# The mechanism of patellamide macrocyclization revealed by study of the *Prochloron sp* PatG macrocyclase domain

Jesko Koehnke<sup>1,2</sup>, Andrew Bent<sup>1,2</sup>, Wael E. Houssen<sup>2,3,4</sup>, David Zollman<sup>1</sup>, Falk Morawitz<sup>1</sup>, Sally Shirran<sup>1</sup>, Jeremie Vendome<sup>5,6</sup>, Ada F. Nneoyiegbe<sup>3</sup>, Laurent Trembleau<sup>3</sup>, Catherine H. Botting<sup>1</sup>, Margaret C. M. Smith<sup>3</sup>, Marcel Jaspars<sup>3</sup>, & James H. Naismith<sup>1</sup>

<sup>1</sup> Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Fife KY16 9ST, Scotland, U.K. <sup>2</sup> Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Meston Walk, Aberdeen AB24 3UE, Scotland, U.K. <sup>3</sup> Institute of Medical Sciences, School of Medicine and Dentistry, University of Aberdeen, Ashgrove Road West, Aberdeen AB25 2ZD, Scotland, U.K. <sup>4</sup> Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA. <sup>5</sup> Howard Hughes Medical Institute, Columbia University, New York, NY 10032, USA.

<sup>2</sup> These authors contributed equally to the work.

### **Correspondance to**

#### James H. Naismith

naismith@st-andrews.ac.uk

or

Marcel Jaspars

m.jaspars@abdn.ac.uk



(a) PatGmac (cyan) with macrocyclization insertion (magenta) superposed onto AkP (orange) (PDB 1DBI). The loop of AkP that contains the macrocyclization insertion in PatGmac is shown in yellow-green, active site serines are shown as spheres.



(b) Close-up of the active sites of PatGmac (cyan, magenta) and AkP (orange) with active site residues shown as sticks (PatGmac: yellow, AkP: light blue).



(c) Difference in PatGmac and AkP disulfide bonds. PatGmac Cys685 and Cys724 (yellow sticks) link two loops, while in AkP Cys137 and Cys139 (light blue sticks) form an intra-loop ring. Active site serines are shown as sticks in the background.

	α1	β1 📩	β2	
PatGmac TenG TruG MicroAPCC7806 MicroANIES298 LynPCC8106 TricholMS101 OsciPCC6506 ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca PiankNIES596	PGGC C C C C C C C C C C C C C C C C C C	PORTIVIAL DGRADLE PRIMINAL PRIMINAL PRI	H H H H H H H H H H H H H H H H H H H	P A E P     576       P A E P     665       P A D P     670       P A D P     708       A I D D     708       A I D D     583       D I E L N D E Y F     62       D I E L N D E Y Y     62
PatGmac TenG TruG MicroAPCC7806 MicroANIES298 LynPCC8106 TrichelMS101 OscIPCC8506 ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca PlankNIES596	С С С С С С С С С С С С С С	2 SIRDKGLKGKE ERDKGLKGKE ERDKGLKGKA ARDOGLKGKA ARDOGLKGKA EDNGNLKGEA EDNGNLKGEA EDNGNLKGEA EDNGNLKGEA C	CAN A LEAN A LEA	a4     +     623       IND HAC HVT     713       IND HAC HVT     713       IND HAC HVT     713       IND HAC HVT     717       IND HAC HVT     717       IND HAC HVT     725       GD YHAC HVT     725       GA FHATHY F     105       GD NHAC HVT     739       GD NHAC HVT     739       GD NHAC HVT     139       GD NHAC HVT     632       INS HATHY S     118       IS SHATHY S     118       IS SHATHY S     118
PatGmac TenG TruG MicroAPCC7806 MicroANE5298 LynPCC8106 TricholMS101 OsciPCC6506 ArthroParaca ArthroParaca ArthroNE539 NodulCCY9114 Anab90 PlankNIE5596	04 5 T V COULE F SPVFFGI ST V F SPVFFGI ST V COULE F SPVFFGI ST V F SPVFFGI S	β3 A P R C C V I NM P H D A V V A P R C C V I NM P H D A V V A P R C C V I NM P H D A V V A P R C C V I NM P H D A V V A P R C C V I NM P H D A V V A P R C C V I NM P H D A V V A P R C C V I NM P H D A V V A P R C C V I NM P I N S F A C A P R C C V I N N P I S F A C A P R C C V I N P I S F A C A P R C C V I N P I S F A C A P R C C V I N P I S F A C	RG	α5 LNLARATEL ARAFEL 769 LNLARAFEL 761 LNLARAFEL 840 INLARAFEL 840 840 INLARAFEL 8
PalGmac TenG MicroAPCC7806 MicroAPCC7806 UpPCC8106 TricholMS101 OsciPCC6506 ArthroParaca ArthroNE4114 Anab30 PlankNiE5596	CA A L BL G A N I H C A F C A L BL G A N I H C A F C A L BL G A N I H C A F C A L BL G A N I H C A F C A L BL G A N I H C A F C A L BL G A N I H C A F C A L BL G A N I H C A A C A L BL G A N I H C A A C A L BL G A N I H C A A C A L BL G A N I H C A A C A L BL G A N I H C A A C A L BL G A N I H C A A C A L C A C C A L B C A N I H C A A C A L C A C C A N I H C A A C A L C A C C A N I H C A A C A L C A C C A N I H C A A C A L C A C C A C C A L C A C C A C C A L C A C C A C C A L C A C C C A C C A L C A C C C A C C A L C A C C C C C C C C C C A L C C A C C C C C C C C C C C C C C C		16 β5 K K C O D N N Y L I V S P T G N N V K C C D D N N Y L I V S P T G N N V K C C D D N N Y L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C C D N N L I V S P T G N N V K C C D D N N L I V S P T G N N V K C C D N N M L V S P G N D V K C C D N N M L V S P G N D V K C C D N N M L V S P G N D V K C C D N N M L V S P G N D V K C C D N N M L V S P S N D V K C C D N N M L V S P S N D V K C C D N N M L V S P S N D V K C C D N N L V S P S N D V K C C D N N M L V S P S N D V K C C D N N M L V S P S N D V K C C D N N L V S P S N D V K C C D N N S N D S N S N S N S N S N S N S N	SNESWCLPA     727       SNKWCLPA     825       LGECWCNPA     826       LGECWCNPA     836       KGECWCNPA     896       KGECWCNPA     897       KGECWCNPA     896       KGECWCNPA     897       KGECWCNPA     897       KGECWCNPA     897       KGECWCNPA     897       KGECWCNPA     896       KGECWCNPA     897       KGECWCNPA     896
PatGmac TenG MicroAPCC7806 MicroANIES298 LynPCC8106 TricholMS101 OsciPCC8506 ArthroParaca ArthroNeS414 Anab30 PlankNiES596			β7 GILAPGEEILGAOPCTTE GILAPGEDVLGAOPYTTE GILAPGEDVLGAOPYTTE GILAPGEDILGAOPCTTE GILAPGEDILGAOPCTTE GILAPGEDILGAOPCTTE GILAPGEEILGAOPCTTE GILAPGEEILGAOPTA GILAPGEEILG	β9     α τ       K P V R L T G S S     881       K P V R L G S S     881       K P V R L G S S     884       K P V R L G S S     884       K P V R L G S S     884       K P V R L G S S     884       K P V R L G S S     825       E V R L G S S     721       A P V R L G S S     721       A P V R L G S S     721       A P V R L G S S     721       A P V R L G S S     721       P V R L G S S     724
PatGmac TenG TruG MicroAPCC7806 MicroAPCC7806 MicroAPCC8186 TricholM8101 OsciPCC6506 ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca ArthroParaca	47 C A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I A A L L M M A A P V M T G I A A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M M A A P V M T G I S A L L M		CAS	PERCLECE     839       PERCLECE     9401       DECE     9401       DECE     9401       DECE     9401       DEE     9401       DE     9401  9401     9401       DE<
PatGmac TenG TruG MicroAVIES298 LynPCC8106 TricholMS101 OsciPCC6506 ArthroParaca ArthroParaca ArthroPasac NodulCCV9114 Anab90 PlankNIES596	C <sup>1</sup> NIPGANKYLEGO 8 NIPGANKYLEGO 8 NIPGANKYLEGO 8 NIPGANKYLEGO 8 NIPGANYULEGO 8 NIPGANULLEGO 8 NIPGANULLEG 8 NIPGANULLEG 8 NIPGANULLEG 8 NIPGAS	11 19 12 12 12 10 10 10 10 10 10 10 10 10 10 10 10 10		

#### Figure S2 Sequence alignment of PatGmac with its homologs

Secondary structure elements are shown in red. Active site residues are indicated by yellow stars, cysteines involved in disulfide bonding as green triangles (matching directions represent disulfide pairs), residues blocking the S3 and S4 sites as blue diamonds, lysines forming salt-bridges with the substrate as purple circles and His and Phe residues involved in substrate binding are marked by a magenta box.





## (a) Fragmentation pattern of cyclo[VGAGIGFP].



(b) LC-MS of macrocyclization reactions with PatGmac $\Delta 1$  (i), PatGmac K598D (ii) and PatGmac triple mutant R589D K594D K598D (iii). Only linear product is observed (curved lines). The error between observed and calculated mass is shown below the  $[M + H]^+$  species.







Figure S5 Macrocyclization of VGAGIGFPAYRG

(a) Ion counts of VGAGIGFPAYRG processed by PatGmac wild-type and PatGmac K598D for linear and macrocyclized products as determined by LC-MS.



## (b) LC-MS of VGAGIGFPAYRG macrocyclization.