Supplementary Information:

Protein ligation for the assembly and study of nonribosomal peptide synthetase megaenzymes

Angelos Pistofidis[1], T. Martin Schmeing[1]

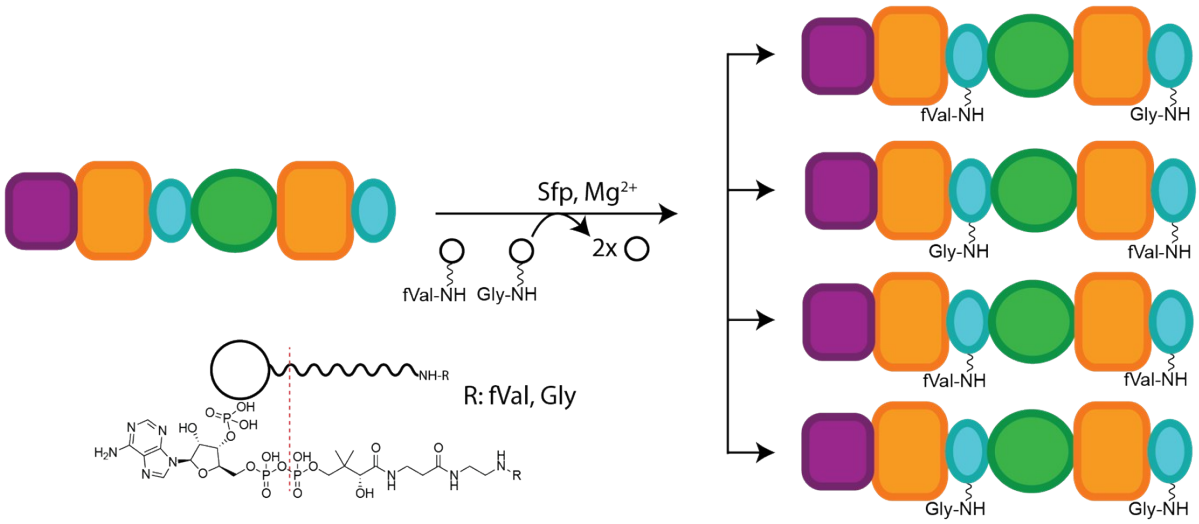
[1]Department of Biochemistry and Centre de recherche en biologie structurale, McGill University, Montréal, QC, H3G 0B1, Canada

Supplementary Table 1: Crystallographic data and refinement statistics

Data collection	**$F_1A_1T_{1(Val)}$-$C_2A_2T_{2(Gly)}$ (PDB: 9MEH)**
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	119.24, 166.28, 178.49
α, β, γ (°)	90, 90, 90
Resolution (Å)	50.00-3.60 (3.66-3.60)
R_{merge}	0.129 (2.49)
$I / \sigma I$	18.6 (0.875)
Completeness (%)	99.9 (100)
Redundancy	11.6 (11.3)
$CC_{1/2}$	1.0 (0.489)
Refinement	
Resolution (Å)	48.38-3.60
No. reflections	40600
R_{work} / R_{free} (%)	26.7 / 27.4
No. atoms	14432
Protein	14419
Ligand/ion	13
Water	0
B-factors	196.89
Protein	196.86
Ligand/ion	230.2
Water	-
R.m.s. deviations	
Bond lengths (Å)	0.002
Bond angles (°)	0.46

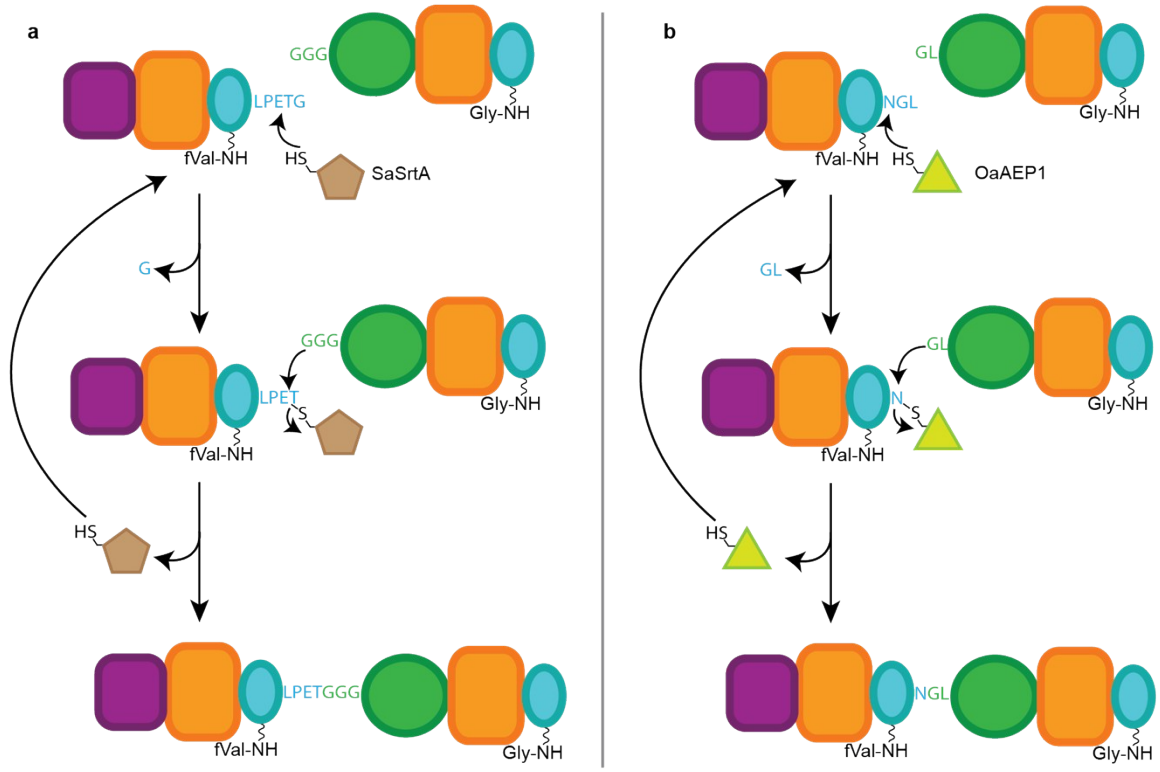
Supplementary Table 2: Primers used for cloning and mutagenesis of constructs

Name	Sequence
pBACt_FAT_NGL_Ins_For	GCGTGAGCAGAGAATTTGTACTTCCAAGGTAAAC
pBACt_FAT_NGL_Ins_Rev	GTACAAATTCTCTGCTCACGCTTAAAGACCATTCACCTGCTTTTGCTCCGTAAGCAGAC
pBACt_GL_CAT_Ins_For	CCGAAAACCTGTATTTTCAGGGCCTCCTGGAGCCTTTGCGAGAGCTGG
pBACt_GL_CAT_Ins_Rev	GCCCTGAAAATACAGGTTTTCGGATCCGG
pBACt_SrtA_FAT_Ins_For	CTTACGGAGCAAAAGCAGGTGCTTCCTGAAACTGGTTAAGCGTGAGCAGAGAATTTGTACTTCCAAGGTAAAC
pBACt_SrtA_FAT_Ins_Rev	CACCTGCTTTTGCTCCGTAAGCAGAC
pBACt_SrtA_CAT_Ins_For	CCGAAAACCTGTATTTTCAGGGCGGCGGGAGCCTGGAGCCTTTGCGAGAG
pBACt_SrtA_CAT_Ins_Rev	CTGAAAATACAGGTTTTCGGATCCGGTACCCAG
LgrABmdB_+NGL_For	AACGTCTGCTTACGGAGCAAAAGCAGGTGAATGGTCTTCTGGAGCCTTTGCGAGAGCTGGACG
LgrABmdB_+NGL_Rev	CGTCCAGCTCTCGCAAAGGCTCCAGAAGACCATTCACCTGCTTTTGCTCCGTAAGCAGACGTT
LgrABmdB_+LPETGGG_For	GTCTGCTTACGGAGCAAAAGCAGGTGCTTCCTGAAACTGGTGGCGGCAGCCTGGAGCCTTTGCGAGAG
LgrABmdB_+LPETGGG_Rev	CTCTCGCAAAGGCTCCAGGCTGCCGCCACCAGTTTCAGGAAGCACCTGCTTTTGCTCCGTAAGCAGAC
LgrABmdB_+SrtA_Δ5_For	GCTTCCTGAAACTGGTGGCGGCCGAGAGCTGGACGAGCAAGCG
LgrABmdB_+SrtA_Δ5_Rev	CGCTTGCTCGTCCAGCTCTCGGCCGCCACCAGTTTCAGGAAGC



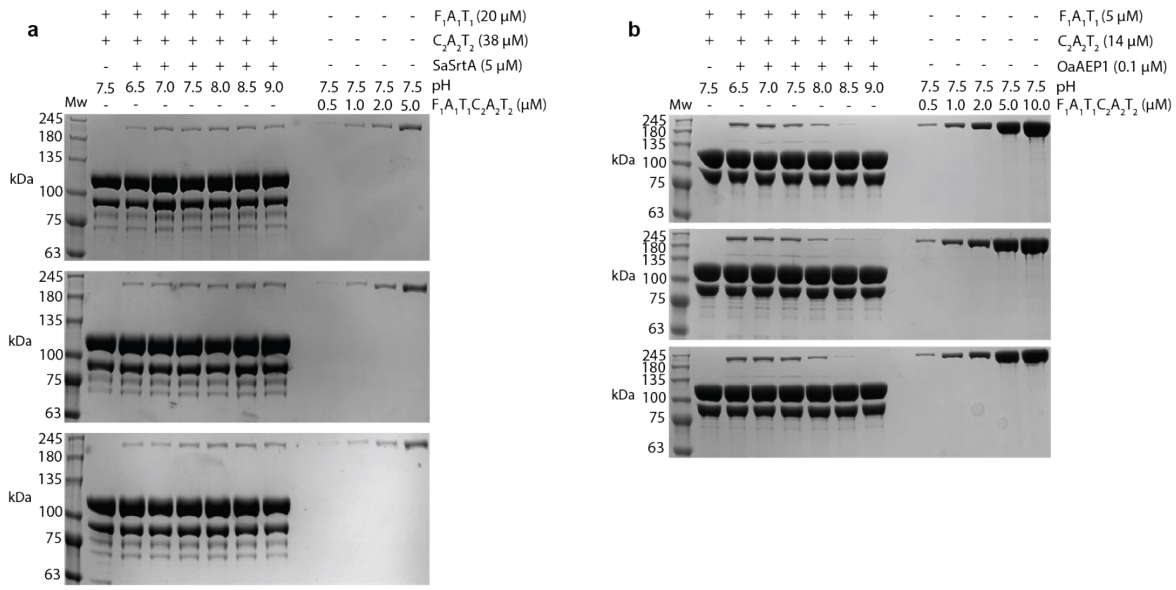
Supplementary Figure 1: Loading $F_1A_1T_1C_2A_2T_2$ with two different substrate analogues would result in a heterogeneous sample

Specifically loading two different amino acyl-ppant analogues onto their respective T domains (fVal-$_{NH}$-ppant on T_1 and Gly-$_{NH}$-ppant on T_2) in $F_1A_1T_1C_2A_2T_2$ is impossible because the promiscuity of the phosphopantetheine transferase enzyme, Sfp [1]. Attempting to do so would produce the mixed population shown on the right.



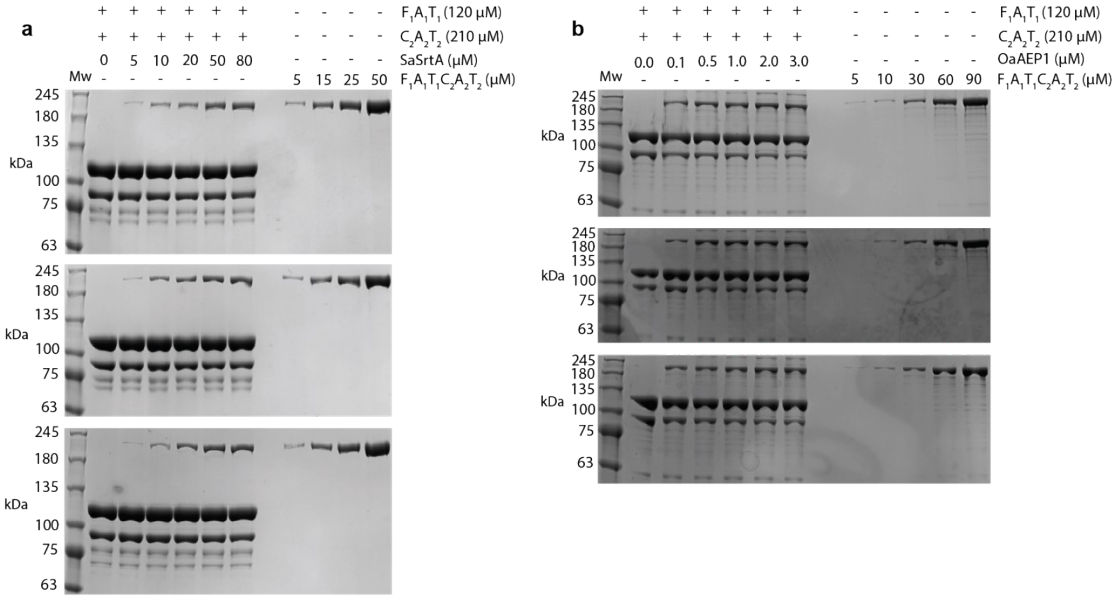
Supplementary Figure 2: The mechanism of protein ligation by SaSrtA and OaAEP1.

a. SaSrtA recognizes a Leu-Pro-Glu-Thr-Gly sequence at the C-terminus of $F_1A_1T_1^{Srt}$ and, via a conserved cysteine, displaces the glycine to form a reactive thioester with the threonine. A Gly-Gly-Gly sequence at the N-terminus of $^{Srt}C_2A_2T_2$ resolves the thioester to form a Leu-Pro-Glu-Thr-Gly-Gly-Gly ligation scar and join the two modules ($F_1A_1T_1$–$C_2A_2T_2$). **b.** OaAEP1 uses a similar mechanism but recognizes an Asn-Gly-Leu sequence at the C-terminus of $F_1A_1T_1^{AEP1}$ and a Gly-Leu sequence at the N-terminus of $^{AEP1}C_2A_2T_2$.



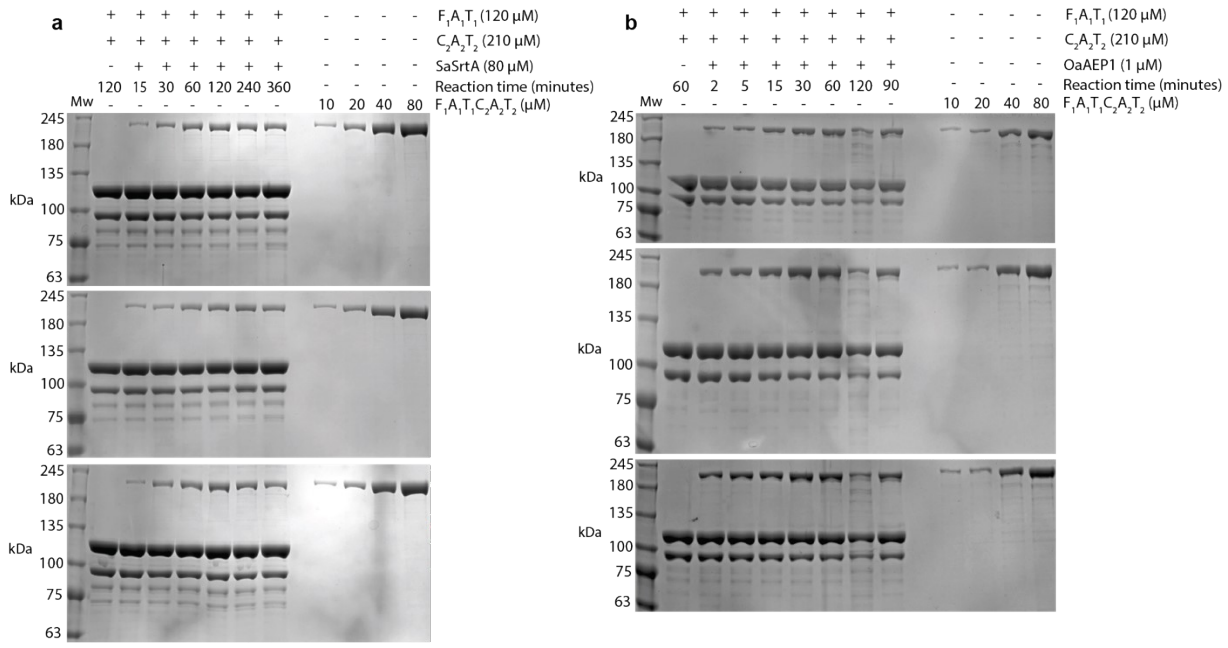
Supplementary Figure 3: SDS-PAGE of pH variation on $F_1A_1T_1$-$C_2A_2T_2$ formation by SaSrtA and OaAEP1

a-b. (**a)** OaAEP1 and (**b)** SaSrtA were incubated with appropriately sorting sequence labelled $F_1A_1T_1$ and $C_2A_2T_2$ at pHs between 6.5 and 9.0. Each gel in **a** and **b** represents a single replicate. On the left side of each gel is the negative control (no OaAEP1 or no SaSrtA) and the reactions at each pH (6.5 – 9.0). On the right side of each gel are known concentrations of $F_1A_1T_1$-$C_2A_2T_2$ used to prepare a calibration curve for the conversion of band intensities to yields (μM).



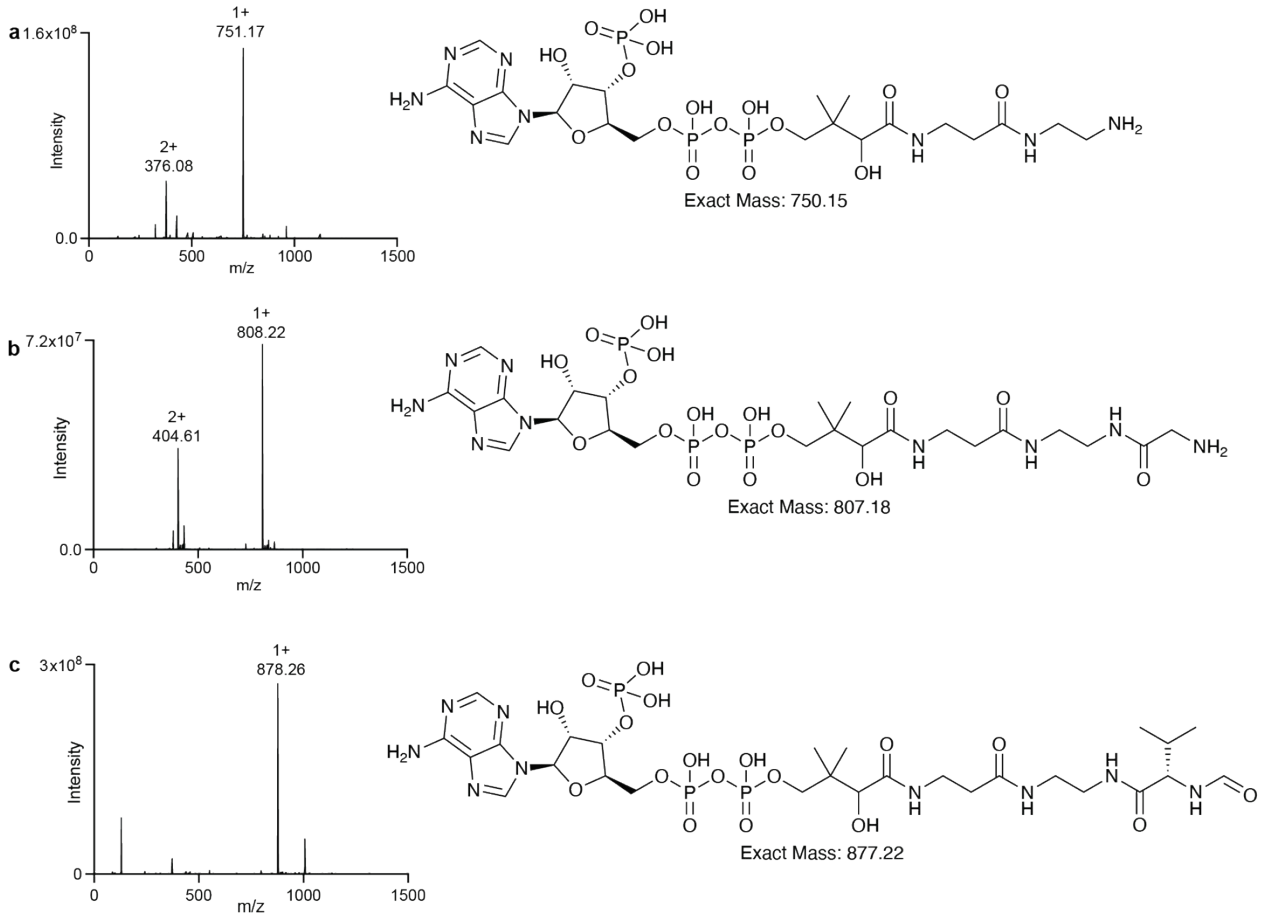
Supplementary Figure 4: SDS-PAGE of $F_1A_1T_1C_2A_2T_2$ formation with varying concentrations of SaSrtA and OaAEP1

a. SaSrtA concentration was varied between 0 – 80 µM. **b.** OaAEP1 concentration was varied between 0 and 3 µM. In **a** and **b,** each gel represents one replicate of a total of three each. On the right side of each gel are known concentrations of $F_1A_1T_1C_2A_2T_2$ used to prepare a calibration curve for the conversion of band intensities to yields (µM).

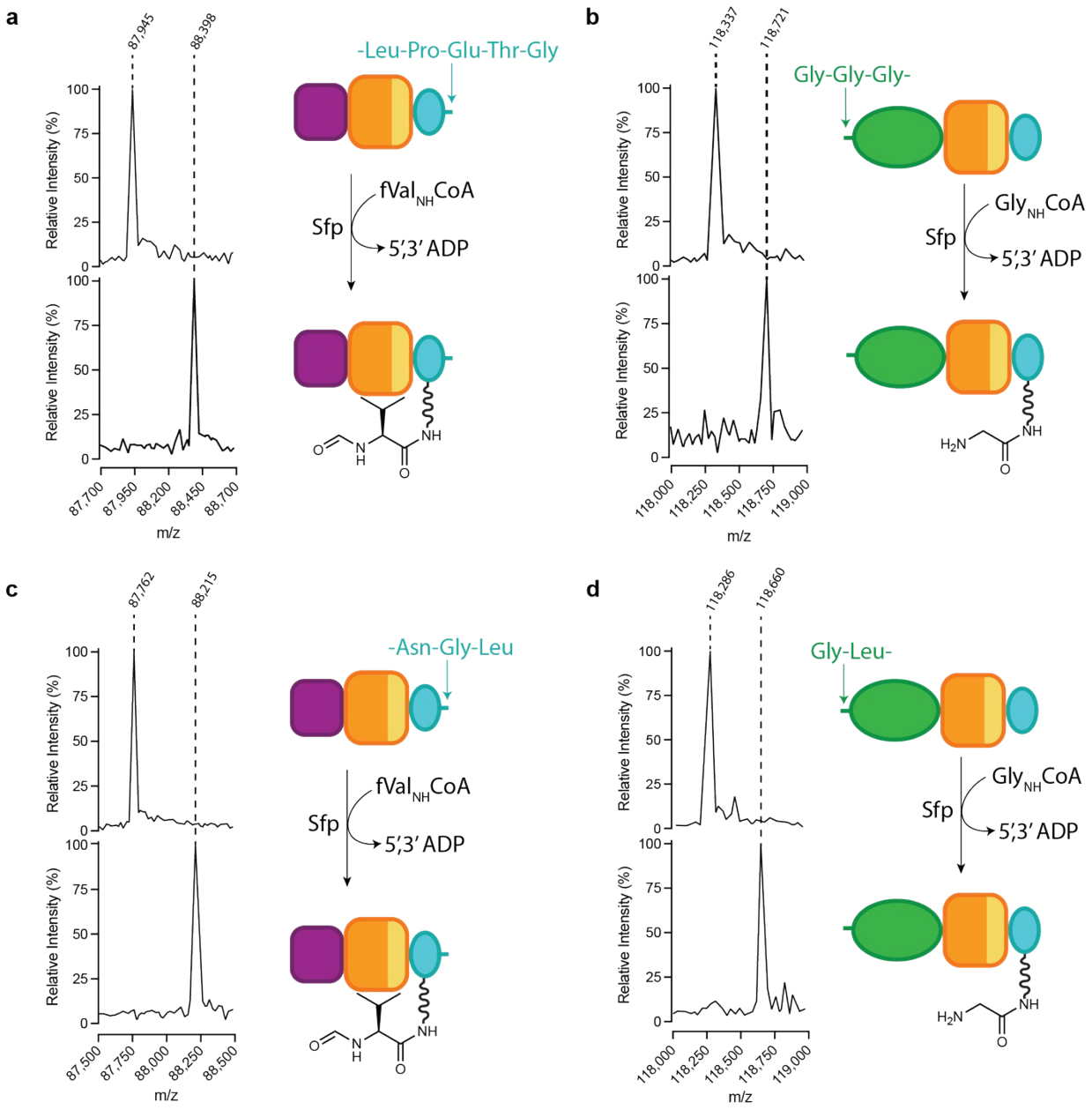


Supplementary Figure 5: SDS-PAGE of $F_1A_1T_1C_2A_2T_2$ formation by SaSrtA and OaAEP1 over time

a. SaSrtA-catalyzed ligation of $F_1A_1T_1^{Srt}$ and $^{Srt}C_2A_2T_2$ over 180 minutes. **b.** OaAEP1-catalyzed ligation of $F_1A_1T_1^{AEP1}$ and $^{AEP1}C_2A_2T_2$ over 120 minutes**.** Proteolysis is evident in the OaAEP1 reactions as increased laddering at 120 minutes. Each SDS gel in **a** and **b** is a replicate of the experiment (conducted in triplicates)**.** On the right side of each gel are known concentrations of $F_1A_1T_1C_2A_2T_2$ used to prepare a calibration curve for the conversion of band intensities to yields (µM).



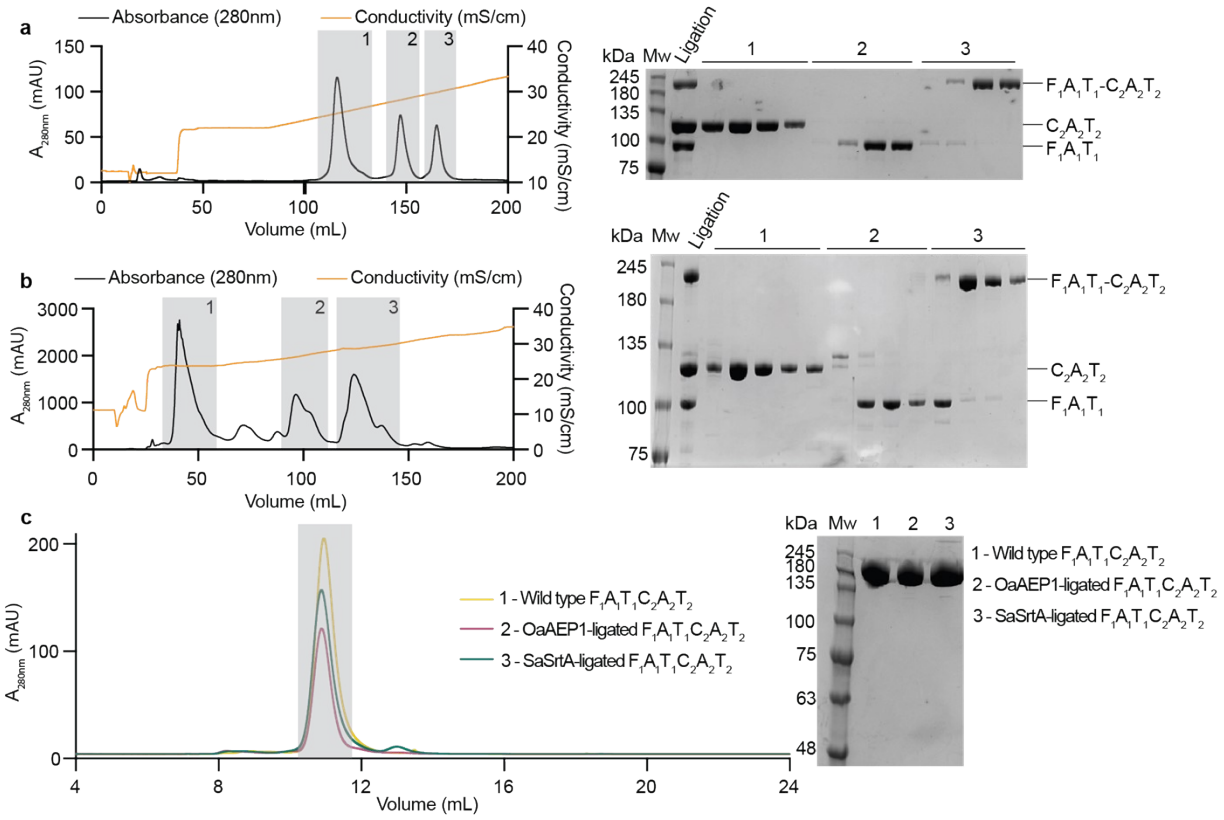
Supplementary Figure 6: Direct infusion of the non-hydrolyzable substrate analogue precursors **a.** $_{NH2\text{-}}$CoA, **b.** Gly$_{\text{-NH-}}$CoA and **c.** fVal$_{\text{-NH-}}$CoA were synthesized using a previously established protocol [2] and were confirmed by direct infusion onto an amaZon speed EDT (Bruker) ion trap mass spectrometer. Full characterization of these molecules by LC-MS and NMR was carried out in a previous study by our lab [3].



Supplementary Figure 7: Loading the amino acyl substrate analogues on the T domains of $F_1A_1T_1$ and $C_2A_2T_2$

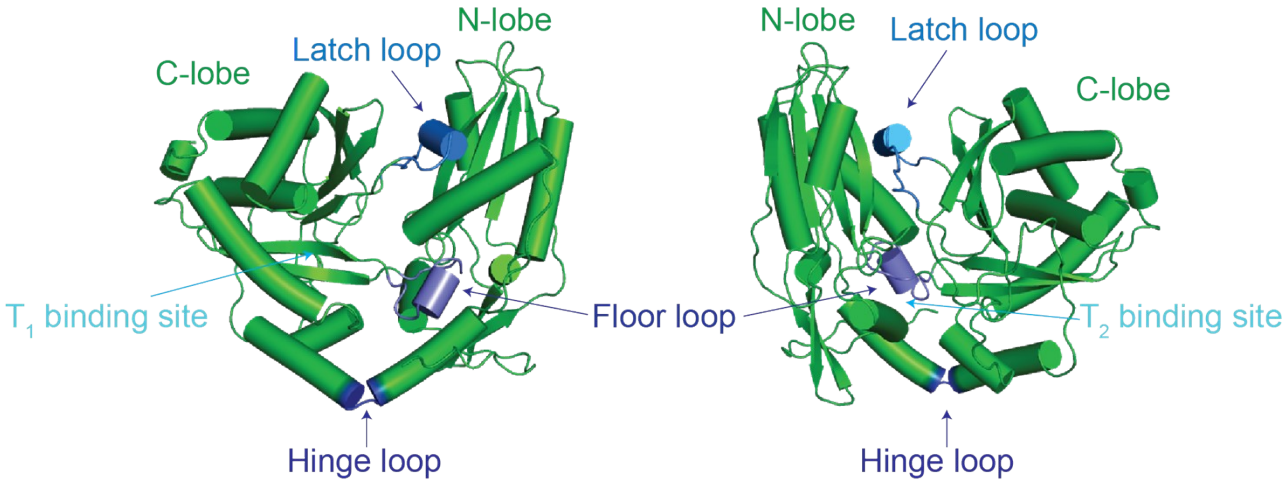
Intact protein LC-MS showing the loading of: **a.** fVal$_{-NH-}$ppant on $F_1A_1T_1^{Srt}$ (expected mass for apo: 87,962, observed mass for apo: 87,945, expected mass for modified: 88,414, observed mass for modified: 88,398), **b.** Gly$_{-NH-}$ppant on $^{Srt}C_2A_2T_2$ (expected mass for apo: 118,355, observed mass for apo: 118,337, expected for mass modified: 118,737, observed mass for modified: 118,721), **c.** fVal$_{-NH-}$ppant on $F_1A_1T_1^{AEP1}$ (expected m/z apo: 87,749, observed m/z apo: 87,762, expected m/z modified: 88,201, observed m/z modified: 88,215), and **d.** Gly$_{-NH-}$ppant on $^{AEP1}C_2A_2T_2$ (expected mass for apo:

118,268, observed mass for apo: 118,286, expected mass for modified: 118,650, observed mass for modified: 118,660).



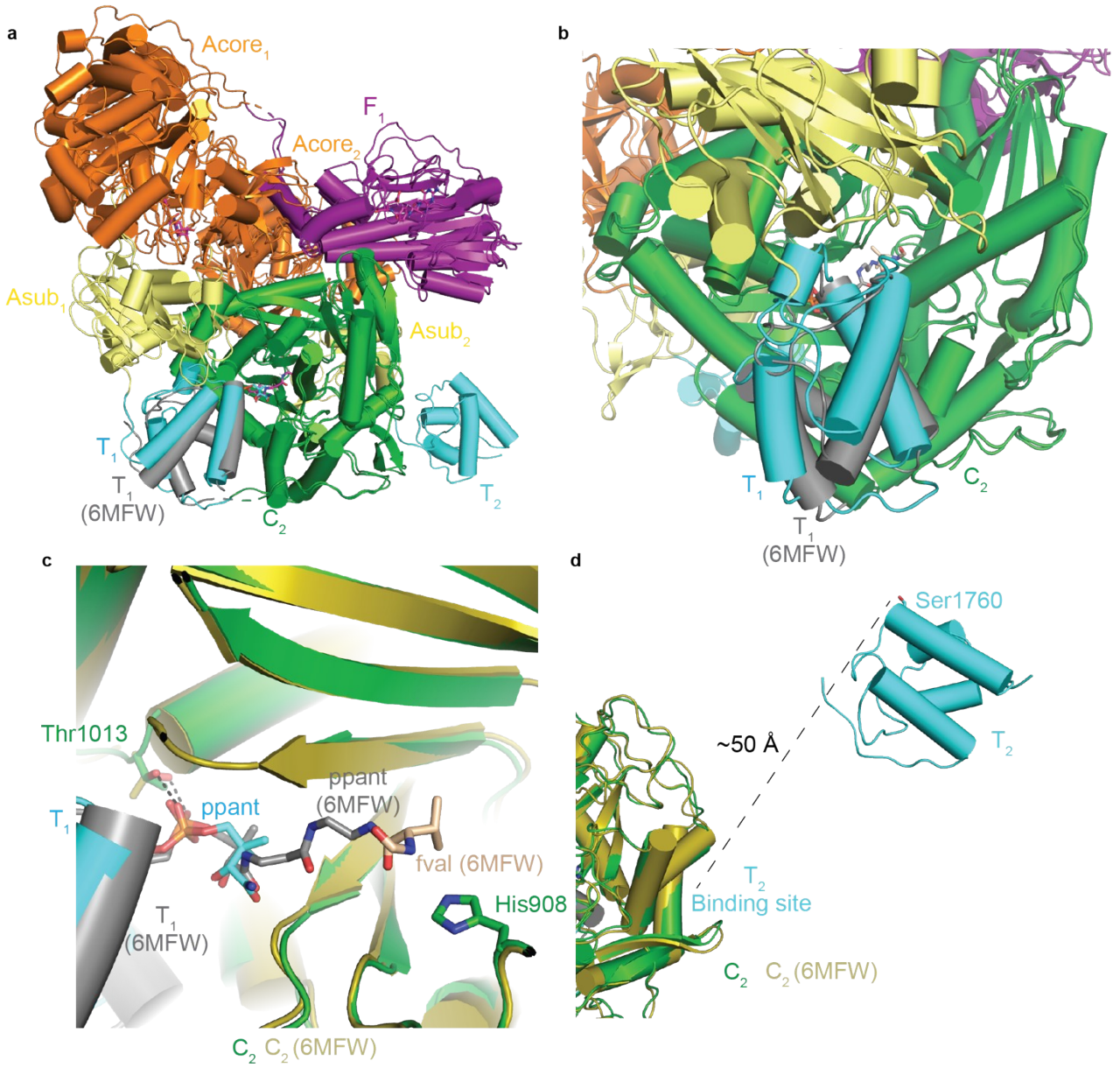
Supplementary Figure 8: Purification of the $F_1A_1T_{1\text{-fVal}}C_2A_2T_{2\text{-Gly}}$ complexes

a. Purification of SaSrtA-ligated $F_1A_1T_{1\text{-fVal}}$–$C_2A_2T_{2\text{-Gly}}$ by anion exchange chromatography. SDS-PAGE analysis of the elution peaks is shown on the right of the trace. **b.** Purification of OaAEP1-ligated $F_1A_1T_{1\text{-fVal}}$-$C_2A_2T_{2\text{-Gly}}$ by anion exchange chromatography. SDS-PAGE analysis of the elution peaks is shown on the right of the trace. In **a** and **b** $F_1A_1T_1$-$C_2A_2T_2$ elutes in peak 3 and $F_1A_1T_1$ and $C_2A_2T_2$ elute in peaks 2 and 1, respectively. **c.** Further purification of the ligated $F_1A_1T_{1\text{-fVal}}$-$C_2A_2T_{2\text{-Gly}}$ complexes by size-exclusion chromatography using a Superdex 200 Increase 10/300 GL (S200) column. SaSrtA-ligated (green line) and OaAEP1-ligated (magenta line) $F_1A_1T_{1\text{-fVal}}$-$C_2A_2T_{2\text{-Gly}}$ eluted as single peaks and eluted at the same volume as wild-type $F_1A_1T_1C_2A_2T_2$ (yellow line). SDS-PAGE (on the right of the S200 trace) of the elution peaks shows that $F_1A_1T_1$-$C_2A_2T_2$ complexes ligated by SaSrtA and OaAEP1 are highly pure.



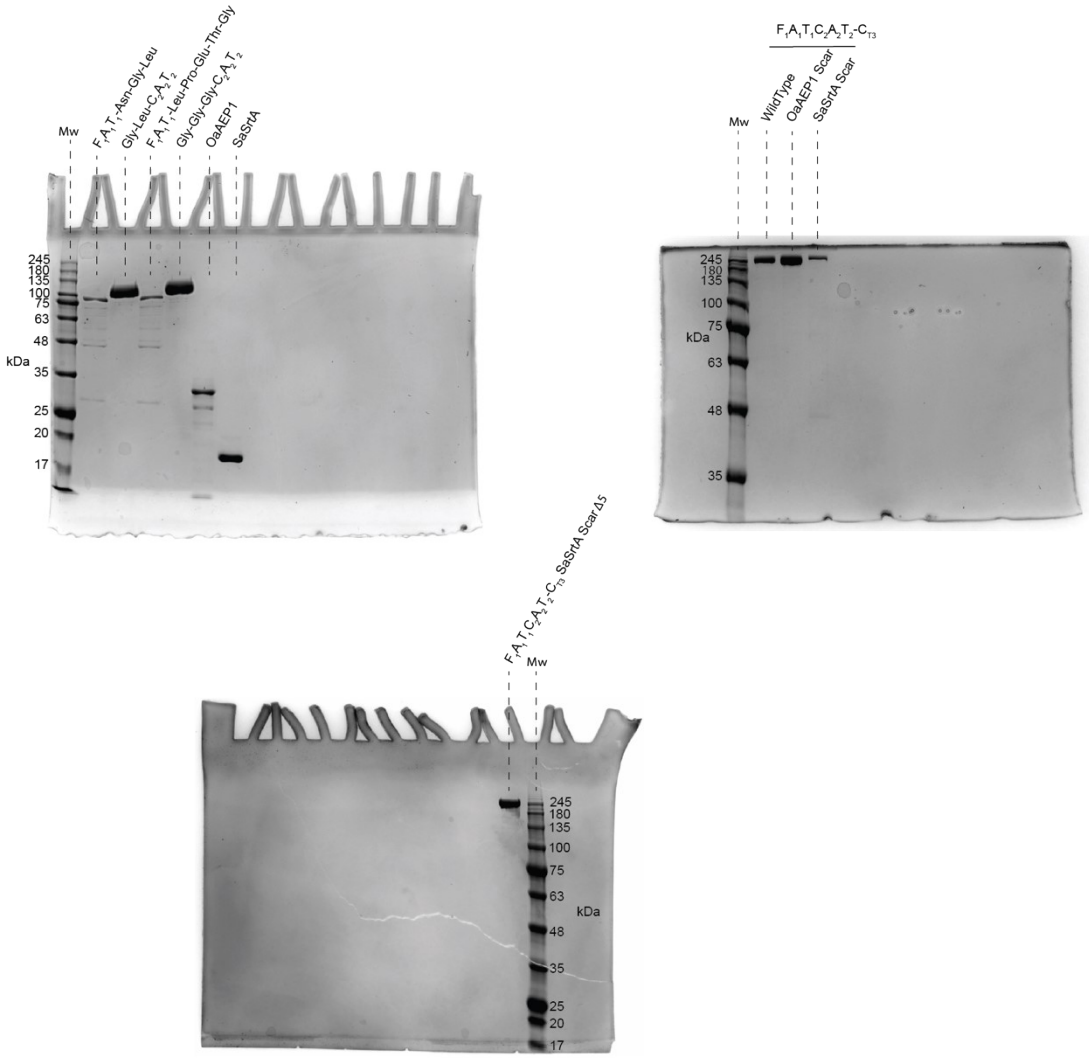
Supplementary Figure 9: Structure of the of $F_1A_1T_{1\text{-}fVal}C_2A_2T_{2\text{-}Gly}$ C domain

Architecture of C_2 domain as seen in the dimodular LgrA solved here. The left panel depicts the donor substrate binding face and the right panel depicts the acceptor substrate binding face. The N- and C-lobes are labelled, the latch loop is shown in blue, the floor loop in light purple and the hinge loop in dark blue.



Supplementary Figure 10: Comparison of $F_1A_1T_{1\text{-}fVal}C_2A_2T_{2\text{-}Gly}$ with the high resolution $F_1A_1T_{1\text{-}fVal}C_2$ structure

a. Alignment of the $F_1A_1T_{1\text{-}fVal}$-$C_2A_2T_{2\text{-}Gly}$ complex with the high-resolution $F_1A_1T_{1\text{-}fVal}C_2$ structure (PDB accession code: 6MFW) [2]. The four domains of $F_1A_1T_{1\text{-}fVal}C_2$ and $F_1A_1T_{1\text{-}fVal}$-$C_2A_2T_{2\text{-}Gly}$ assume a very similar conformation. **b**. Close-up view of T_1 from $F_1A_1T_{1\text{-}fVal}$-$C_2A_2T_{2\text{-}Gly}$ and $F_1A_1T_{1\text{-}fVal}C_2$ overlayed to highlight the similar T_1 orientation observed in the two structures. **c.** Zoom into the T_1 binding site showing the short portion of the ppant from the $F_1A_1T_{1\text{-}fVal}$-$C_2A_2T_{2\text{-}Gly}$ structure aligning well with the $fVal_{\text{-}NH\text{-}}$ppant substrate analogue observed in $F_1A_1T_{1\text{-}fVal}C_2$. T_1 ppant phosphate interacts with Thr1013 in the C-lobe. **d.** T_2 is ~50Å away from the expected binding site on the acceptor face of C_2.



Supplementary Figure 11: SDS gels showing all proteins used in this study.

Supplementary References

1. L. E. Quadri, P. H. Weinreb, M. Lei, M. M. Nakano, P. Zuber and C. T. Walsh, *Biochemistry*, 1998, **37**, 1585-1595.
2. J. M. Reimer, M. Eivaskhani, I. Harb, A. Guarne, M. Weigt and T. M. Schmeing, *Science*, 2019, **366**.
3. A. Pistofidis, P. Ma, Z. Li, K. Munro, K. N. Houk and T. M. Schmeing, *Nature*, 2024, DOI: 10.1038/s41586-024-08417-6.