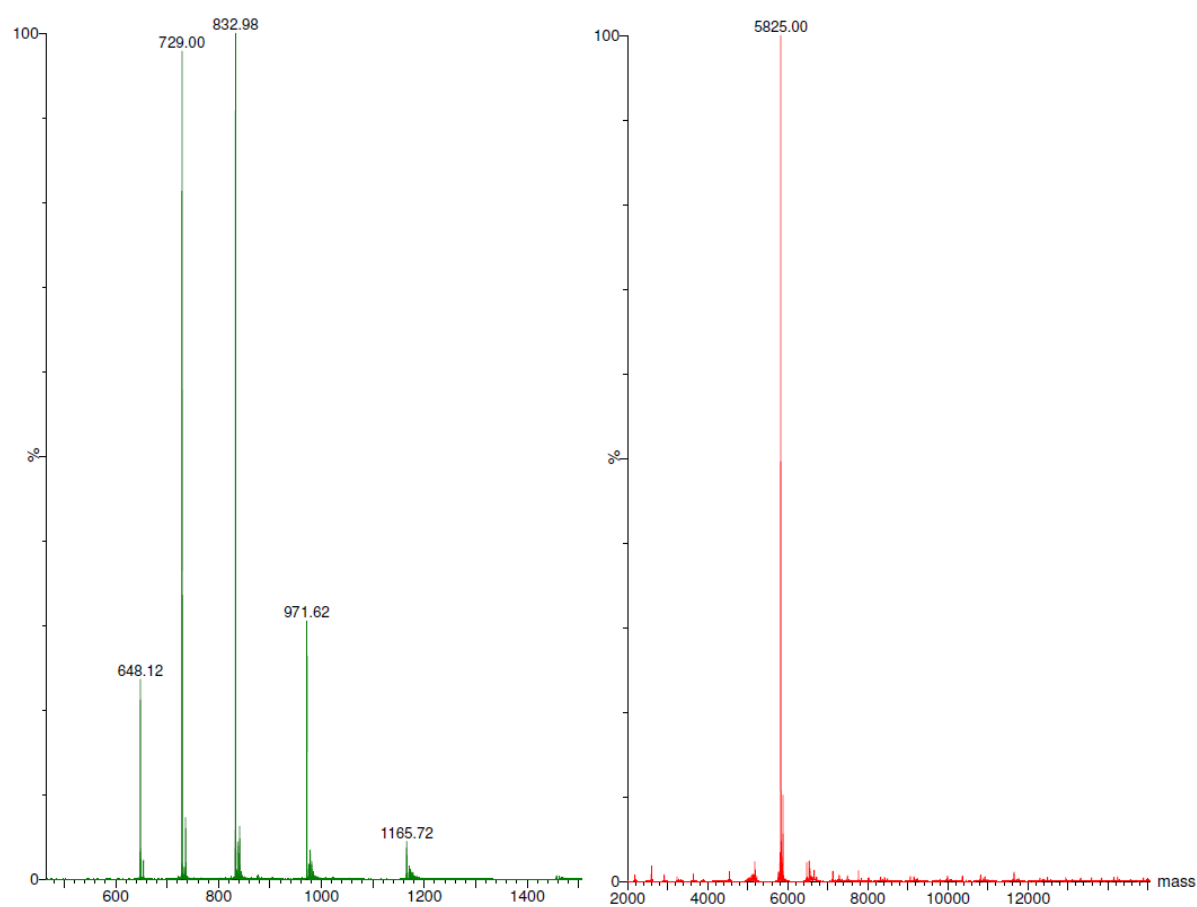


Figure S7.61: High-definition ESI+ mass spec of isolated L-lactacin Q W23L (Met¹-Lys⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.7 [M+5H]⁵⁺, 971.6 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5825.0.

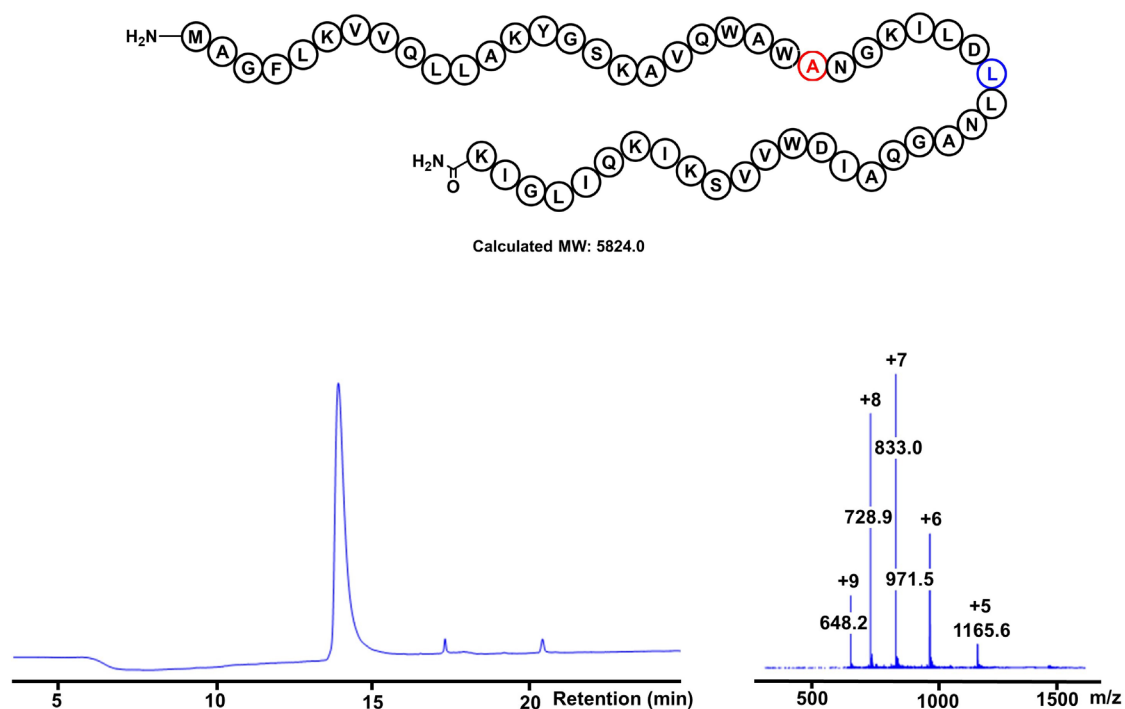
7.3.25. L-lacticin Q W32L (Met¹-Lys⁵³)

Figure S7.62: LC trace at UV210nm (left) of isolated L-lacticin Q W32L (Met¹-Lys⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.6 [M+5H]⁵⁺, 971.5 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 728.9 [M+8H]⁸⁺, 648.2 [M+9H]⁹⁺).

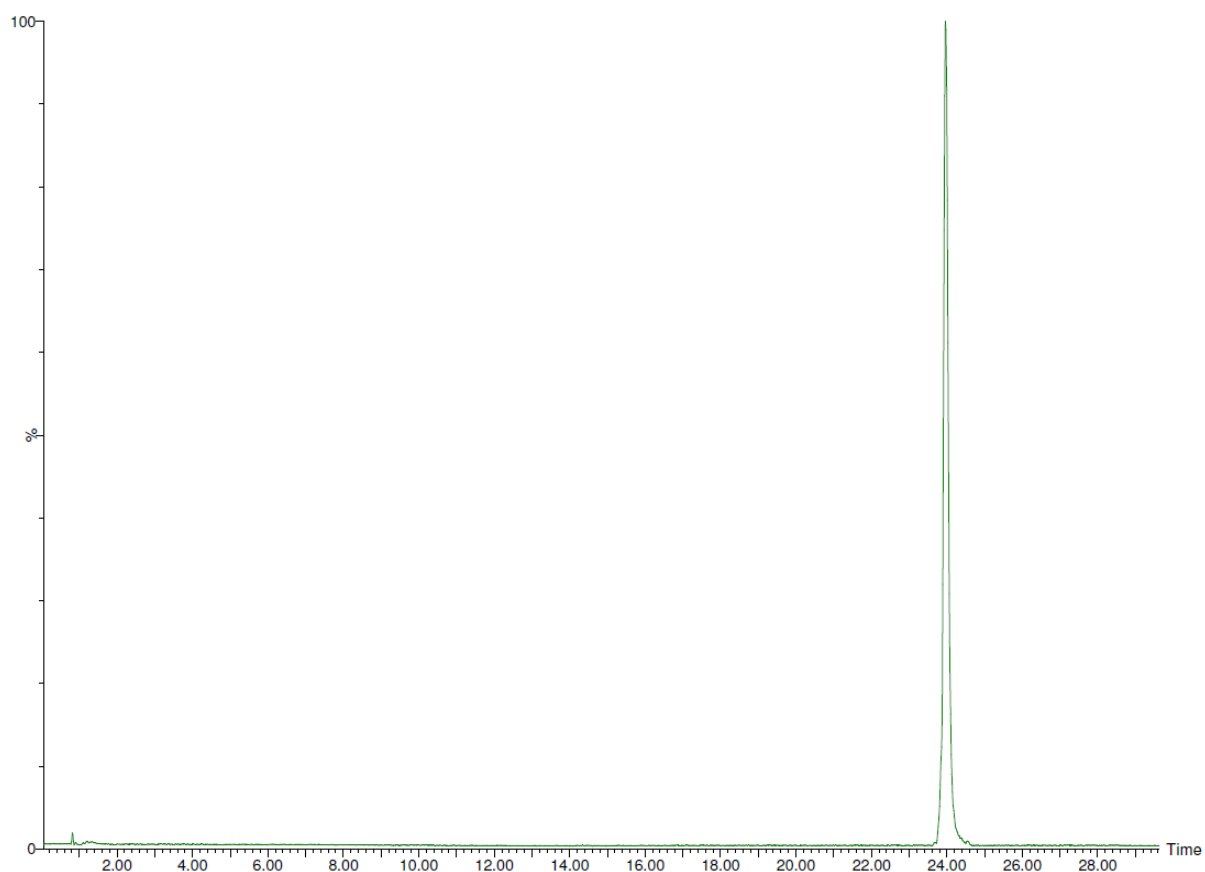


Figure S7.63: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W32L (Met¹-Lys⁵³) [23.96 min] using a 5-95% gradient of A/B over 30 minutes.

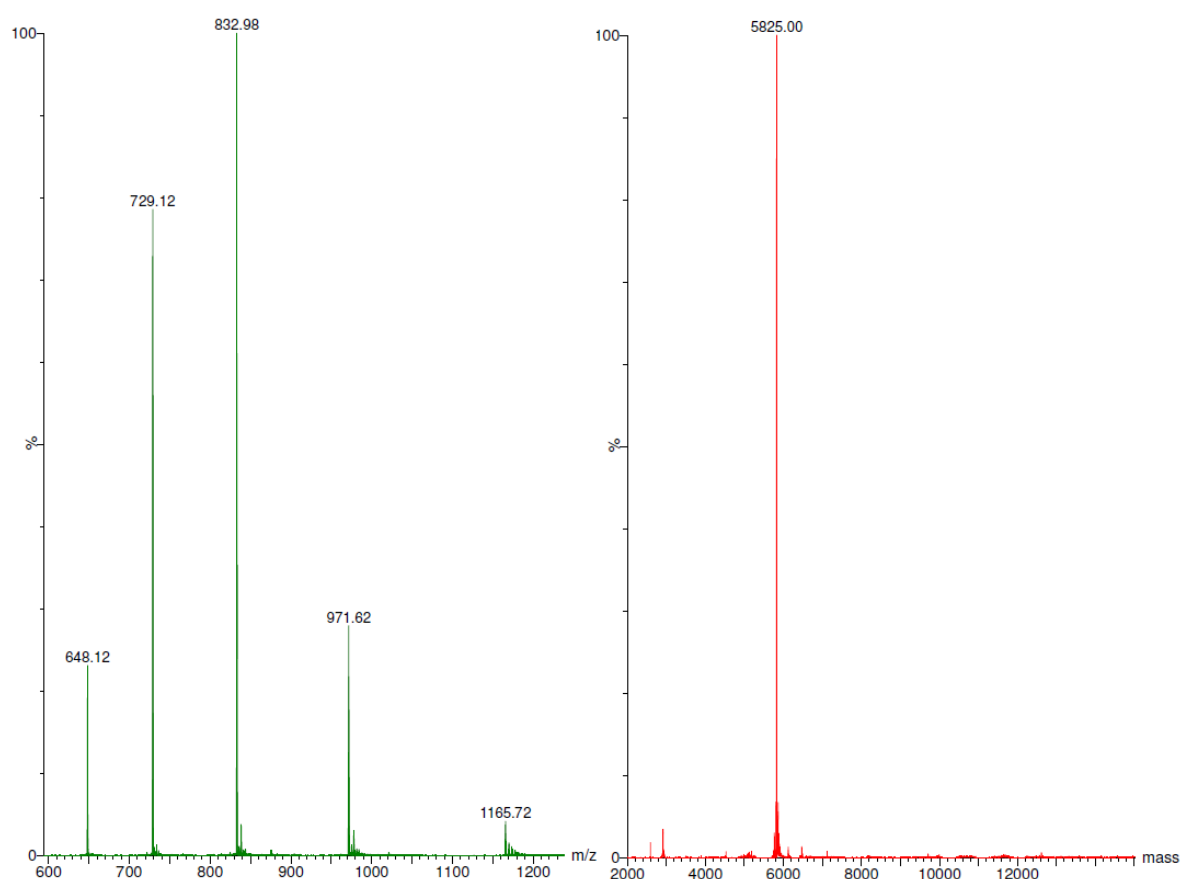


Figure S7.64: High-definition ESI+ mass spec of isolated L-lactacin Q W32L (Met¹-Lys⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.7 [M+5H]⁵⁺, 971.6 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.1 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5825.0.

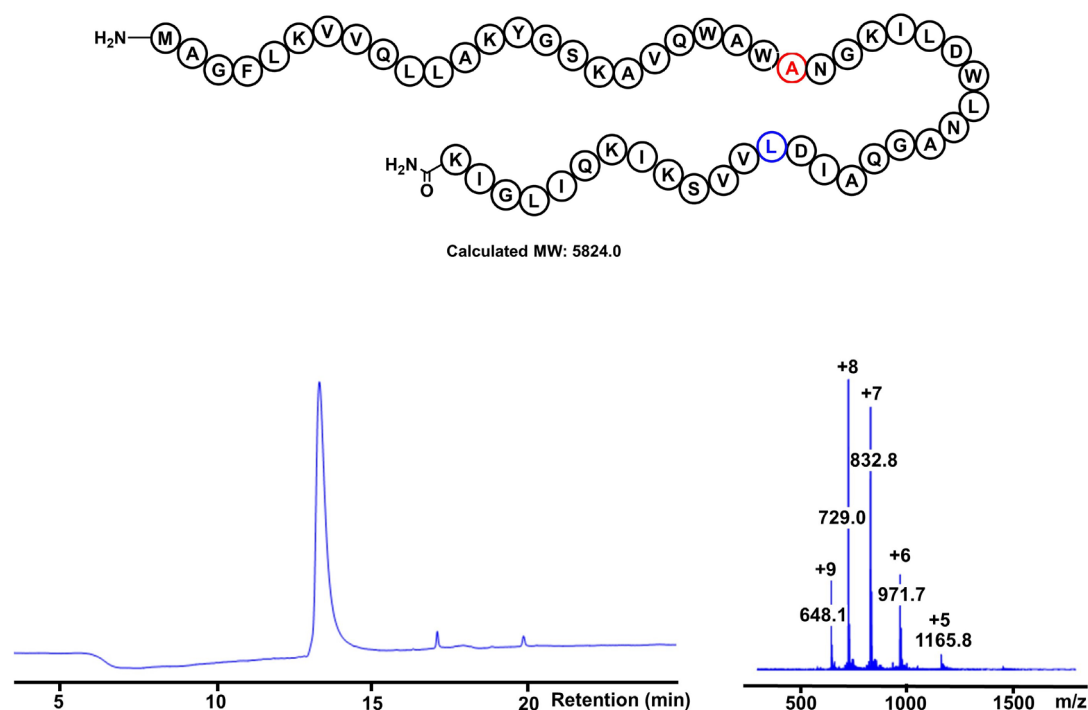
7.3.26. L-lactacin Q W41L (Met¹-Lys⁵³)

Figure S7.65: LC trace at UV210nm (left) of isolated L-lactacin Q W41L (Met¹-Lys⁵³) using a 20-99% gradient of A/B over 30 minute. Right - (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 832.8 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺).

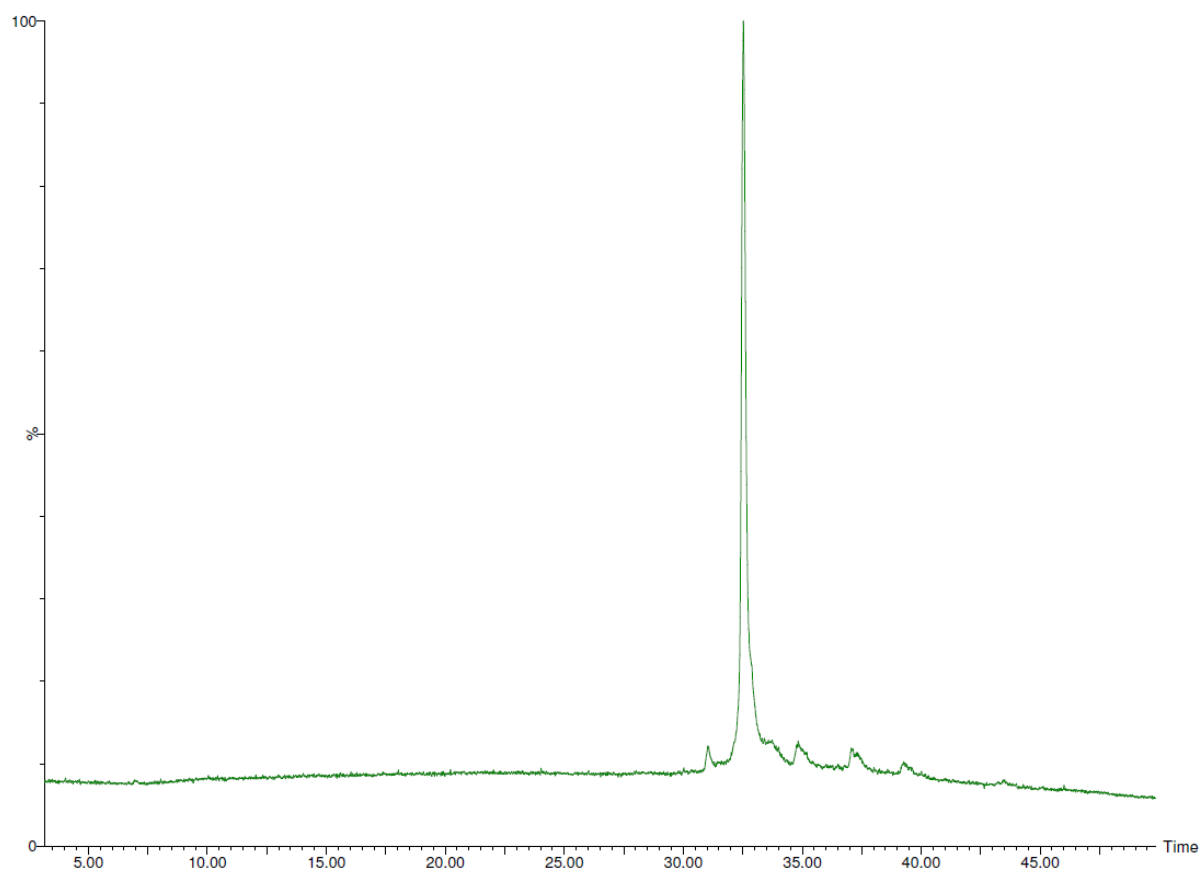


Figure S7.66: UPLC total ion chromatogram in ESI+ mode of isolated L-lactacin Q W41L (Met¹-Lys⁵³) [32.53 min].

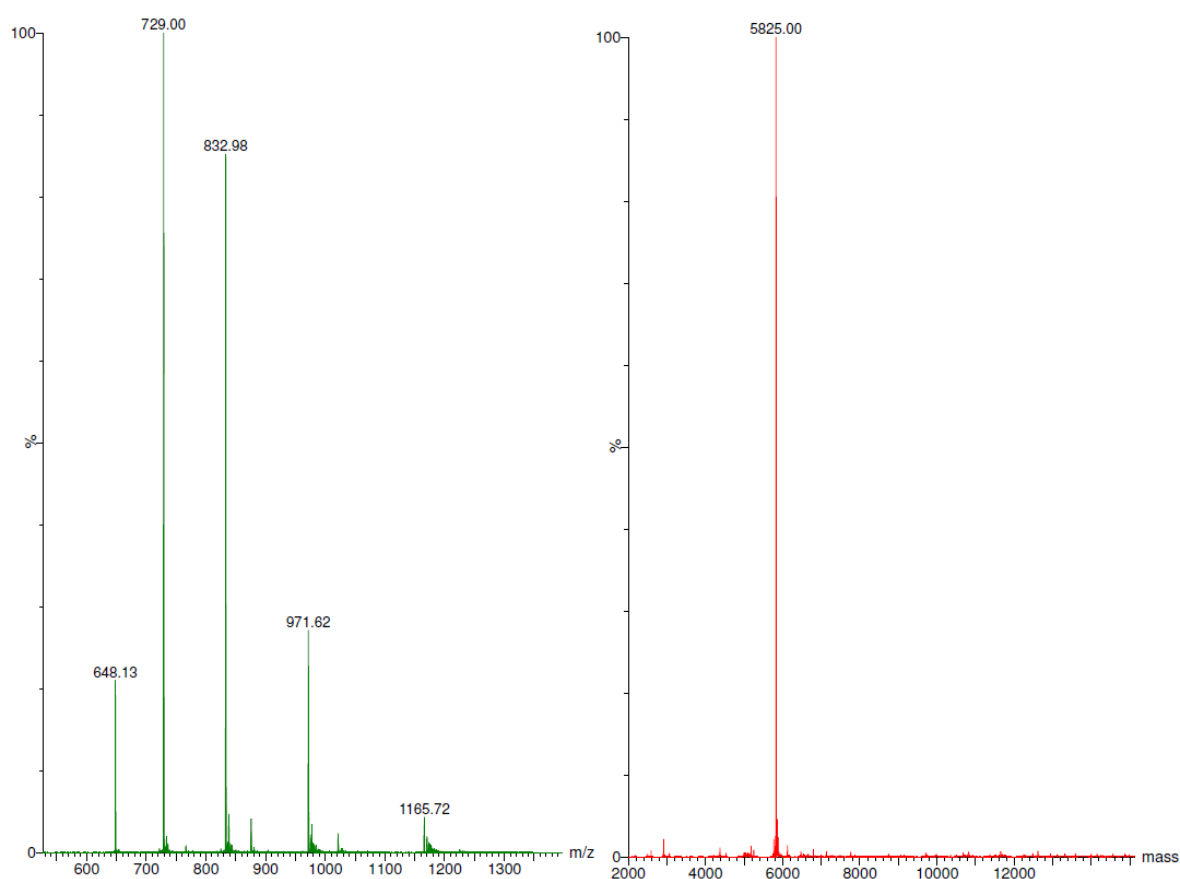
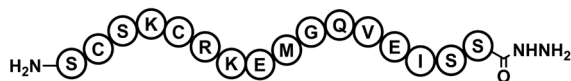


Figure S7.67: High-definition ESI+ mass spec of isolated L-lactacin Q W41L (Met¹-Lys⁵³) (left), (ESI-MS (m/z): calculated 1165.8 [M+5H]⁵⁺, 971.7 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺, observed 1165.7 [M+5H]⁵⁺, 971.6 [M+6H]⁶⁺, 833.0 [M+7H]⁷⁺, 729.0 [M+8H]⁸⁺, 648.1 [M+9H]⁹⁺). (Right – deconvoluted mass), calculated 5824.0, observed 5825.0.

7.3.27. L-TNRCD2 (Ser⁷²-Ser⁸⁷) peptide hydrazide



Calculated MW: 1788.0

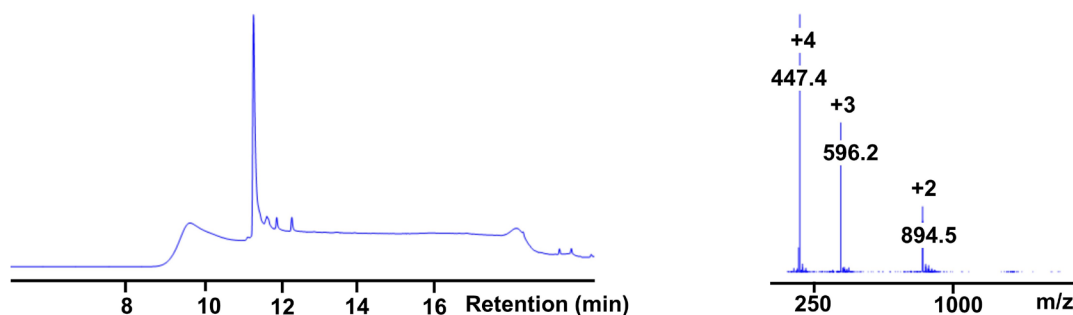
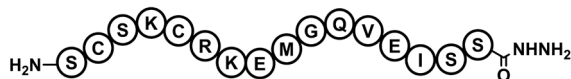


Figure S7.68: LC trace at UV210nm (left) of isolated L-TNRCD2 (Ser⁷²-Ser⁸⁷) peptide hydrazide using a 0-60% gradient of A/B over 30 minutes. Right - (ESI-MS (m/z): calculated 895.0 [M+2H]²⁺, 597.0 [M+3H]³⁺, 448.0 [M+4H]⁴⁺, 894.5 [M+2H]²⁺, 596.2 [M+3H]³⁺, 447.4 [M+4H]⁴⁺).

7.3.28. D-TNRCD2 (Ser⁷²-Lys⁸⁷) peptide hydrazide



Calculated MW: 1788.0

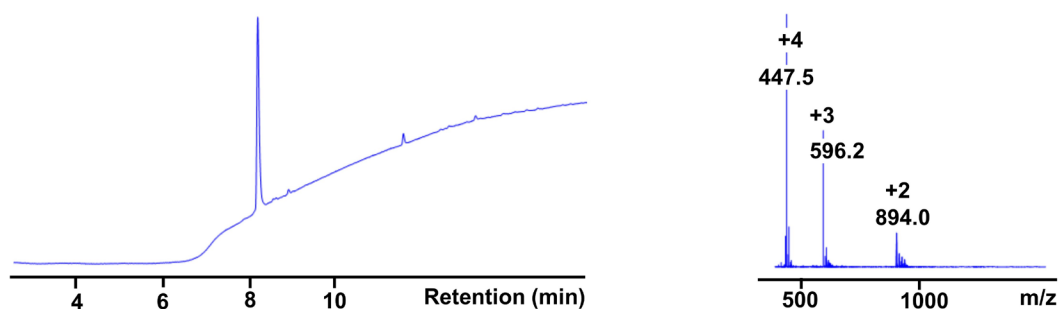


Figure S7.69: LC trace at UV210nm (left) of isolated L-TNRCD2 (Ser⁷²-Ser⁸⁷) peptide hydrazide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 895.0 [M+2H]²⁺, 597.0 [M+3H]³⁺, 448.0 [M+4H]⁴⁺, 894.5 [M+2H]²⁺, 596.2 [M+3H]³⁺, 447.4 [M+4H]⁴⁺).

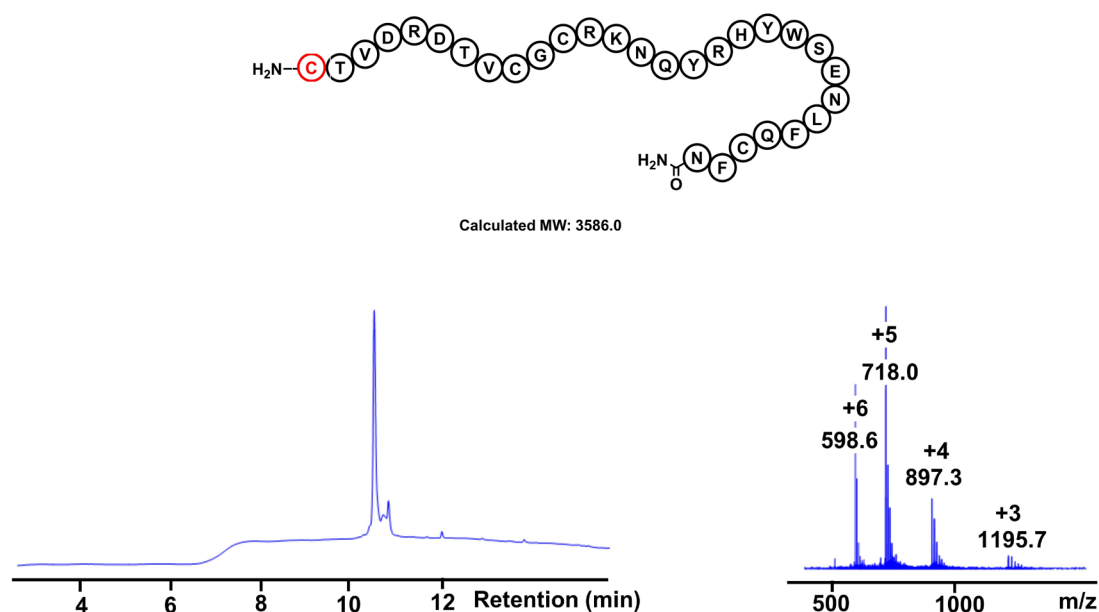
7.3.29. L-TNRCD2 (Cys⁸⁸-Asn¹¹⁶) N-cysteine peptide

Figure S7.70: LC trace at UV210nm (left) of isolated L-TNRCD2 (Cys⁸⁸-Asn¹¹⁶) N-cysteine peptide using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1196.3 [M+3H]³⁺, 897.5 [M+4H]⁴⁺, 718.2 [M+5H]⁵⁺, 598.7 [M+6H]⁶⁺, observed 1195.7 [M+3H]³⁺, 897.3 [M+4H]⁴⁺, 718.0 [M+5H]⁵⁺, 598.6 [M+6H]⁶⁺).

7.3.30. D-TNRCD2 (Cys⁸⁸-Asn¹¹⁶) N-cysteine peptide

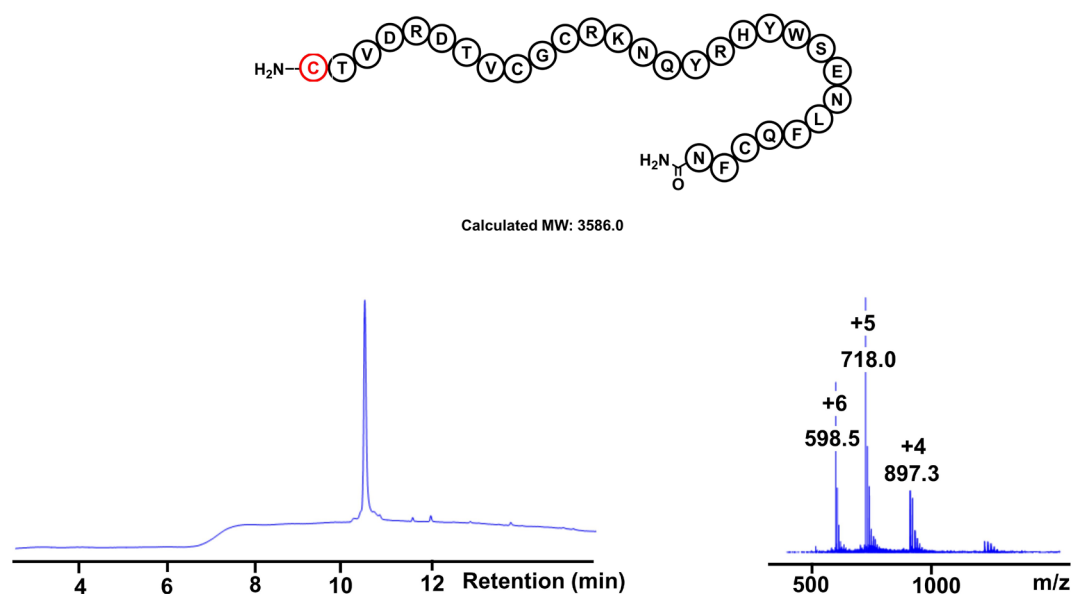


Figure S7.71: LC trace at UV210nm (left) of isolated D-TNRCD2 (Cys⁸⁸-Asn¹¹⁶) N-cysteine peptide using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 897.5 [M+4H]⁴⁺, 718.2 [M+5H]⁵⁺, 598.7 [M+6H]⁶⁺, observed 897.3 [M+4H]⁴⁺, 718.0 [M+5H]⁵⁺, 598.5 [M+6H]⁶⁺).

7.3.31. D-TNRCD2 (Cys⁸⁸-Asn¹¹⁶-Gly¹¹⁷-Gly¹²¹-Lys(biotin)¹²²) N-cysteine peptide

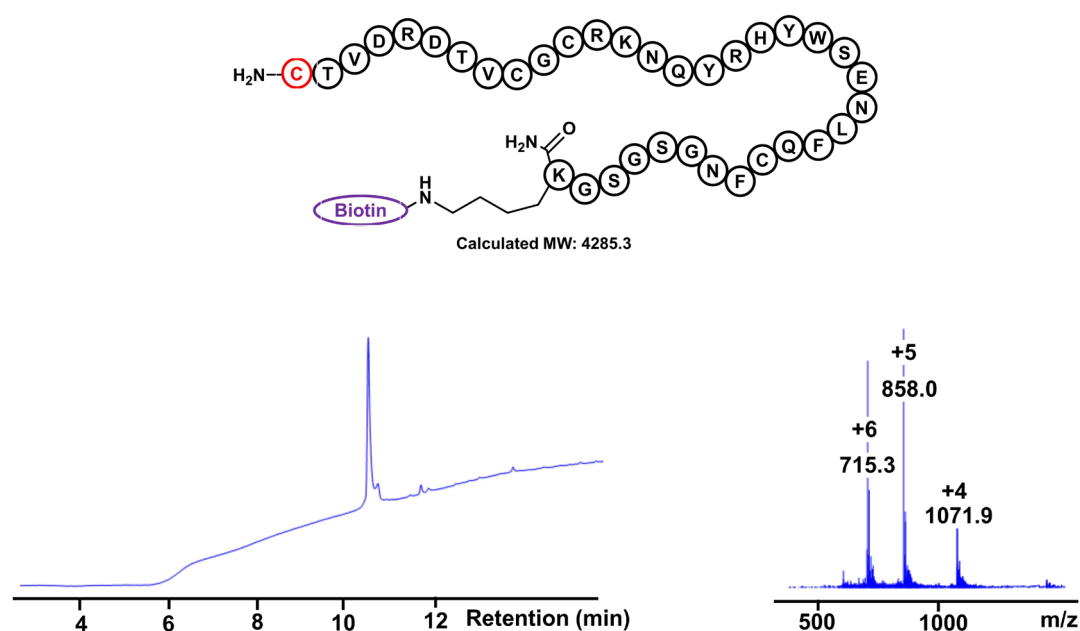


Figure S7.72: LC trace at UV210nm (left) of isolated D-TNRCD2 (Cys⁸⁸-Asn¹¹⁶-Gly¹¹⁷-Gly¹²¹-Lys(biotin)¹²²) N-cysteine peptide using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1072.3 [M+4H]⁴⁺, 858.1 [M+5H]⁵⁺, 715.2 [M+6H]⁶⁺, observed 1071.9 [M+4H]⁴⁺, 858.0 [M+5H]⁵⁺, 715.3 [M+6H]⁶⁺).

7.3.32. L-TNRCD2 (Ser⁷²-Asn¹¹⁶) reduced (not isolated)

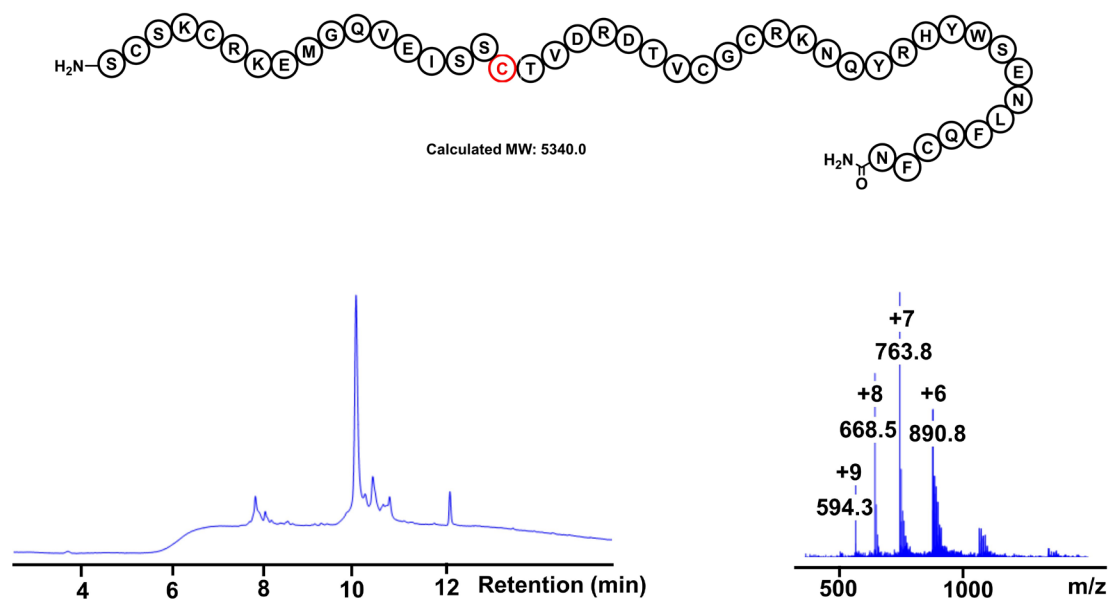


Figure S7.73: LC trace at UV210nm (left) of reduced L-TNRCD2 (Ser⁷²-Asn¹¹⁶) following native chemical ligation (without isolation). Right - (ESI-MS (m/z): calculated 897.5 [M+4H]⁴⁺, 718.2 [M+5H]⁵⁺, 598.7 [M+6H]⁶⁺, observed 897.3 [M+4H]⁴⁺, 718.0 [M+5H]⁵⁺, 598.5 [M+6H]⁶⁺).

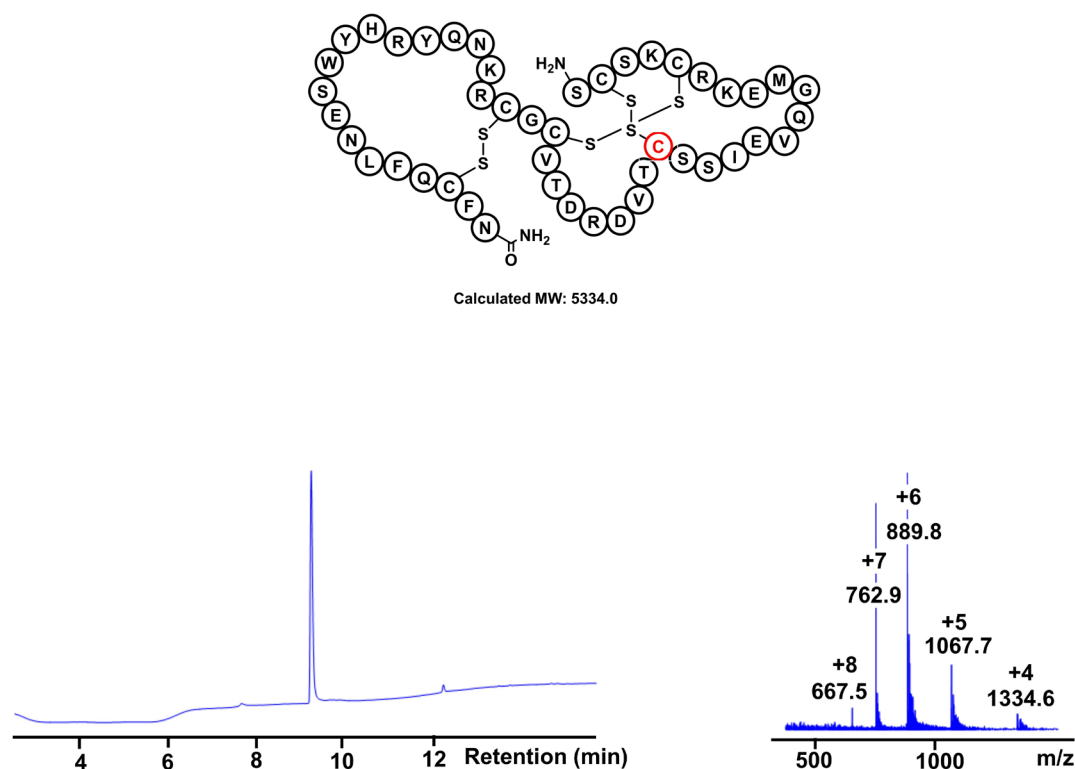
7.3.33. L-TNRCD2 (Ser⁷²-Asn¹¹⁶)

Figure S7.74: LC trace at UV210nm (left) of isolated L-TNRCD2 (Ser⁷²-Asn¹¹⁶). Right - (ESI-MS (m/z): calculated 1334.5 [M+4H]⁴⁺, 1067.8 [M+5H]⁵⁺, 890.0 [M+6H]⁶⁺, 763.0 [M+7H]⁷⁺, 667.7 [M+8H]⁸⁺, observed 1334.6 [M+4H]⁴⁺, 1067.7 [M+5H]⁵⁺, 889.8 [M+6H]⁶⁺, 762.9 [M+7H]⁷⁺, 667.5 [M+8H]⁸⁺).

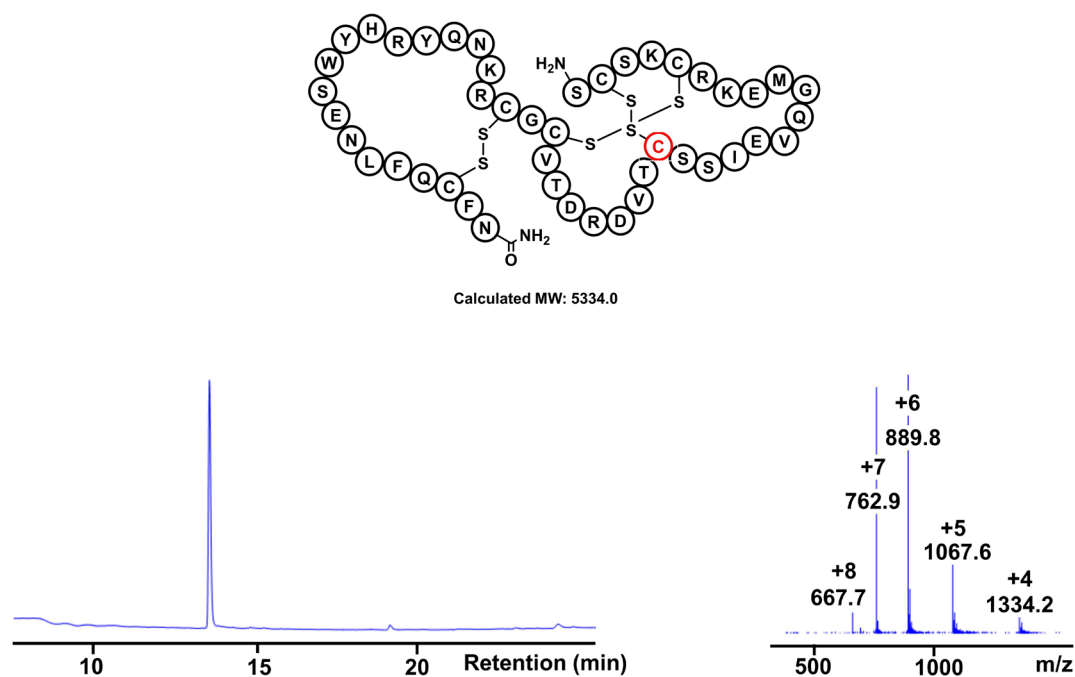
7.3.34. D-TNRCD2 (Ser⁷²-Asn¹¹⁶)

Figure S7.75: LC trace at UV210nm (left) of isolated D-TNRCD2 (Ser⁷²-Asn¹¹⁶) using a 5-70% gradient of A/B over 30 minutes on a RP-C4 column (ACE, 4.6 mm x 250 mm, 300 Å, 5 µm). Right - (ESI-MS (m/z): calculated 1334.5 [M+4H]⁴⁺, 1067.8 [M+5H]⁵⁺, 890.0 [M+6H]⁶⁺, 763.0 [M+7H]⁷⁺, 667.7 [M+8H]⁸⁺, observed 1334.2 [M+4H]⁴⁺, 1067.6 [M+5H]⁵⁺, 889.8 [M+6H]⁶⁺, 762.9 [M+7H]⁷⁺, 667.7 [M+8H]⁸⁺).

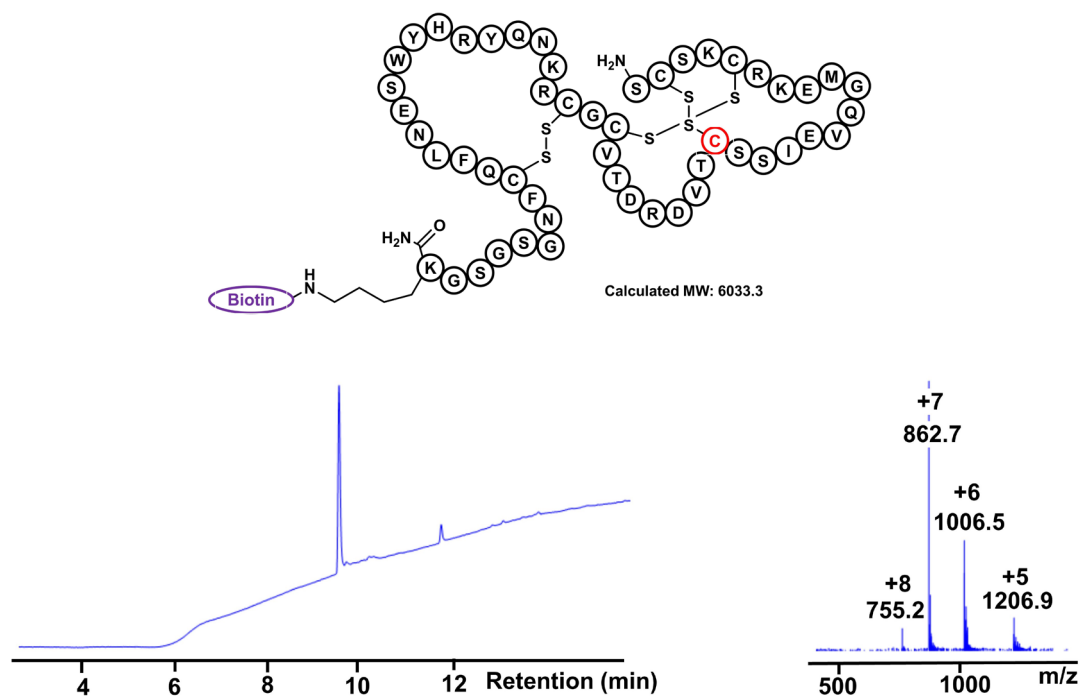
7.3.35. D-TNRCD2 (Ser⁷²-Asn¹¹⁶-Gly¹¹⁷-Gly¹²¹-Lys(biotin)¹²²)

Figure S7.76: LC trace at UV210nm (left) of isolated D-TNRCD2 (Ser⁷²-Asn¹¹⁶-Gly¹¹⁷-Gly¹²¹-Lys(biotin)¹²²) using a 5-70% gradient of A/B over 30 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1207.7 [M+5H]⁵⁺, 1006.6 [M+6H]⁺⁶, 862.9 [M+7H]⁺⁷, 755.2 [M+8H]⁺⁸, observed 1207.9 [M+5H]⁵⁺, 1006.5 [M+6H]⁺⁶, 862.7 [M+7H]⁺⁷, 755.2 [M+8H]⁺⁸).

7.3.36. L-TCPB-E-C1

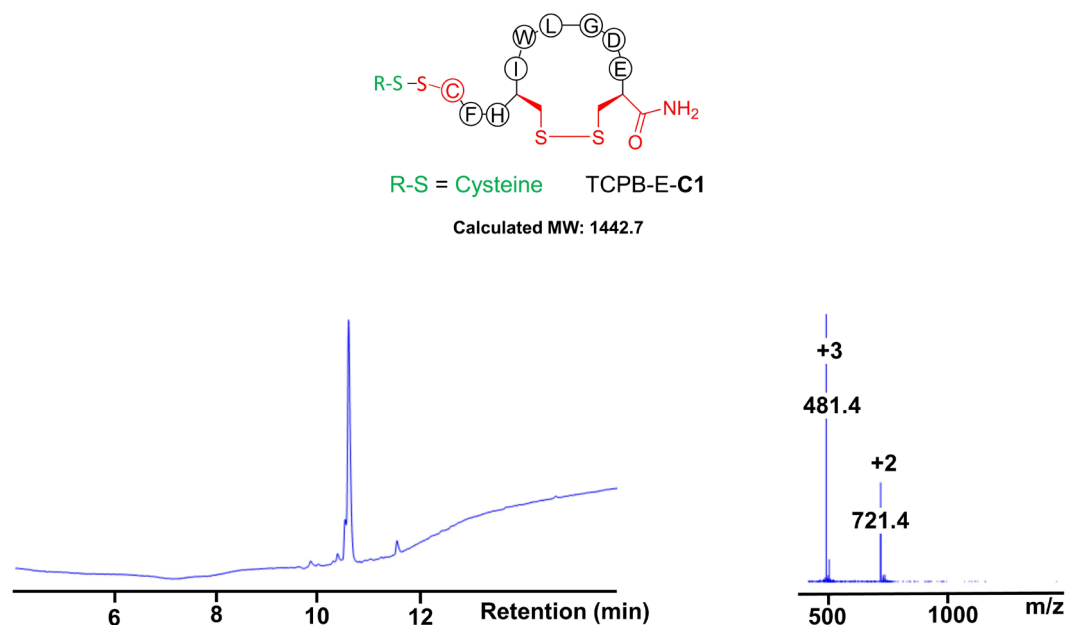


Figure S7.77: LC trace at UV210nm (left) of L-TCPB-E-C1 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 $[M+2H]^{2+}$, 481.9 $[M+3H]^{+3}$, observed 721.4 $[M+2H]^{2+}$, 481.4 $[M+3H]^{+3}$).

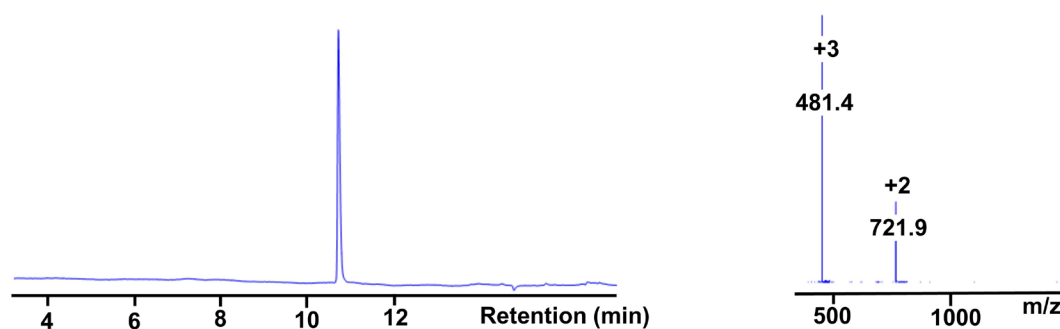


Figure S7.78: LC trace at UV210nm (left) of isolated L-TCPB-E-C1 (prepared by solid-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 $[M+2H]^{2+}$, 481.9 $[M+3H]^{+3}$, observed 721.9 $[M+2H]^{2+}$, 481.4 $[M+3H]^{+3}$).

7.3.37. L-TCPB-E-C4

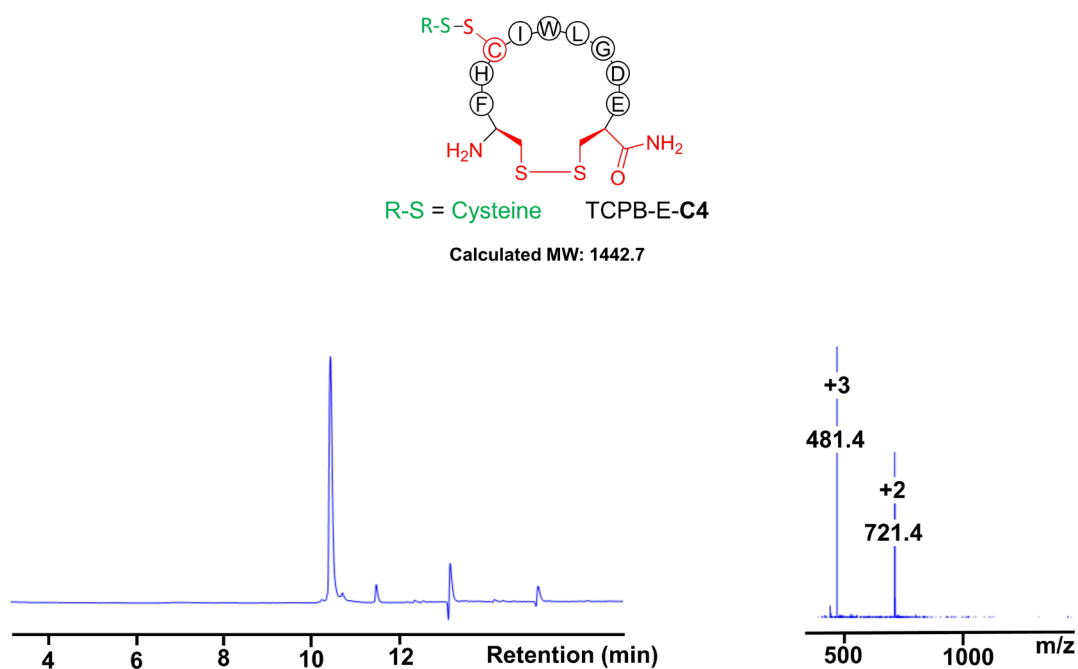


Figure S7.79: LC trace at UV210nm (left) of isolated L-TCPB-E-C4 (prepared by solid-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 722.4 $[M+2H]^{2+}$, 481.9 $[M+3H]^{+3}$, observed 721.4 $[M+2H]^{2+}$, 481.4 $[M+3H]^{+3}$).

7.3.38. L-TCPB-E-C11

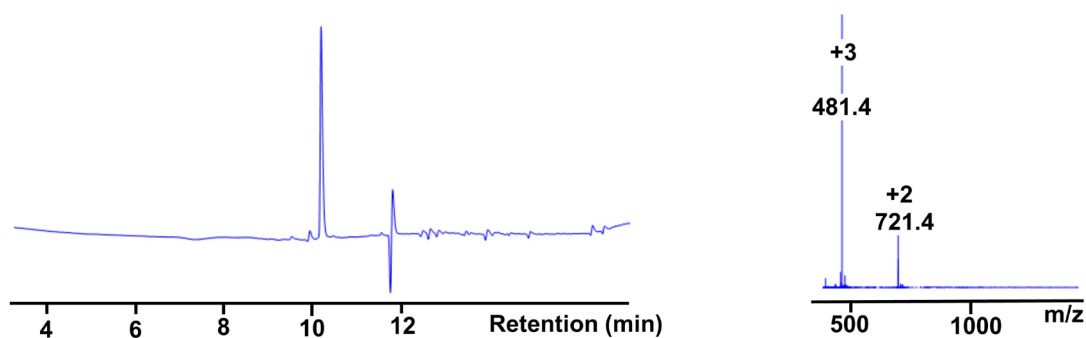
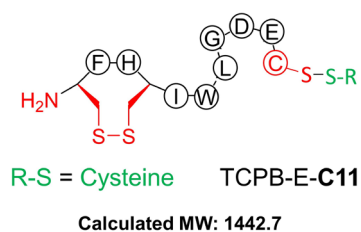


Figure S7.80: LC trace at UV210nm (left) of L-TCPB-E-C11 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 μm). Right - (ESI-MS (m/z): calculated 722.4 $[M+2H]^{2+}$, 481.9 $[M+3H]^{+3}$, observed 721.4 $[M+2H]^{2+}$, 481.4 $[M+3H]^{+3}$).

7.3.39. L-TCPB-E-G1

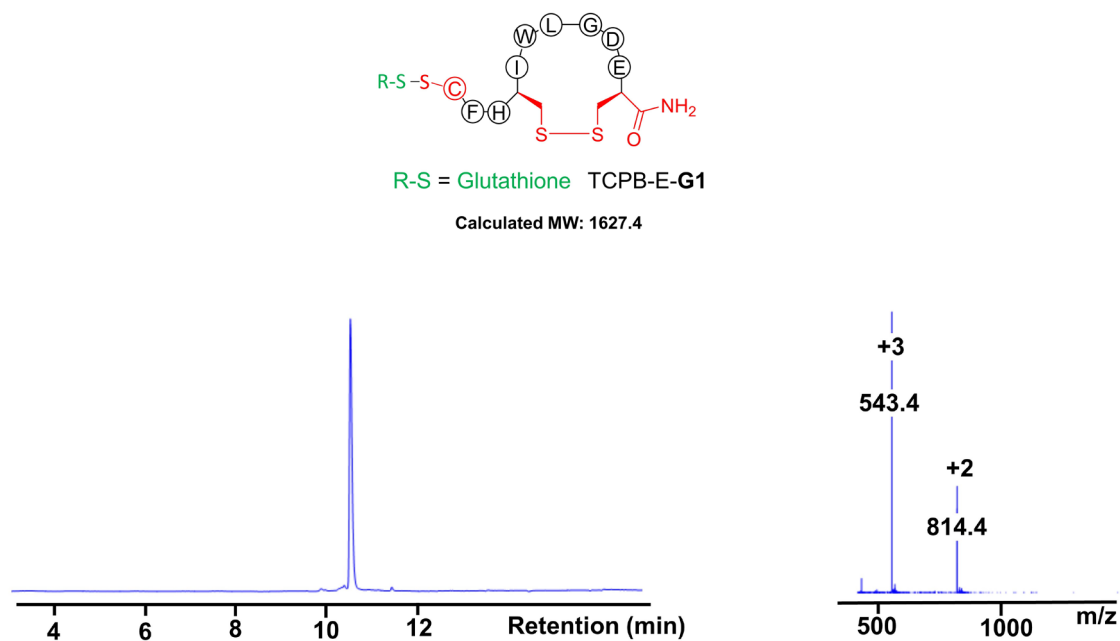


Figure S7.81: LC trace at UV210nm (left) of L-TCPB-E-G1 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 814.7 $[M+2H]^{2+}$, 543.5 $[M+3H]^{+3}$, observed 814.4 $[M+2H]^{2+}$, 543.4 $[M+3H]^{+3}$).

7.3.40. L-TCPB-E-G11

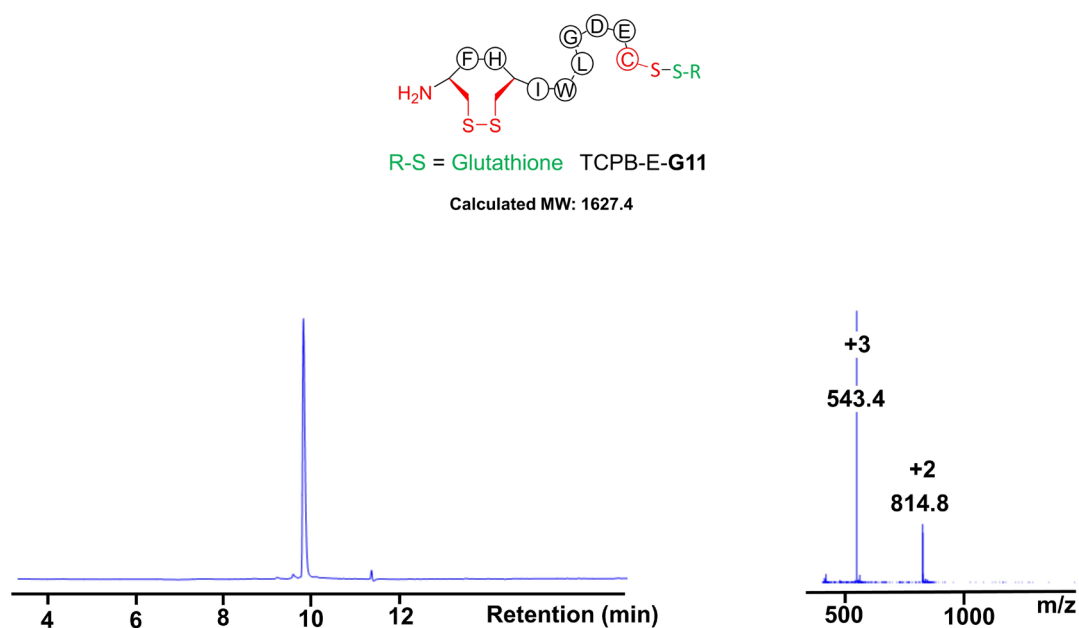


Figure S7.82: LC trace at UV210nm (left) of L-TCPB-E-G11 (isolated from solution-phase oxidation) using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 814.7 $[M+2H]^{2+}$, 543.5 $[M+3H]^{+3}$, observed 814.8 $[M+2H]^{2+}$, 543.4 $[M+3H]^{+3}$).

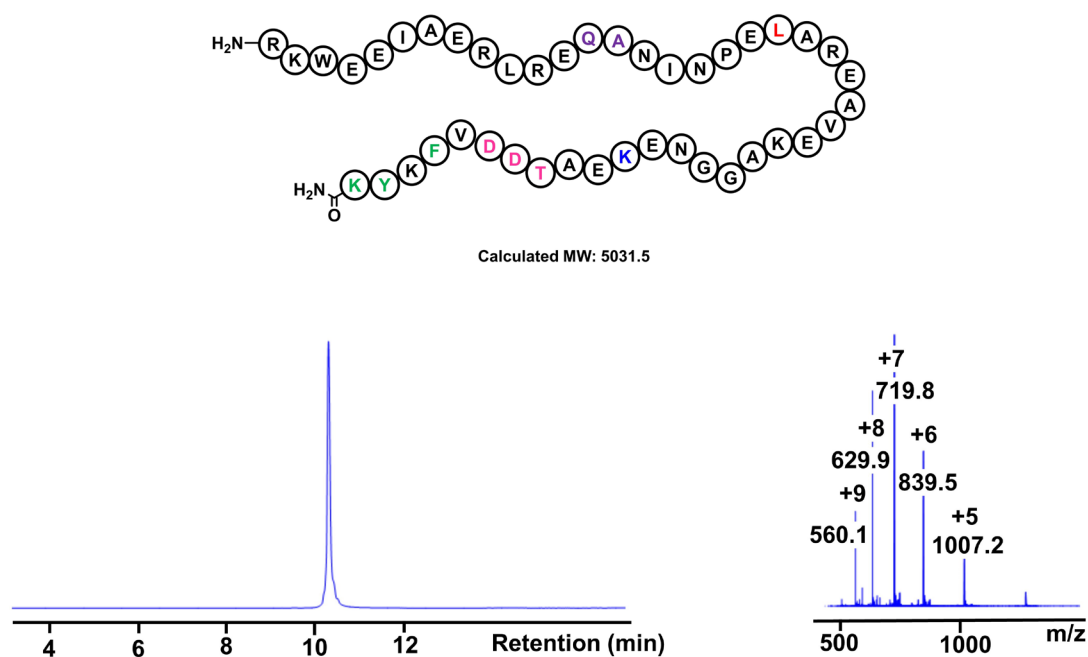
7.3.41. L-TMPB-T2 (Arg¹-Lys⁴³)

Figure S7.83: LC trace at UV210nm (left) of isolated L-TMPB-T2 using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1007.3 [M+5H]⁵⁺, 839.6 [M+6H]⁶⁺, 719.8 [M+7H]⁷⁺, 629.9 [M+8H]⁸⁺, 560.1 [M+9H]⁹⁺, observed 1007.2 [M+5H]⁵⁺, 839.5 [M+6H]⁶⁺, 719.8 [M+7H]⁷⁺, 629.9 [M+8H]⁸⁺, 560.1 [M+9H]⁹⁺).

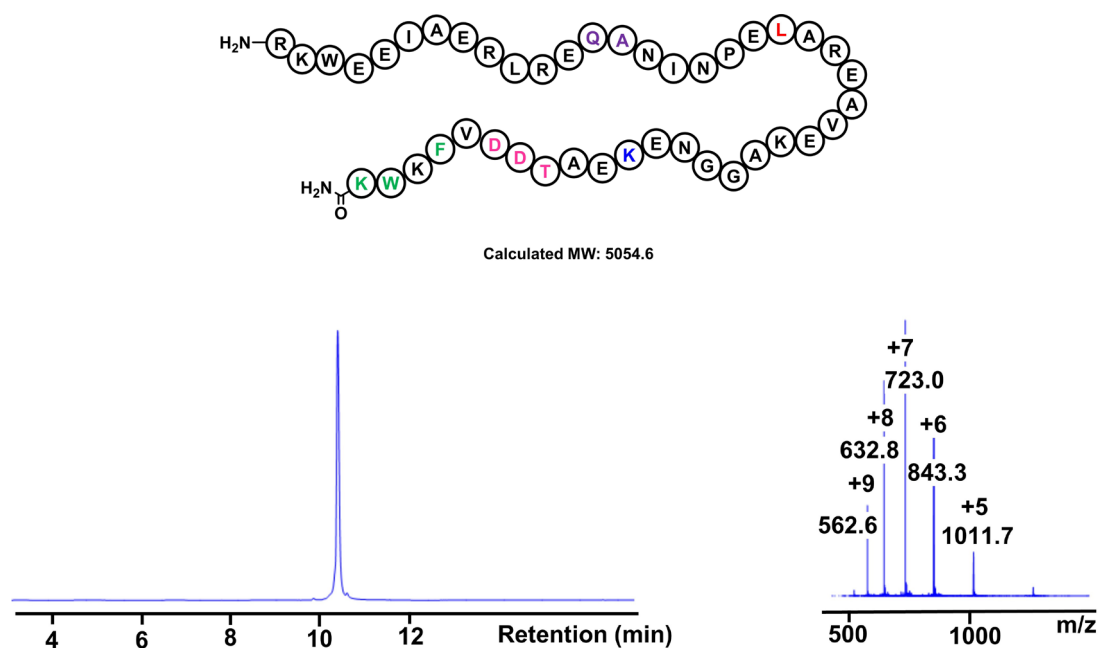
7.3.42. L-TMPB-T3 (Arg¹-Lys⁴³)

Figure S7.84: LC trace at UV210nm (left) of isolated L-TMPB-T3 using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1011.9 [M+5H]⁵⁺, 843.4 [M+6H]⁶⁺, 723.1 [M+7H]⁷⁺, 632.8 [M+8H]⁸⁺, 562.6 [M+9H]⁹⁺, observed 1011.7 [M+5H]⁵⁺, 843.3 [M+6H]⁶⁺, 723.0 [M+7H]⁷⁺, 632.8 [M+8H]⁸⁺, 562.6 [M+9H]⁹⁺).

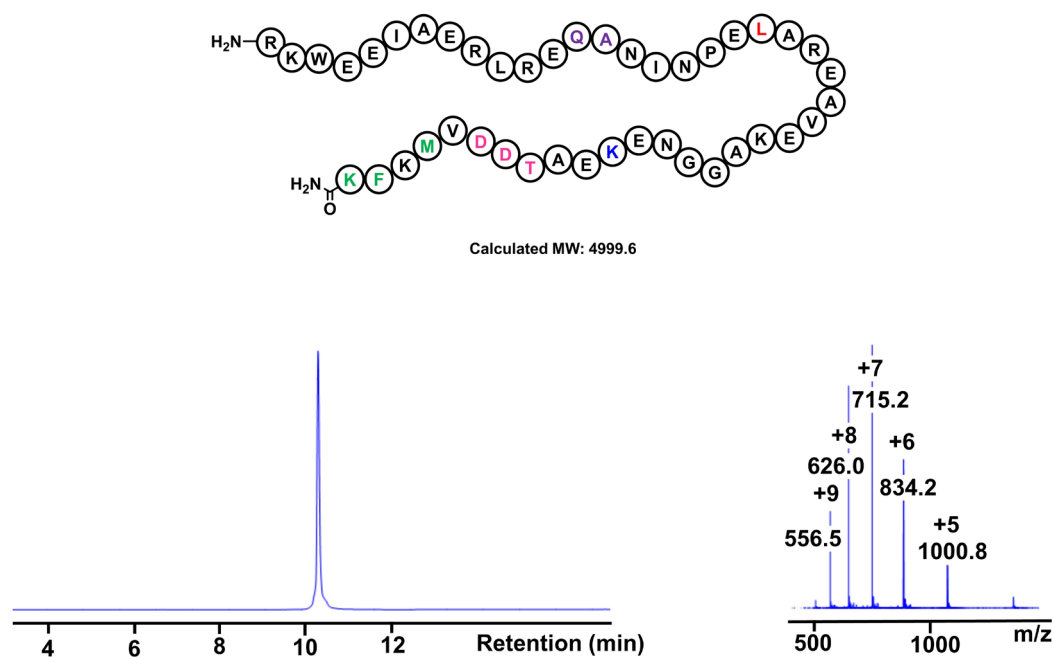
7.3.43. L-TMPB-T4 (Arg¹-Lys⁴³)

Figure S7.85: LC trace at UV210nm (left) of isolated L-TMPB-T4 using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1000.9 [M+5H]⁵⁺, 834.3 [M+6H]⁶⁺, 715.2 [M+7H]⁷⁺, 625.9 [M+8H]⁸⁺, 556.5 [M+9H]⁹⁺, observed 1000.8 [M+5H]⁵⁺, 834.2 [M+6H]⁶⁺, 715.2 [M+7H]⁷⁺, 626.0 [M+8H]⁸⁺, 556.5 [M+9H]⁹⁺).

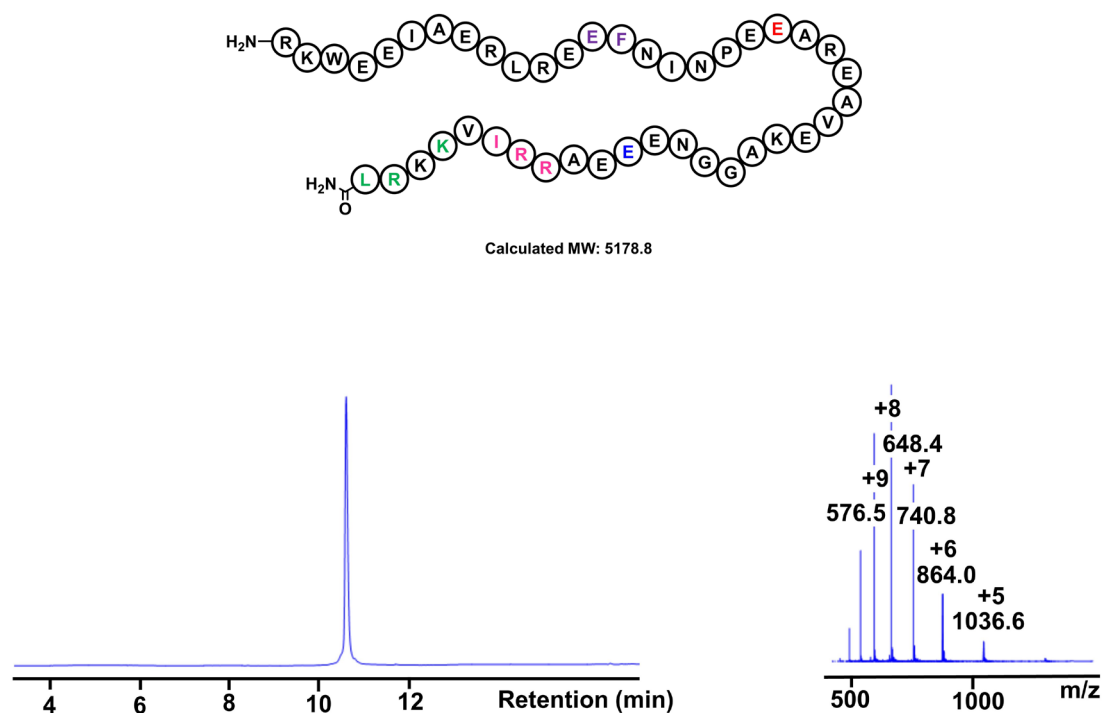
7.3.44. L-TMPB-Sc (Arg¹-Leu⁴³)

Figure S7.86: LC trace at UV210nm (left) of isolated L-TMPB-Sc using a 5-70% gradient of A/B over 20 minutes on a RP-C18 column (Zorbax SB, 2.1 mm x 100 mm, 300 Å, 3.5 µm). Right - (ESI-MS (m/z): calculated 1036.8 [M+5H]⁵⁺, 864.1 [M+6H]⁶⁺, 740.8 [M+7H]⁷⁺, 648.3 [M+8H]⁸⁺, 576.4 [M+9H]⁹⁺, observed 1036.6 [M+5H]⁵⁺, 864.0 [M+6H]⁶⁺, 740.8 [M+7H]⁷⁺, 648.4 [M+8H]⁸⁺, 576.5 [M+9H]⁹⁺).

7.4. Crystallography data refinement statistics

Table S7.7: X-ray data collection, processing, and refinement statistics. [a] TNRC2 not yet deposited in PDB.

	AucA-SO₄²⁻	AucA apo	AucA-G3P	AucA dimer	LnqQ	TNRC2
PDB accession	8AVR	8AVS	8AVT	8AVU	7P5R	[a]
Data collection						
Space group	C12 ₁	C12 ₁	C12 ₁	C12 ₁	P1	P12 ₁ 2
Cell dimensions						
a, b, c (Å)	72.7, 36.6, 106.2	73.1, 37.3, 114.0	73.7, 37.2, 114.1	79.6, 23.1, 52.6	24.2, 27.8, 70.5	20.4, 50.5, 46.2
α, β, γ (°)	90.0, 96.0, 90.0	90.0, 93.1, 90.0	90.0, 93.0, 90.0	90.0, 114.351, 90.0	95.0, 90.4, 115.4	90.0, 92.9, 90.0
Resolution (Å)	36.17 – 1.13 (1.15 - 1.13)	36.48 – 1.21 (1.23 - 1.21)	34.51 – 1.20 (1.22 - 1.20)	21.99 – 0.89 (0.91 - 0.89)	35.05 – 0.96 (0.98 - 0.96)	34.06 – 1.40 (1.42 - 1.40)
R _{sym} or R _{merge}	0.060 (0.354)	0.051 (0.599)	0.119 (1.093)	0.030 (0.624)	0.024 (0.803)	0.090 (1.076)
I / σI	9.7 (1.4)	8.8 (1.0)	8.2 (1.3)	17.2 (1.0)	11.3 (1.0)	11.3 (1.0)
Completeness	86 (38.7)	96.6 (80.3)	99.9 (99.6)	92.7 (46.4)	93.0 (84.2)	100 (100)
Redundancy	3.0 (1.2)	3.1 (1.8)	13.4 (13.6)	5.9 (3.5)	3.5 (3.0)	6.0 (5.7)
Refinement						
Resolution (Å)	36.2 – 1.13	36.51 – 1.21	34.53 – 1.20	22.00 – 0.89	35.05-0.96	[a]
No. reflections	89417(4466)	90456(4505)	96947(4912)	63267(3106)	93924(4211)	[a]
R _{work} / R _{free}	0.183/0.209	0.187/0.221	0.176/0.203	0.194/0.211	0.185/0.211	[a]
No. atoms						
Protein	1716	1710	1722	874	1733	[a]
Ligand/ion	50	56	35	43	21	[a]
Water	428	393	406	153	190	[a]
B-factors						
Protein	13.94	19.5	18.61	10.59	13.60	[a]
Ligand/ion	14.26	21.71	19.06	11.05	23.10	[a]
Water	25.35	30.5	31.81	21.48	22.40	[a]
R.m.s deviations						
Bond lengths (Å)	0.0085	0.0100	0.0106	0.0099	0.0115	[a]
Bond angles (°)	1.404	1.339	1.703	1.713	1.817	[a]

7.5. ROSETTA protein *de novo* protein design scripts

7.5.1. Example command line for running ROSETTA scripts in Linux

Alex_lander_PC:~/Documents/Folder_containing_files\$
rosetta_scripts.linuxgccrelease FlexDesign.xml @DesignFlag

7.5.2. Relax script

Text included in *red italic* are annotations and are not part of the script.

```
<ROSETTASCRIPITS>
  <SCOREFXNS>
    <ScoreFunction name="st" weights="ref2015_cst"/>
  </SCOREFXNS>
  <RESIDUE_SELECTORS>
  </RESIDUE_SELECTORS>
  <TASKOPERATIONS>
    <IncludeCurrent name="IC" />
    <RestrictToRepacking name="repack" />
  </TASKOPERATIONS>
  <FILTERS>
    <PackStat name="packstat" threshold="0.58" confidence="0"/> ideally, ~0.65
    is required for good packing
    <BuriedUnsatHbonds name="buns" scorefxn="st" confidence="0"/>
  </FILTERS>
  <MOVERS>
    <VirtualRoot name="root"/>
```

```

    <AtomCoordinateCstMover name="coord" sidechain="true"/>
    <FastRelax name="relax" scorefxn="st" task_operations="IC"/>
    <MinMover name="min" scorefxn="st" chi="1" bb="1" cartesian="false"/>
</MOVERS>
<PROTOCOLS>
    <Add mover_name="root"/>
    <Add mover_name="coord"/>
    <Add mover_name="relax"/>
    <Add filter_name="packstat"/>
    <Add filter_name="holes"/>
    <Add filter_name="buns"/>
</PROTOCOLS>
<OUTPUT />
</ROSETTASCRIPITS>

```

7.5.3. Relax flag

```

-in:file:s scaffold_protein_5UOI.pdb
-parser:protocol relax_cst.xml

-nstruct 1
-relax:default_repeats 5
-out:prefix relaxed_

```

7.5.4. GlobalDock script

Text included in *red italic* are annotations and are not part of the script.

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
</SCOREFXNS>
  <RESIDUE_SELECTORS>
    Define the residues where we want binding. Here, the key residues for
    TNF-alpha binding are listed
    <Index name="targets" resnums="77B-80B,107B,110B,111B" />
    Select the residues which are 'X' distance from the target residues
    <Neighborhood name="neighbours" selector="targets" distance="7.0"/>
    <Chain chains="A" name="scaffold"/> Chain of scaffold protein to add
    <And name="near_residues" selectors="neighbours"/>
    <Or name="interface1" selectors="targets,near_residues"/>
    <Not name="not_interface1" selector="interface1"/>
  </RESIDUE_SELECTORS>
  <TASKOPERATIONS>
    <OperateOnResidueSubset name="int2" selector="not_interface1">
      <PreventRepackingRLT/>
    </OperateOnResidueSubset>
  </TASKOPERATIONS>
  <FILTERS>
    How well do the 3D shapes of protein and target pack together.
    <ShapeComplementarity name="shape_complementarity"
    min_sc="0.3" jump="1" confidence="0" />
    Measure solvent accessible surface area, how much of the total
    surface area is covered by the interface.
    <Sasa name="sasa" threshold="1000"
    upper_threshold="10000000000000000" jump="1" confidence="0" />
```

```
<AverageDegree name="connectivity" threshold="0"
distance_threshold="7.0" task_operations="int2" confidence="0"/>
```

Total residues at binding interface, more residues is better (more design possible).

```
<ResInInterface name='InterfaceRes' confidence='0'/>
```

Total buried surface area of "targets". The main filter referred to in chapter 3. This defines which scaffold are interacting with the TNF-alpha face.

```
<BuriedSurfaceArea name="BurResTNF" filter_out_low="true"
cutoff_buried_surface_area="1600" atom_mode="all_atoms"
residue_selector="targets" confidence="0"/>
```

```
</FILTERS>
```

```
<MOVERS>
```

```
<DockSetupMover name="dockset" partners="A_B"/> these are the
chains corresponding to the docking and the scaffold.
```

```
<DockingInitialPerturbation randomize1="1" randomize2="1"
center_at_interface="1" slide="1" name="initdockpert"/> randomize the
docking.
```

```
<Docking name="dockprot" score_low="score_docking_low"
fullatom="0" local_refine="0" jumps="1"/> This is where the GlobalDock
is differed from the FullDock.
```

```
</MOVERS>
```

```
<APPLY_TO_POSE>
```

```
</APPLY_TO_POSE>
```

```
<PROTOCOLS>
```

```
<Add mover_name="dockset"/>
```

```
<Add mover_name="initdockpert"/>
```

```
<Add mover_name="dockprot"/>
```

```
<Add filter_name="shape_complementarity"/>
```

```
<Add filter_name="connectivity"/>
```

```
<Add filter_name="sasa"/>
```

```
<Add filter='InterfaceRes'/>  
<Add filter='BurRes'/>  
  
</PROTOCOLS>  
<OUTPUT />  
</ROSETTASCRIPITS>
```

7.5.5. GlobalDock flag

```
-in:file:s TNR_5uoi_relaxed.pdb  
  
-out:file:silent docked_structure_data.silent  
-out:file:scorefile scores_for_docking.sc  
-parser:protocol GlobalDock.xml  
  
-nstruct 1000  
  
-ex1  
-ex2aro
```

7.5.6. FlexDesign script

Text included in *red italic* are annotations and are not part of the script.

```
<ROSETTASCRIPTS>
```

```
<SCOREFXNS>
```

```
<ScoreFunction name="sfxn" weights="ref2015" />
```

```
<ScoreFunction name="sfxn_design" weights="ref2015_cst" >
```

```
<Reweight scoretype="approximate_buried_unsat_penalty" weight="5.0" />
```

```
</ScoreFunction>
```

```
<ScoreFunction name="vdw_sol" weights="empty" >
```

```
<Reweight scoretype="fa_atr" weight="1.0" />
```

```
<Reweight scoretype="fa_rep" weight="0.55" />
```

```
<Reweight scoretype="fa_sol" weight="1.0" />
```

```
</ScoreFunction>
```

```
</SCOREFXNS>
```

```
<RESIDUE_SELECTORS>
```

```
<Index name="good_target_res" resnums="77,79"
error_on_out_of_bounds_index="true" reverse="true" /> Define the residues
involved in TNF-alpha binding that are "good" targets.
```

```
<Chain name="chainA" chains="A"/>
```

```
<Chain name="chainB" chains="B"/>
```

```
<Neighborhood name="interface_chA" selector="chainB" distance="8.0" />
```

```
<Neighborhood name="interface_chB" selector="chainA" distance="8.0" />
```

```
<And name="AB_interface" selectors="interface_chA,interface_chB" />
```

```
<Not name="Not_interface" selector="AB_interface" />
```



```
<ResidueName name="pro_and_gly_positions" residue_name3="PRO,GLY"
/> Used to ban substitutions, to avoid disruption of protein fold.
```

```
<!-- Layer Design --> Adapted from the literature. Here, the parts of the
protein are defined, so that allowed designable resiudes can be specified.
```

```
<Index name="Nter" resnums="1" error_on_out_of_bounds_index="true"
reverse="false" />
```

```
<Index name="Cter" resnums="40" error_on_out_of_bounds_index="true"
reverse="true" />
```

```
<Layer name="surface" select_core="false" select_boundary="false"
select_surface="true" use_sidechain_neighbors="true"/>
```

```
<Layer name="boundary" select_core="false" select_boundary="true"
select_surface="false" use_sidechain_neighbors="true"/>
```

```
<Layer name="core" select_core="true" select_boundary="false"
select_surface="false" use_sidechain_neighbors="true"/>
```

```
<SecondaryStructure name="sheet" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="E"/>
```

```
<SecondaryStructure name="entire_loop" overlap="0" minH="3" minE="2"
include_terminal_loops="true" use_dssp="true" ss="L"/>
```

```
<SecondaryStructure name="entire_helix" overlap="0" minH="3" minE="2"
include_terminal_loops="false" use_dssp="true" ss="H"/>
```

```
<And name="helix_cap" selectors="entire_loop">
```

```
<PrimarySequenceNeighborhood lower="1" upper="0" selector="entire_helix"/>
```

```
</And>
```

```
<And name="helix_start" selectors="entire_helix">
```

```
<PrimarySequenceNeighborhood lower="0" upper="1" selector="helix_cap"/>
```

```
</And>
```

```
<And name="helix" selectors="entire_helix">
```

```
  <Not selector="helix_start"/>
```

```
</And>
```

```
<And name="loop" selectors="entire_loop">
```

```
  <Not selector="helix_cap"/>
```

```
</And>
```

</RESIDUE_SELECTORS>

<TASKOPERATIONS>

<ProteinInterfaceDesign name="pack_long" design_chain1="0"
design_chain2="0" jump="1" interface_distance_cutoff="15"/>

###Rosetta recommended design task ops

<InitializeFromCommandline name="init" />

<IncludeCurrent name="current" /> *dont throw away input rotamers*

<LimitAromaChi2 name="limitchi2" chi2max="110" chi2min="70"
include_trp="True" /> *Prevents unrealistic rotamers of F, Y, H and W(if True)*

<ExtraRotamersGeneric name="ex1_ex2" ex1="1" ex2aro="1" />

<ExtraRotamersGeneric name="ex2" ex2="1" />

<ConsensusLoopDesign name="consensus_loop" />

Change a.a. preference of core layers

<DesignRestrictions name="layer_design">

<Action selector_logic="Nter AND chainA"
aas="ADEF GHIKLMNPQRSTVWY" />

<Action selector_logic="Cter AND chainA"
aas="ADEF GHIKLMNPQRSTVWY" />

<Action selector_logic="surface AND helix_start AND chainA"
aas="EHKPQRDNST"/>

<Action selector_logic="surface AND helix AND chainA"
aas="EHKQRDNST"/>

<Action selector_logic="surface AND sheet AND chainA"
aas="DEHKNQRST"/>

<Action selector_logic="surface AND loop AND chainA"
aas="DEGHKNPQRST"/>

<Action selector_logic="boundary AND helix_start AND chainA"
aas="ADEIKLMNPQRSTVWY"/>

<Action selector_logic="boundary AND helix AND chainA"
aas="ADEFIKLMNPQRSTVWY"/>

<Action selector_logic="boundary AND sheet AND chainA"
aas="DEFIKLNQRSTVWY"/>

```

    <Action selector_logic="boundary AND loop AND chainA"
    aas="ADEFGIKLNPQRSTVWY"/>

    <Action selector_logic="core AND helix_start AND chainA"
    aas="AFILMPVWY"/>

    <Action selector_logic="core AND helix AND chainA"
    aas="AFILMVWYDENQTS"/>

    <Action selector_logic="core AND sheet AND chainA"
    aas="FILMVWYDENQST"/>

    <Action selector_logic="core AND loop AND chainA"
    aas="AFGILMPVWY"/>

    <Action selector_logic="helix_cap AND chainA"
    aas="DNST"/>

</DesignRestrictions>

```

#####restrict to interface, we only want to design the residues of the scaffold that are near the target.

```

<SelectBySASA name="PR_monomer_core" mode="sc" state="monomer"
probe_radius="2.2" core_asa="15" surface_asa="30" core="0" boundary="1"
surface="1" verbose="1" />

<OperateOnResidueSubset name="restrict_to_interface"
selector="Not_interface">

    <PreventRepackingRLT/>

</OperateOnResidueSubset>

<OperateOnResidueSubset name="restrict_target" selector="chainB">

    <PreventRepackingRLT/>

</OperateOnResidueSubset>

    <OperateOnResidueSubset name="restrict_target2repacking"
    selector="chainB">

        <RestrictToRepackingRLT/>

    </OperateOnResidueSubset>

<DisallowIfNonnative name="disallow_GLY" resnum="0" disallow_aas="G" />

<DisallowIfNonnative name="disallow_PRO" resnum="0" disallow_aas="P" />

    <OperateOnResidueSubset name="restrict_PRO_GLY"
    selector="pro_and_gly_positions">

```

```

    <PreventRepackingRLT/>
  </OperateOnResidueSubset>

  <ProteinProteinInterfaceUpweighter name="up_ppi" interface_weight="2"
skip_loop_in_chain="A" />

</TASKOPERATIONS>

<MOVERS>
  <TaskAwareMinMover name="min" scorefxn="sfxn" bb="1" chi="1"
task_operations="pack_long" />
</MOVERS>

<FILTERS>

  <Sasa name="interface_buried_sasa" confidence="0" />

  Ddg is the commonly used filter discussed in Chapter 3, which provides a
simulated binding energy.

  <Ddg name="ddg" threshold="-10" jump="1" repeats="3" repack="1"
relax_mover="min" confidence="0" scorefxn="sfxn" />

  <ShapeComplementarity name="interface_sc" verbose="0" min_sc="0.55"
write_int_area="1" write_median_dist="1" jump="1" confidence="0"/>

  <ContactMolecularSurface name="contact_area_target" verbose="0"
distance_weight="0.5" confidence="0" target_selector="chainB"
binder_selector="chainA"/>

  <ContactMolecularSurface name="contact_area_binder" verbose="0"
distance_weight="0.5" confidence="0" target_selector="chainA"
binder_selector="chainB"/>


  <BuriedUnsatHbonds name="vbuns_heavy_all" use_reporter_behavior="true"
report_all_heavy_atom_unsats="true" scorefxn="sfxn"
residue_selector="AB_interface" ignore_surface_res="false"
print_out_info_to_pdb="true" confidence="0" use_ddG_style="true"
dalphaball_sasa="false" probe_radius="1.1" atomic_depth_selection="5.5"
burial_cutoff="1000" burial_cutoff_apo="0.2" />

  <BuriedUnsatHbonds name="vbuns_heavy_sc" use_reporter_behavior="true"
report_sc_heavy_atom_unsats="true" scorefxn="sfxn"
residue_selector="AB_interface" ignore_surface_res="false"
print_out_info_to_pdb="true" confidence="0" use_ddG_style="true"
dalphaball_sasa="false" probe_radius="1.1" atomic_depth_selection="5.5"
burial_cutoff="1000" burial_cutoff_apo="0.2" />

```

```
<BuriedUnsatHbonds name="vbuns_heavy_bb" use_reporter_behavior="true"
report_bb_heavy_atom_unsats="true" scorefxn="sfxn"
residue_selector="AB_interface" ignore_surface_res="false"
print_out_info_to_pdb="true" confidence="0" use_ddG_style="true"
dalphaball_sasa="false" probe_radius="1.1" atomic_depth_selection="5.5"
burial_cutoff="1000" burial_cutoff_apo="0.2" />
```

```
</FILTERS>
```

```
<MOVERS>
```

```
<FavorSequenceProfile name="favour_seq" weight="1" use_current="true"
scaling="prob" matrix="MATCH"/> Penalize mutations, keep as much scaffold
native as possible!
```

```
<FastDesign name="FastDesign" scorefxn="sfxn_design" repeats="2"
task_operations="init,current,limitchi2,ex1_ex2,ex2,layer_design,restrict_to_inte
rface,restrict_target2repacking,disallow_GLY,disallow_PRO,PR_monomer_cor
e,up_ppi,consensus_loop,restrict_PRO_GLY" batch="false"
ramp_down_constraints="false" cartesian="false" bondangle="false"
bondlength="false" min_type="dfpmin_armijo_nonmonotone" >
```

```
<MoveMap name="MM" >
```

```
<Chain number="1" chi="true" bb="true" />
```

```
<Chain number="2" chi="true" bb="true" />
```

```
<Jump number="1" setting="true" />
```

```
</MoveMap>
```

```
</FastDesign>
```

```
<FastRelax name="FastRelax" scorefxn="sfxn" repeats="5" batch="false"
ramp_down_constraints="false" cartesian="false" bondangle="false"
bondlength="false" min_type="dfpmin_armijo_nonmonotone"
task_operations="restrict_target,limitchi2,init,current,ex2,ex1_ex2" >
```

```
<MoveMap name="MM" >
```

```
<Chain number="1" chi="true" bb="true" />
```

```
<Chain number="2" chi="true" bb="true" />
```

```
<Jump number="1" setting="true" />
```

```

    </MoveMap>
  </FastRelax>
</MOVERS>
<APPLY_TO_POSE>
</APPLY_TO_POSE>
<PROTOCOLS>
  <Add mover="favour_seq" />
  <Add mover="FastDesign" />
  <Add mover="FastRelax" />
  <Add filter_name="interface_buried_sasa" />
  <Add filter_name="contact_area_target" />
  <Add filter_name="contact_area_binder" />
  <Add filter_name="ddg" />
  <Add filter_name="interface_sc" />
  <Add filter_name="vbuns_heavy_all" />
    <Add filter_name="vbuns_heavy_sc" />
      <Add filter_name="vbuns_heavy_bb" />
</PROTOCOLS>
<OUTPUT />
</ROSETTASCRIPTS>

```

7.5.7. Design flag

```

-in:file:s docked_TNR_5uoi.pdb
-out:file:scorefile scores_for_design.sc

-parser:protocol FlexDesign.xml

-nstruct 100

```

7.5.8. FullDock script

Text included in *red italic* are annotations and are not part of the script.

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="sfxn" weights="ref2015" />
    <ScoreFunction name="vdw_sol" weights="empty" >
      <Reweight scoretype="fa_atr" weight="1.0" />
      <Reweight scoretype="fa_rep" weight="0.55" />
      <Reweight scoretype="fa_sol" weight="1.0" />
    </ScoreFunction>
  </SCOREFXNS>
  <RESIDUE_SELECTORS>
    <Chain name="chainA" chains="A"/>
    <Chain name="chainB" chains="B"/>
    <Neighborhood name="interface_chA" selector="chainB" distance="8.0" />
    <Neighborhood name="interface_chB" selector="chainA" distance="8.0" />
    <And name="AB_interface" selectors="interface_chA,interface_chB" />
    <Not name="Not_interface" selector="AB_interface" />
  </RESIDUE_SELECTORS>
  <TASKOPERATIONS>
    <ProteinInterfaceDesign name="pack_long" design_chain1="0"
    design_chain2="0" jump="1" interface_distance_cutoff="15"/>

    ###Rosetta recommended design task ops

    <InitializeFromCommandline name="init" />
    <IncludeCurrent name="current" /> dont throw away input rotamers
    <LimitAromaChi2 name="limitchi2" chi2max="110" chi2min="70"
    include_trp="True" /> prevents unrealistic rotamers of F, Y, H and W(if True)
```

```
<ExtraRotamersGeneric name="ex1_ex2" ex1="1" ex2aro="1" />
```

```
<ExtraRotamersGeneric name="ex2" ex2="1" />
```

```
</TASKOPERATIONS>
```

Same filters as design incl. task aware min mover for ddg (only use interface to calculate ddg)

```
<MOVERS>
```

```
<TaskAwareMinMover name="min" scorefxn="sfxn" bb="1" chi="1"
task_operations="pack_long" />
```

```
</MOVERS>
```

```
<FILTERS>
```

```
<Sasa name="interface_buried_sasa" confidence="0" />
```

```
<Ddg name="ddg" threshold="-10" jump="1" repeats="3" repack="1"
relax_mover="min" confidence="0" scorefxn="sfxn" />
```

```
<ShapeComplementarity name="interface_sc" verbose="0" min_sc="0.55"
write_int_area="1" write_median_dist="1" jump="1" confidence="0"/>
```

```
<ContactMolecularSurface name="contact_area_target" verbose="0"
distance_weight="0.5" confidence="0" target_selector="chainB"
binder_selector="chainA"/>
```

```
<ContactMolecularSurface name="contact_area_binder" verbose="0"
distance_weight="0.5" confidence="0" target_selector="chainA"
binder_selector="chainB"/>
```

```
<BuriedUnsatHbonds name="vbuns_heavy_all" use_reporter_behavior="true"
report_all_heavy_atom_unsats="true" scorefxn="sfxn"
residue_selector="AB_interface" ignore_surface_res="false"
print_out_info_to_pdb="true" confidence="0" use_ddG_style="true"
dalphaball_sasa="false" probe_radius="1.1" atomic_depth_selection="5.5"
burial_cutoff="1000" burial_cutoff_apo="0.2" />
```

```
<BuriedUnsatHbonds name="vbuns_heavy_sc" use_reporter_behavior="true"
report_sc_heavy_atom_unsats="true" scorefxn="sfxn"
residue_selector="AB_interface" ignore_surface_res="false"
print_out_info_to_pdb="true" confidence="0" use_ddG_style="true"
```



```
dalphaball_sasa="false" probe_radius="1.1" atomic_depth_selection="5.5"
burial_cutoff="1000" burial_cutoff_apo="0.2" />
```

```
<BuriedUnsatHbonds name="vbuns_heavy_bb" use_reporter_behavior="true"
report_bb_heavy_atom_unsats="true" scorefxn="sfxn"
residue_selector="AB_interface" ignore_surface_res="false"
print_out_info_to_pdb="true" confidence="0" use_ddG_style="true"
dalphaball_sasa="false" probe_radius="1.1" atomic_depth_selection="5.5"
burial_cutoff="1000" burial_cutoff_apo="0.2" />
```

```
</FILTERS>
```

```
<MOVERS>
```

```
<DockSetupMover name="dockset" partners="A_B"/>
```

```
<DockingInitialPerturbation randomize1="0" randomize2="0"
use_ellipsoidal_randomization="0" dock_pert="true" trans="3" rot="8" spin="0"
center_at_interface="0" slide="0" name="initdockpert"/>
```

```
<DockingProtocol name="Docking" docking_score_low="score_docking_low"
docking_score_high="sfxn" low_res_protocol_only="0" docking_local_refine=
"0" dock_min="1" ignore_default_docking_task="0"
task_operations="ex2,ex1_ex2" partners="A_B"/> This differs from
GlobalDock. Here, we first conduct docking in centroid mode, followed by
rounds off docking and minimization considering full amino acid side chains.
```

```
<FastRelax name="FastRelax" scorefxn="sfxn" repeats="5" batch="false"
ramp_down_constraints="false" cartesian="false" bondangle="false"
bondlength="false" min_type="dfpmin_armijo_nonmonotone"
task_operations="limitchi2,init,current,ex2,ex1_ex2" >
```

```
<MoveMap name="MM" >
```

```
<Chain number="1" chi="true" bb="true" />
```

```
<Chain number="2" chi="true" bb="true" />
```

```
<Jump number="1" setting="true" />
```

```
</MoveMap>
```

```
</FastRelax>
```

```
</MOVERS>
```

```
<APPLY_TO_POSE>
```

```
</APPLY_TO_POSE>
```

```

<PROTOCOLS>
  <Add mover="dockset" />
  <Add mover="initdockpert" />
  <Add mover="Docking" />
  <Add mover="FastRelax" />
  <Add filter_name="interface_buried_sasa" />
  <Add filter_name="contact_area_target" />
  <Add filter_name="contact_area_binder" />
  <Add filter_name="ddg" />
  <Add filter_name="interface_sc" />
  <Add filter_name="vbuns_heavy_all" />
    <Add filter_name="vbuns_heavy_sc" />
    <Add filter_name="vbuns_heavy_bb" />

</PROTOCOLS>
<OUTPUT />
</ROSETTASCRIPTS>

7.5.9. FullDock flag

-in:file:s randomized_TMPB_T2_and_D-TNRCD2.pdb
-in:file:native designoutput_TMPB_T2_bound_to_D-TNRCD2.pdb

-out:file:scorefile TMPB_T2_FullDock_results_for_ddg_vs_RMSD.sc
-out:file:silent FullDock_coordinate_data.silent

-parser:protocol FullDock.xml

-nstruct 10000

```