A suite of kinetically superior AEP ligases can cyclise an intrinsically disordered protein

Karen S. Harris¹, Rosemary F. Guarino¹, Ravindu S. Dissanayake¹, Pedro Quimbar¹, Owen C. McCorkelle¹, Simon Poon¹, Quentin Kaas², Thomas Durek², Edward K. Gilding², Mark A. Jackson², David J. Craik², Nicole L. van der Weerden¹, Robin F. Anders¹, Marilyn A. Anderson^{1*}

¹La Trobe Institute for Molecular Science, Melbourne, Vic. 3086, Australia, ²Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072, Australia.

*To whom correspondence should be addressed, Tel: +61 3 9479 1255, E-mail: <u>m.anderson@latrobe.edu.au</u>

Supplementary Figures S1-S11, Tables S1-S2

Table S1. Percent identity matrix of OaAEP1 _b	OaAEP3, OaAEP4 and OaAEP5, as
determined by Clustal Omega.	

	OaAEP1b	OaAEP3	OaAEP4	OaAEP5
OaAEP1b	100	81.4	81.84	82.24
OaAEP3	81.4	100	94.2	94.07
OaAEP4	81.84	94.2	100	88.2
OaAEP5	82.24	94.07	88.2	100

OaAEP1b OaAEP4 OaAEP3 OaAEP5	↓ ¹ MVRYLAGAVLLLVVLSVAAAVSGARDGDYLHLPSEVSRFFRPQETNDDHGEDSVGTR ¹ MVRYPAGAVLLLVVLSV-VA-VDGARDGYLKLPSEVSDFFRPRNTNDGDDSVGTR ¹ MVRYLAGAFQVVLLVVILSDIAI-SEERTDGYLKLPTEVSRFFRTPEQSSDGGDDSIGTR ¹ MVRYLAGAFQVVLLVVILSDIAI-SEERTDGYLKLPTEVSRFFRTPEQSSDGGDDSIGTR	57 53 59 59
OaAEP1b OaAEP4 OaAEP3 OaAEP5	<pre>58 wavliagskgyanyrhqagvchayqilkrgglkdenivvfmyddiaynesnprpgviins ⁵⁴ wavllagsngywnyrhqadlchayqilkrgglkdenivvfmyddiayneenprpgviins ⁶⁰ wavliagskgydnyrhqadvchayqilkrgglkdenivvfmyddiaynesnprpgviins ⁶⁰ wavliagskgydnyrhqadvchayqilkrgglkdenivvfmyddiaynesnprpgviins</pre>	117 113 119 119
OaAEP1b OaAEP4 OaAEP3 OaAEP5	<pre>* * 118 PHGSDVYAGVPKDYTGEEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGA 114 PHGSDVYAGVPKDYTGDEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGA 120 PHGSDVYAGVPKDYTGEEVNAKNFLAAILGNKSAITGGSGKVVDSGPNDHIFIYYTDHGA</pre>	177 173 179 179
OaAEP1b OaAEP4 OaAEP3 OaAEP5	* [*] ¹⁷⁸ AGVIGMPSKPYLYADELNDALKKKHASGTYKSLVFYLEACESGSMFEGILPEDLNIYALT ¹⁷⁴ PGVIGMPSKPYLYADELNDALRKKHASGTYKSMVFYLEACEAGSMFDGLLPDGLNIYALT ¹⁸⁰ AGVIGMPSKPYLYADELNDALRKKHASGTYKSLVFYLEACEAGSMFEGLLTDDLNIYALT	237 233 239 239
OaAEP1b OaAEP4 OaAEP3 OaAEP5	238 STNTTESSWCYYCPAQ-ENPPPPEYNVCLGDLFSVAWLEDSDVQNSWYETLNQQYHHVDK 234 ASNTTEGSWCYYCPGQ-DAGPPPEYSVCLGDFFSIAWLEDSDVHNLRSETLNQQYHNVKN 240 ASNTTEGSWCYYCPGQ-DAGPPPEYSVCLGDFFSIAWLEDSDVHNLRSETLNQQYHNVKN 240 ASNATEGSCPYYCPGDLNYSPPPEYDVCLGDFFSIAWLEDSDVHNLRSETLNQQYHNVKN	296 292 298 299
OaAEP1b OaAEP4 OaAEP3 OaAEP5	296 RISHASHATQYGNLKLGEEGLFVYMGSNPANDNYTSLDGNALTPSSIVVNQRDADLLHLW 293 RISYASHATQYGDLKRGVEGLFLYLGSNPENDNYTFVDDNVVRPSSKAVNQRDADLVHFW 299 RISYASHATQYGDLKRGVEGLFLYLGSNPENDNYTFVDDNVVRPSSKAVNQRDADLVHFW 300 RISYASHATQYGDLKRGVEGLFLYLGSNPENDNYTFVDDNVVRPSSKAVNQRDADLVHFW	356 352 358 359
OaAEP1b OaAEP4 OaAEP3 OaAEP5	<pre>357 EKFRKAPEGSARKEVAQTQIFKAMSHRVHIDSSIKLIGKLLFGIEKCTEILNAVRPAGQP 353 EKFRKAPEGSSKKEEAQKQILEAMSHRVHIDSSINLIGKLLFGIEKGHKILTAVRSAGHP 360 EKFRKAPEGSSKKEEAQKQILEAMSHRVHIDSSINLIGKLLFGIEKGHKILTAVRSAGHP</pre>	416 412 418 419
OaAEP1b OaAEP4 OaAEP3 OaAEP5	417 LVDDWACLRSLVGTFETHCGSLSEYGMRHTRTIANICNAGISEEQMAEAASQACASIP 413 LVDDWACLRSLVRTFETHCGSLSQYGMKHTRTLANICNAGITEEQMAEAASQACVSIPSN 419 LVDDWACLRSLVRTFETHCGSLSQYGMKHTRTLANICNAGISEDQMKEAASQACASVPSN 420	474 472 478 479
OaAEP1b OaAEP4 OaAEP3 OaAEP5	473 PWSSHDGGFSA 479 PWSSHDGGFSA 480 SWSSLXKGFHARLAKIIA	483 489 497

Supplementary Figure S1: *Multiple alignment of AEPs using Clustal Omega.* Colour shows regions of identity. * denotes the catalytic dyad. ↓ denotes the putative signal peptide cleavage site and the first AEP residue of the recombinant protein. The final residue in each recombinantly expressed protein was Pro474 (OaAEP1_b), Ala489 (OaAEP3), Ala484 (OaAEP4) or Leu484 (OaAEP5). The recombinant AEPs are self-processed during incubation at low pH. ▼ and ∇ show the N- and C-terminal self-processing sites respectively that were previously identified experimentally for recombinant OaAEP1_b (1), defining the likely borders of the auto-activated proteins.



Supplementary Figure S2: Active site titration for quantitation of active enzyme. (a) Activity of $OaAEP1_b$ against an IQF peptide (Abz-STRNGLPS-Y(3NO₂)) in the presence of increasing concentrations of inhibitor. The assays were conducted in activity buffer (50 mM sodium acetate buffer, pH 5.0, up to 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) at room temperature. $OaAEP1_b$ is shown as a representative example. Other recombinant AEPs were quantitated in the same manner. (b) The initial rates (V_i) were calculated during the linear portion of the progress curve shown in (a). This initial rate was expressed relative to the initial rate of the no inhibitor control (V₀). V_i /V₀ was then plotted against inhibitor concentration. The x-intercept of the linear portion of the curve gives the concentration of active enzyme.

Table	S2. 3	Summary	data	showing	concentration	of	active	enzyme	determin	ed b	y active	site	titration
comp	ared	to concen	tratio	ns based	on BCA assay.								

Enzyme	Concentration (μM) (BCA assay)	Concentration (µM) (active site titration) (n≥3; ±sem)
OaAEP1b	31.25	6.17 ±0.135
OaAEP3	184	5.93 ±0.375
OaAEP4	9.375	0.33 ±0.04
OaAEP5	31.25	2.65 ±0.069



Supplementary Figure S3: *Products of enzymatic processing of the IQF peptide are cyclic.* Enzymes (OaAEP1_b, 60 nM; OaAEP3, 7.4 nM; OaAEP4, 6.6 nM; OaAEP5, 13 nM to reflect the concentrations used in the kinetic assays) were incubated with the R1 IQF peptide (30 μ M) for 30 min and the products were analysed by MALDI MS. The assays were carried out in activity buffer (50 mM sodium acetate buffer, pH 5.0, 0.5 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) at room temperature. The mass of the product observed is consistent with cyclised peptide. The observed average masses are listed (Da; [M+H]⁺). The expected average mass of cyclic product (cyc) is 3194.64 Da. The expected average mass of precursor (pre) is 3775.2 Da. -16 Da peaks present in the precursor and product spectra are likely to represent a synthesis-derived modification. The mass of the product is consistent with cyclic peptide indicating that the real-time fluorescence assay is providing a read out of cyclic peptide production. A very small peak +18 Da from the cyclic product is observed, consistent with linear product. A single representative experiment of two technical replicates is shown.



Supplementary Figure S4: Michaelis-Menten kinetics of recombinant AEP ligases processing the $R1_{IQF}$ peptide. A single representative Michaelis-Menten plot of at least three technical replicates is shown for OaAEP1_b, OaAEP3 and OaAEP5. The assays were carried out in activity buffer (50 mM sodium acetate buffer, pH 5.0, 50 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) The average kinetic parameters from all experiments is reported. It is clear that OaAEP1_b has a far higher K_m on this substrate that the other AEP enzymes. The very low K_m of OaAEP4 made determining accurate kinetic parameters challenging and estimates are reported in Table 1.

a S2' binding pocket residues

b Predictive residues for AEP ligase activity

residue # (as in OaAEP1 _b):):	ۍ، 6	<i>%</i>	`^%	⁵ ્ર્ઝ	20	200	⁵ (ર્ગ	૾૾ૺૺ૾ૢ	ڹ ٷؠۯ	b b b b b b b b b b b b b b b b b b b	3	× ~0
	OaAEP1 _b	:	K	D	K	D	С	Y	Q	N	V	HVDKRISH	E	G
Canonical AEP	OaAEP3 [~]	:	ĸ	D	K	D	С	Y	Q	A	V	NVKNRISY	V	G
ligases	OaAEP4	:	к	D	K	D	С	Y	Q	A	V	NVKNRISY	V	G
	OaAEP5	:	к	D	K	D	Ρ	Y	D	Y	V	NVKNRISY	V	G
	AtLEGY	:	D	D	S	N	G	т	Е	Ρ	т	LVKRRTAPV-GYSY	к	N
pH-dependent	RcAEP1	:	K	D	L	М	G	т	Μ	Ρ	т	SVKARTSNYNTYAA	Α	К
ligases	HaAEP1	:	D	D	Е	к	G	т	т	Ρ	т	KVKKRTSNSNTYNT	Ρ	L
	AtLEGß	:	Α	Α	Т	Α	G	т	М	Ρ	т	TVKMRTSNYNTYAA	Α	K
												MLA		

Supplementary Figure S5: Comparison of AEP protein sequences at ligase predictive residues and the S2' pocket. (a) AEP sequences in the S2' binding pocket, as identified in the AtLEGy structure (Zauner et al, *JBC*. 2018). Residue numbering is as for OaAEP1_b. Hydrophobic residues replaced with a charged His are shown in purple. (b) AEP sequences at the 12 ligase predictive sites and the MLA (Jackson et al, *Nat. Comms.* 2018). Residues in blue are typically identified in AEP ligases. Residues in magenta are typically identified in AEP proteases. The MLA typically harbours a deletion (or increased hydrophobicity) in AEP ligases.

OaAEP3





Supplementary Figure S6: Solvent tolerance of OaAEP3 and OaAEP5. Recombinant OaAEP3 (0.132 μ M) and OaAEP5 (0.132 μ M) were incubated with the R1 peptide (KL---NGL variant) (280 μ M) at pH 5 for 60 min in the presence of increasing concentrations of solvents. The assays were carried out in activity buffer (50 mM sodium acetate buffer, pH 5.0, 0.5 mM NaCl, 1 mM EDTA, 0.5 mM TCEP) at room temperature. The products were assessed by MALDI-MS. The proportion of cyclic product was determined relative to all peaks attributed to the processed or unprocessed target peptide. Reactions were stopped by heating at 70 °C for 5 min. The average of two technical replicates is shown and error bars report the range.



Supplementary Figure S7: *Size exclusion analysis of MSP2.* (a) MSP2_{AEP} and recombinant MSP2 lacking the AEP recognition site were analysed by size exclusion (superdex 75 column). The two versions of recombinant MSP2 had identical retention volumes and their apparent Mw was much higher than their predicted Mw of 25 kDa. The experimentally determined retention volumes of the globular proteins conalbumin (75 kDa, 9.4 mL), and cytochrome C (12 kDa, 13.4 mL) are shown for comparison. The void volume was 7.6 mL, as determined using blue dextran. (b) MSP2_{AEP} was processed by OaAEP1_b (0.68 μ M, 60 min) as in Fig. 5b and then applied to a superdex 200 size exclusion column to separate the precursor and cyclic product (i). Fractions were analysed by SDS-PAGE followed by Coomassie staining (ii). Fractions containing predominantly precursor or predominantly cyclic product were pooled.



Supplementary Figure S8: *AEP-mediated processing of an intrinsically disordered protein.* (a) SDS PAGE analysis of MSP2 (3 μ g) incubated with and without recombinant OaAEP1_b. The assays were conducted in activity buffer without reducing agent (50 mM sodium acetate buffer, pH 5.0, 50 mM NaCl, 1 mM EDTA). MSP2 only and enzyme only controls are also shown. A single representative experiment of two technical replicates is shown. (b) SDS PAGE analysis of MSP2 (3 μ g) incubated with recombinant OaAEP1_b, OaAEP3, OaAEP4 and OaAEP5 at 0.132 μ M for 30 min. Enzyme only controls are also shown. The grouping of different parts of the same gel is indicated by the white space. A single representative experiment of two technical replicates is shown. The gels are the same as those in Fig.5 b and c respectively and are displayed here in their entirety.



401 NKGTGQHGHM HGSRNNHPQN TSDSQKECTD GNKENCGAAT SLLSNSSNGL PSLAAHHHHH H

Supplementary Figure S9: Sequence coverage of backbone cyclised $MSP2_{AEP}$ after tryptic digestion and tandem MS. Two recombinant $MSP2_{AEP}$ sequences are listed consecutively. The first sequence lacks the leaving group released after enzyme processing (GLPSLAAHHHHHH, black underline in the second sequence) and is ligated to the N-terminus of a second MSP2 sequence This reproduces the sequence of backbone cyclised MSP2 across the cyclisation point (arrowed). The sequence coverage attained via tandem MS after tryptic digestion of the gel-extracted cyclic product is shown (blue underline). Multiple peptides spanning the cyclisation point were identified confirming the cyclic nature of the extracted protein.



Supplementary Figure S10: Cyclic $MSP2_{AEP}$ has greater electrophoretic mobility than linear $MSP2_{AEP}$. (a) Linear precursor (Lin) and putatively cyclic (Cyc) versions of $MSP2_{AEP}$ were separated by size exclusion (superdex 200, 16/60) and analysed by SDS-PAGE. The grouping of different parts of the same gel is indicated by the white space. (b) The mass of the putative cyclic $MSP2_{AEP}$ was analysed by ESI-MS. The observed dominant average mass of 23,115.0 Da was consistent with the expected average mass of 23,115.0 Da of cyclic product with a single disulfide bond.



Supplementary Figure S11: Circular dichroism spectroscopy of cyclic $MSP2_{AEP}$ in aqueous solution. Difference in mean residue ellipticity, $\Delta\epsilon(M^{-1}cm^{-1})$, plotted as a function of wavelength (nm) for cyclic $MSP2_{AEP}$ solubilised in 10 mM sodium acetate pH 6.2 at a concentration of 0.2 mg mL⁻¹. The nonlinear least squares best fit using the CONTINLL algorithm with the SDP48 database is shown. The spectra is consistent with a predominantly disordered protein with a small portion of secondary structure.

References:

1. Harris, K. S., Durek, T., Kaas, Q., Poth, A. G., Gilding, E. K., Conlan, B. F., Saska, I., Daly, N. L., van der Weerden, N. L., Craik, D. J., and Anderson, M. A. (2015) Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. *Nat. Commun.* **6**, 10199