

Supplementary Information

Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module

Jack Davison¹, Jonathan Dorival¹, Hery Rabeharindranto^{1,2}, Hortense Mazon¹, Benjamin Chagot^{1,*}, Arnaud Gruez^{1,*} & Kira J Weissman^{1,*}

¹Molecular and Structural Enzymology Group, Université de Lorraine, IMoPA, UMR 7365, Vandœuvre-lès-Nancy, F-54500, France. CNRS, IMoPA, UMR 7365, Vandœuvre-lès-Nancy, F-54500, France.

²Present address: Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés – INSA, UMR INSA/CNRS 5504 – UMR INSA/INRA 792, 135 Avenue de Rangueil, 31077 Toulouse CEDEX 4, France.

*Correspondence should be addressed to B.C. (benjamin.chagot@univ-lorraine.fr), A.G. (arnaud.gruez@univ-lorraine.fr) or K.J.W. (kira.weissman@univ-lorraine.fr)

Supplementary Figures

a

| | | | |
|-----------|---|---|------------------------------|
| CurL | 1 | ----- | LP |
| CurI | 1 | ----- | MGWKSVA |
| CurJ | 1 | ----- | ILSSNSTA |
| CurG | 1 | ----- | VKVGLEFRTEIDQ |
| CurH | 1 | ----- | IP-IEFSSNLDE |
| CurK | 1 | ----- | MP-IDFSSHLDQ |
| RhiC | 1 | -----HRAELATW-----LNGVHQ-----AGSAARGAPK-PY-RQF----- | |
| MmpA | 1 | ----- | |
| CurA | 1 | -----IAAT-----GSKLF-----QNGNGGHPQ-ES----- | |
| JamE | 1 | -----ITAT-----GSSKLP-----QNGNGGYSQ-ES-----S | |
| StiB | 1 | ----- | LSREIPT-ATSEASPPTRT |
| TaiE | 1 | -----LGARLPAAQAARA-----G-AF-AGVEPGEFDARA-PA-AA-RAAAP----- | |
| Tal | 1 | ----- | IQK-TV-SAGA |
| AjuA | 1 | ----- | APTANGATV-RGPERT-----TTS |
| ElaK | 1 | ----- | MTN |
| AjuE | 1 | -----IATV-----SAGHDDAREPQE----- | |
| VirA | 1 | ----- | AAPVPA-----AAEPAAL |
| SnaE2 | 1 | ----- | TAPQ |
| TaiL | 1 | LALARRRAGGAPLPDAEAASATTATTASASGTAHAPNAAGDAHAPNAAGARA-ATPAPAENAAGAVHA-ARAANPAN | |
| VirFG | 1 | ----- | |
| StiA | 1 | ----- | GIQLNGVEE-----PP-QAVE-SPSV-S |
| StiE | 1 | ----- | GLVLSVAEA-----PS-QAAEQSGSRAA |
| StiG | 1 | ----- | -----AIHSAAPEN |
| AjuB | 1 | ----- | AEPV-----PA-TSSPTL--IAV |
| AjuG | 1 | ----- | APT-----PE-TAAPPSPDASS |
| consensus | 1 | ----- | .. |

| | | | | | |
|-----------|----|--|------------|------------|--|
| | | | α_1 | α_2 | |
| CurL | 3 | ---EG---QE ^{EE} EVDL ^{LD} DSQ---NTQNL ^{LS} SAQIE ^{IQ} CK ^{VP} HAI ^A AEE ^{-LQ} E ^E KNL ^L KEGN--- | | | |
| CurI | 10 | ---ESN---I ^{PE} ---RKEVDVDEQ ^{TL} LPV ^{IED} ISE ^{EE} EE ^{EE} ALAA ^{QQ} -LEK ^L KSMI--- | | | |
| CurJ | 9 | ---ASE---AETL ^{DM} V ^{EG} SLESEL ^{SE} SLAAE ^{IN} QLS ^{ED} EM ^D LAVS ^{QA} -V ^{SQ} L ^D QL--- | | | |
| CurG | 13 | ---D---QE ^{DD} SI ^L ---EAK ^{LD} E ^L K ^Q SS ^N OE ^{LE} SS ^{IQ} I ^L EST ^{IN} --- | | | |
| CurH | 12 | ---IEN---LE ^{QE} EID ⁻ ---N ^{TF} R ⁻ ---F ^Q EM ^{SE} DE ^{MA} NI ^L ARK ^{-LE} S ^{EG} KKS--- | | | |
| CurK | 12 | ---TEK---ALDEE ^{DV} AD ⁻ ---GS ^Q LS ^D TE ^{LS} E ^{LE} AS ^{VL} Q ^E -IEA ^L E ^K LI--- | | | |
| RhiC | 30 | ---EP ^H PLD ^H Q ^L E ^K Q ^F LS ^G ---E ^L S ^D SL ^N LV ^S I ^G T ^V R--- | | | |
| MmpA | 1 | ---V ^Q V ^V H ^{SS} ---D ^D L ^L D ^T I ^L N ^Q V ^Q S ^G ALA ^{AD} S ^{AC} A ^V VED ^V ---LA ^{QR} --- | | | |
| CurA | 23 | ---I ^P N ^{NG} S ^P V ^N S ^S E ^{EM} DP ⁻ ---K ^L R ^L R ^A L ^L N ^K V ^A K ^E L ^T I ^Q E ^A N ^K L ^V Q ^Q I ⁻ ---K ^K Q ^V T ^V --- | | | |
| JamE | 24 | ---S ^E N ^H A ^S Q ^G D ^S E ^{EM} S ^Q ---K ^V R ^L R ^A L ^L N ^K V ^A N ^N E ^L T ^V Q ^E A ^N Q ^L V ^Q Q ^L ---K ^Q Q ^V T ^V --- | | | |
| StiB | 19 | K ^D S ^A P ^T I ^E P ^Q V ⁻ DA ^L S ⁻ ---G ^T A ⁻ ---I ^A E ^L F ^E Q ^L S ^A I ^D A ^L ---I ^D N ⁻ T--- | | | |
| TaiE | 39 | ---FA---HT---D ^A A ^A ---H ^T D ^T D ^A L ^L R ^A I ^E R ^G E ^L D ^A G ^D A ^D A ^L W ^{RR} ---V ^Q S ^R A ^A R ^E P ^E L ^A Q ^F --- | | | |
| Tal | 10 | ---P---A ^Q G ^E ---P ^P S ^L D ^E L ^L R ^Q V ^E A ^G E ^L D ^P S ^V A ^Q Q ^L T ^N ---S ^Q S--- | | | |
| AjuA | 21 | ---VR---S ^V L---G ^E F ^L D ⁻ ---L ^P A ^E L ^S D ^S E ^A E ^L L ^V E ⁻ ---E ^L E ^R L ^N Y--- | | | |
| ElaK | 4 | ---I ^S T ^Y E ^P V ^Q ---S ^L Q ^Q ---E ^F S ^L D ^D V ^V R ^Q V ^E K ^E M ^D A ^E A ^A N ^L L ^A V ⁻ ---S ^K Q ⁻ --- | | | |
| AjuE | 17 | ---SR ^D ---R ^A A ^L D ^A A ^L H ^E V ⁻ ---E ^Q L ^S E ^D D ^A L ^A A ^L M ^K ---G ^T --- | | | |
| VirA | 15 | ---P ^L P ^V T ⁻ G ^A ---G ^E P ^S ---Q ^A D ^L D ^A L ^L S ^A V ^R ---N ^R L ^S I ^E Q ^A V ^T L ^L T ^P ---R ^R --- | | | |
| SnaE2 | 5 | ---P ^A G ^A V ⁻ P ^A ---G ^S V ^S ---A ^A D ^L D ^A L ^L A ^A V ^R ---N ^R L ^T V ^E Q ^A L ^L L ^P Q ⁻ ---H ^T --- | | | |
| TaiL | 76 | ---P ^A T ^L A ⁻ N ^P ---A ^P G ^G ---G ^V E ^L D ^D V ^L A ^R V ^H R ^G E ^L S ^V E ^A E ^A L ^L A ^G ---A ^L G ⁻ --- | | | |
| VirFG | 1 | ---M ^D A ^K E ^I L ^T R ^F K ^D G ^G L ^D R ^A A ^A Q ^A L ^L A ^G ---R ^T --- | | | |
| StiA | 21 | E ^E D ^G A ^A M ^L K ^M L ⁻ E ^Q F ^Q ---E ^L S ^N E ^D L ^L G ^M L ^S D ⁻ D ^K E ^D ---E ^V S ^S --- | | | |
| StiE | 23 | G ^D L ^E A ^L A ^E G ^L L ⁻ D ^E L ^E ---G ^L S ^D D ^E S ^G ---R ^L A ^S G ^T A ^Q R ^S P ^R E ^D I ^N E | | | |
| StiG | 10 | T ^P V ^G N ^P S ^R G ^I P ⁻ D ^E L ^E ---G ^L P ^E E ^E L ^L A ^L F ^R ---E ^M A ^A L ^E R ^E T ⁻ S ^G E ^K S--- | | | |
| AjuB | 16 | T ^P S ^E V ^T A ^L P ^A L ⁻ P ^S L ^E ---N ^L S ^E A ^E L ^S D ^L L ^A A ⁻ E ^L S ^A A ^L M ⁻ G ^P G ^M D ⁻ --- | | | |
| AjuG | 17 | A ^P S ^E T ^G S ^E T ^A P ⁻ D ^P L ^D ---A ^L T ^E D ^E L ^V A ^L L ^A A ⁻ E ^L W ^P A ^R A ^G L ⁻ S ^D A ^R S ^G T ^L S ^E G ^P Q ^R E ^P H ^T A ⁻ --- | | | |
| consensus | 81 | | H H H | C | |

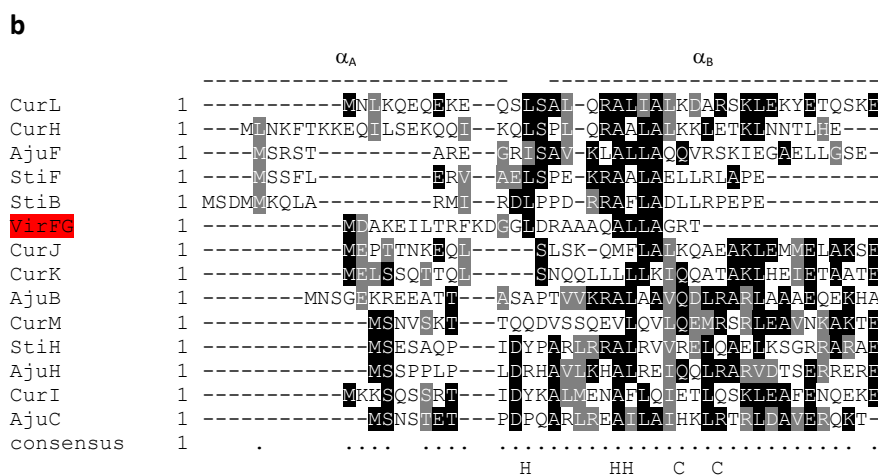


Fig. S1 Sequence analysis of putative dimerization/docking regions in various PKS. (a) Multiple alignment of the C-terminal regions of PKS subunits from *trans*-AT PKS (in red), with those of *cis*-AT PKS¹ (black, except for those of subunit CurA and the closely-related JamE (in pink)). The sequences from the *trans*-AT PKS (with the exception of VirFG) shown lie downstream of the conserved sequences of single (Ta1) or tandem acyl carrier protein (ACP) domains associated with β -modification reactions, which are located at the ends of subunits. The VirFG sequence represents the N-terminus of module 6, which we propose to be a docking element interacting with the C-terminus of VirA. The *trans*-AT PKS regions were identified manually, as BLAST analysis using the VirA region did not yield any homologous sequences. The positions of the two α -helices in the solved structure of the CurG and CurK C-terminal docking domains are indicated, as are the hydrophobic (H) and charged (C) residues which contribute to the docking interface.¹ (b) Multiple alignment of the N-terminal regions from the partner *cis*-AT PKS subunits to those shown in (a), alongside that of VirFG (in red). The positions of the two α -helices in the solved structure of the CurH and CurL N-terminal docking domains are indicated, as are the hydrophobic (H) and charged (C) residues which contribute to the docking interface.¹ Key: Cur, curacin (*cis*-AT PKS); Rhi, rhizoxin (*trans*-AT); Mmp, mupirocin (*trans*-AT); Jam, jamaicamide (*cis*-AT); Sti, stigmatellin (*cis*-AT); Tai, thailandamide (*trans*-AT); Ta, myxovirescin (antibiotic Ta; *trans*-AT); Aju, ajudazol (*cis*-AT); Ela, elansolid (*trans*-AT); Vir, virginiamycin (*trans*-AT); Sna, pristinaamycin (*trans*-AT).

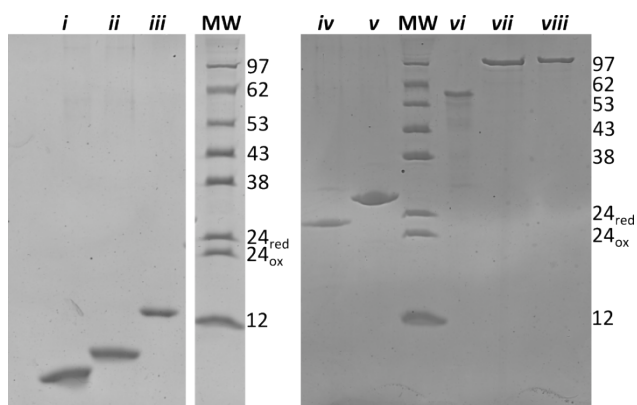


Fig. S2 SDS-PAGE analysis of protein preparations used in this study. Key to lanes: i) ACP_{5a} (calc'd: 8.9 kDa); ii) ACP_{5b} (calc'd: 9.2 kDa); iii) ACP_{5b}-DD (calc'd: 14.3 kDa); iv) ACP_{5a}-ACP_{5b} (calc'd: 19.4 kDa); v) ACP_{5a}-ACP_{5b}-DD (calc'd: 24.5 kDa); vi) KS₅-linker (calc'd: 62.2 kDa); vii) KS₅-ACP_{5a}-ACP_{5b} (calc'd: 98.5 kDa); viii) KS-ACP_{5a}-ACP_{5b}-DD (calc'd: 103.6 kDa). The molecular weights of the markers are indicated.

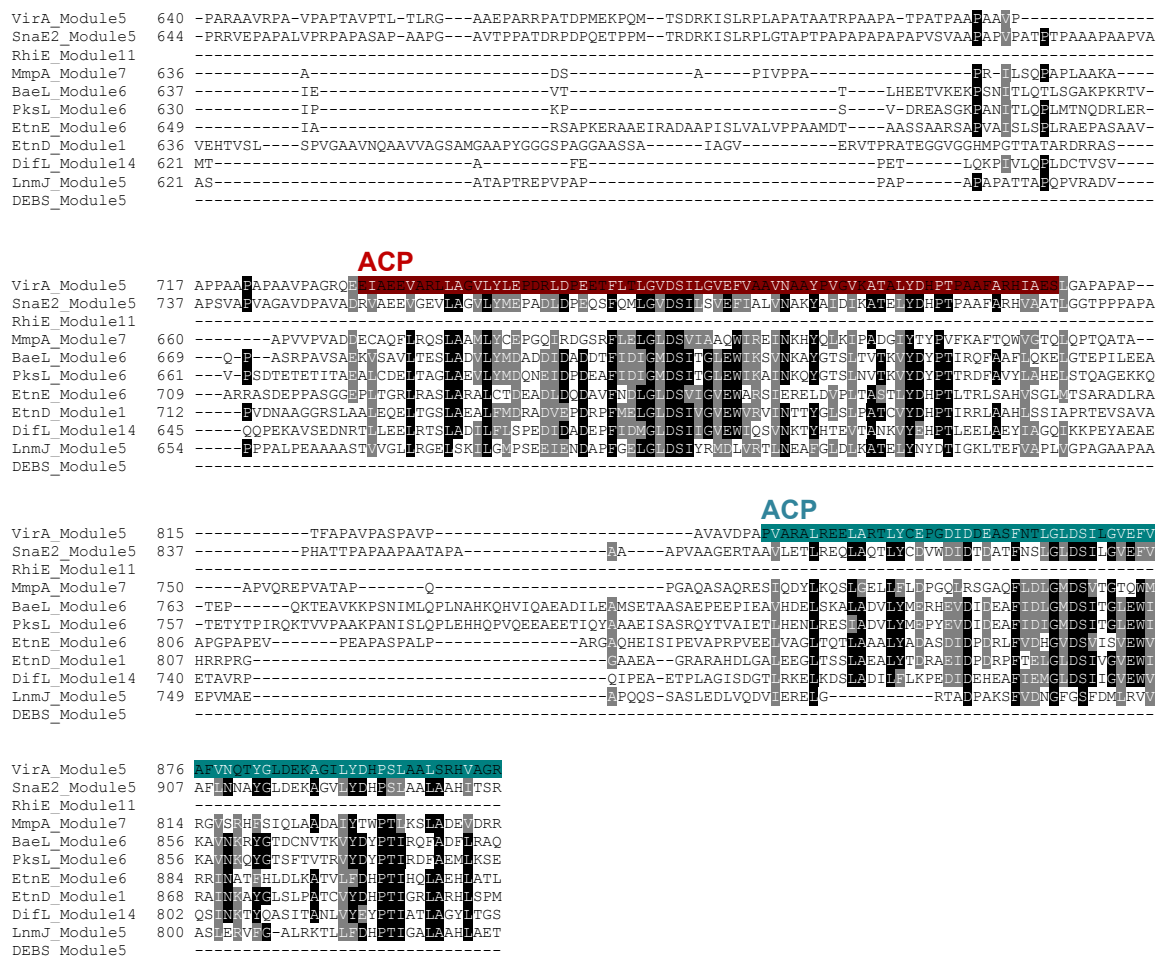


Fig. S3 Sequence analysis of a selection of modules (type KS-ACP-ACP and KS-ACP-ACP-ACP) from *trans*-AT PKS showing the homology to the KS-AT and post-AT linker regions of a typical *cis*-AT PKS module. The sequence of the KS-AT and post-AT linkers from module 5 of the erythromycin (DEBS) *cis*-AT PKS, for which structures are known,² are shown within and above the alignment (in blue and green, respectively), while the corresponding regions from the recently-published KS-B didomain structure of RhiE³ are shown in light blue and olive. Comparison of the KS-AT linker regions between the VirA, RhiE and DEBS modules shows that the RhiE sequence contains a large insertion. The positions of the conserved domains (based on the structure of the KS₅-AT₅ didomain of DEBS² and the NMR structures of VirA ACP_{5a} and ACP_{5b} (this work)) within the modules are also indicated in color. Key: Dif, diffidicin; Vir, virginiamycin; Sna, pristinemycin; Mmp, mupirocin; Bae, bacillaene from *Bacillus amyloliquefaciens*; Pks, bacillaene from *B. subtilis*; Lnm, leinamycin; Etn, etnangien. In each case, the subunit name and module number are indicated.

a

ACP

```

MmpB_Module8_ACP-ACP-ACP_2 1 AALLDEIVALLCOQLRTVAGIDPHPTDLHDGFDSSLLTQLLAQSSSYGVLEDFEGSVLEDAIACGLVQVQQRHGAEPAASLRVPQVQ--ERRA-----
MlnB_Module2 1 SEKSTLREITGVLKTEERFDFEPISEYGFDSIMFTMGELNETLGLSMPDDEGLTDFNLELEKSEYSEHLGCSFSSDE--SAA-----
MmpB_Module8_ACP-ACP-ACP_1 1 DPSDILRGLYCOQLKVAQEDDTTAFSDYGFDSVMLTETATLNRITLLELSTAALEFHEPTQALAAHQOGARTAESQPP--APG-----
BaeJ_Module2 1 KWHGALHLLSSLLKYGQDELDLDELSEYGFDSVFTVPTGLNEANQKLAETLFEHGLTRSLAAHLLTDEAEAG--PS--QPEEK---HTA-----
MlnH_Module11 1 TITDELSSLGEATNTPSNKARHQSEYVGLDLSKQLHHSKMLFEELDTQLVPTLLFNNHHDLAALBEKQDVRNTPAFAAFKKEKTQTKEP-----S
AlbI_Module1 1 SPLDDEVRKLLGRLLKMDVDTLDSHPLEYGVDSIVAETAMALRETPG-FEVSLEFETQSIDLLDESQEAP---LLA-----T-
XabB_Module1 1 SPLDDEVRKLLGRLLKMDHATLDSHPLEYGVDSIVAETAMALRETPG-FEVSLEFETQSIDLLDESQEAP---LLA-----T-
DifF_Module1 1 KSVDFLKKVIGCALKVPAPHRLLAHPLEIAYGLDSILVVRITNALKQVFGS-VSSTLFEFYPTIDELADVFLSRRKAEERQFG--NGRAHPENA---S-
DifF_Module6 1 KTAAEFLKAAAKVLLKMPASEENLAPLLELYGLDSILVVRITNALKQVFGS-VSSTLFEFYPTIDELADVFLSRRKAEERQFG--NGRAHPENA---S-
MlnF_Module8 1 ITEGDIQQLIADIKGTAFETISVVKGVYVGLDSTLLQAVKALREKICRELVPTLLFEYPTIQKLANFEEGKETSVEKD----LPLH---YEN
ChiC_Module5 1 AKARLRQLVAGRIGRDEFSVPTIRGEVYVGLSDLLDLVRELEALLCEQLVPTLLFEYPTIQKLANFEEGKETSVEKD----LPLH---YEN
ChiC_Module4 1 RALACIRBELFARELLAEHEIHDNTRFEPCVDSILLGLAKREKIVYCKLFESEIFFEYPTIRAAALQLSSVKKPAPAPSA--AALAARFFASACHA
LkcC_Module2 1 AVEGFREBFACTRRLVYSEVYVGLDSSVDFEFTGVDSTFAVEAGVYVGLSLLDTRTALENYPTARAAAEHTAATFAPSEARPAR--APEFAAQPREQL-PS
DisA_Module2 1 AVEGRILDSLASTLQDRSRLLSSVDFEFTGVDSTFAVEAGVYVGLSLLDTRTALENYPTARAAAEHTAATFAPSEARPAR--APEFAAQPREQL-PS
MmpA_Module7 1 ECAQRLRQSLAAMLYCFGCHRDGSRFTEGLDSTFAVQWIRREINNKYQOKPADGTYTYFVFKAFPTQVVGTCQVPTQATAAP--VOREPVAT-----
OnnB_Module2_ACP-ACP-ACP_1 1 QVRQRLRRLLAEELHMTFDMVEDETEVFKMGLDSIIAVSWVKINQAGCALGATVYVYTNLDDLOHFPEIAKAPSSVIP--EPELAVSS---SDY
OnnB_Module 1 DWHVLRQLLAEELHMTFDMVEDETEVFKMGLDSIIAVSWVKINQAGSEATVYVYTNLDDLOHFPEIATITFTSTT--R-----
VirA_Module5 1 I E V R L L A V L L P P D L D E F G V D S I L G V A V N Y V V T L Y D P T A A T A S G A P -----
OnnI_Module7 1 QNTQLLSQLAQTLEVPENCITLTPFSDYGVDSILGVNFTQINDEGLIENNTVIFPHHTHHHTHTTRVYVSRIL-----CLS
RhiC_Module9 1 AATAVALNGLAKTLQLEFERRMTDPEFAYGLDSILGVSEVGLLEALGIELNTALFLFYPTLDSVLDLGGQPAESARLAG--GQGDVDESANGLSAA
consensus 1

```

ACP

```

MmpB_Module8_ACP-ACP-ACP_2 93 -----PQR--VVT-----FAEPEQRADRQSPVVVLLSAGSAAQDROVADNLLHIDGQADFDVLDHTLARVTQAGSGMLPVRL
MlnB_Module2 92 -----KDR--REN-----HIE-----GAKMSSLLKDRKTEDEPTTF-----SDY
MmpB_Module8_ACP-ACP-ACP_1 85 -----LT--RAQ-----VAQ-----GVRVWAEALKRLEDIGDDDF-----SDY
BaeJ_Module2 89 -----DKS--LQT-----LHT-----ATAMVSGLLKDRREDITLDTL-----SEY
MlnH_Module11 95 EKPVLV---ESLYLQKEW-----TESPLKSEVPEKNTLLDFDSSVSD-----EIR-----KYI
AlbI_Module1 78 -----PAPQDDLQQLKQLARLTKKIDTQDTSKTL-----ESY
XabB_Module1 78 -----PAPQDDLQQLKQLARLTKKIDTQDTSKTL-----ESY
DifF_Module1 93 -----RVQAPAE-----SAPIPNVRERGITVTEHLGKTKLKKIPSGRDPVSVL-----TAY
DifF_Module6 93 -----SLSLEKSTEVYKQVSAALLKMPEDDQVHTL-----SDY
MlnF_Module8 91 HTAL-----TYLTHWRQKA-----AAWDPNC---GNKKRCTLFGYDRDFHMRN-----REN
ChiC_Module5 97 -----GA---AGALD-----TPDTLVTLDTPGTFTAFAPEDALRRKLAGLGGPPAQRSDSGC-----YEL
ChiC_Module4 99 -----AA---AAR-----EGSATLQWVGGQLAGLVKMPAAEDASTRI-----BEL
LkcC_Module2 91 -----A---VA-----SEAGAVVIRGAVEGEVRVAGVHLIDSEVGRDPS-----BEM
DisA_Module2 98 SPQPAPGAPPR---ER--QAT-----SQVQAP-APERPPPPQPGAQQRVRLALGALAEVMAIDREPSDAIL-----AEC
MmpA_Module7 92 -----FA-----DFGCAASAQRESQDYLLKSLLELLEDPGQSRGACF-----DEL
OnnB_Module2_ACP-ACP-ACP_1 92 SDSDIYPGFKP-IALQETVAAKKTGNANGSSLAHEKQKPTPAGRPEAIGQDLKDVHVKRQLLAEELHMTFDMVDSVF-----VEM
OnnB_Module 85 -----QSA-----NAVGEPE-----EKAVAMABDRQLLAEELHMAABATDDVNF-----VEM
VirA_Module5 78 -----APAPTFAVAVPASAVFAV-----AVDPAPVARLRELARLYCEPGDIDEAFC-----NTL
OnnI_Module7 82 NAPQLQLEVP-----DHGELVHRLVSHLAKLLDWAESTIEGDVDF-----SDY
RhiC_Module9 99 HTPPEVSAHTPELVLAELDLPVVT-----AQVMQASQEAALAVVGLASALCLDAAQIDNEQEF-----SDY
consensus 101

```

```

MmpB_Module8_ACP-ACP-ACP_2 164 GLSVSSLEAEQ-----FRAYLOTIE-HSAYAQVGNALVWASVNR
MlnB_Module2 127 GDSITFTTRGGR-----INQOYCFELPSVFEFCNLAEMTDYVIEQ
MmpB_Module8_ACP-ACP-ACP_1 119 GMSVSSVQITGL-----LNEAFDCLQADTFQAGNWWELTAAADI
BaeJ_Module2 124 GDSVSTVFTVQ-----LNEAYQELAFIPIFEHGLHSGLAGYLAIE
MlnH_Module11 141 GSENVYAEAAA--YEMVNEKHYKFNFI-----QSAHYSRLLEDYKKS
AlbI_Module1 115 GDSIVLIEEANA-----FRYYP-SLDAQMELSLIPRLVAQWQAT
XabB_Module1 115 GDSIVLIEEANA-----FRYYP-SLDAQMELSLIPRLVAQWQAT
DifF_Module1 140 GLDSILVVRITNA-----FRWVDP-QMSTLFEFYPTIEEADYVLS
DifF_Module6 129 GDSILVVRITNA-----FRWVDP-QMSTLFEFYPTIEEADYVLS
MlnF_Module8 140 GFENTQITLIVISGNRFSYHGQGVYCFHR-----EAHFRLAADLKRR
ChiC_Module5 152 GLSDLLSARE-----EALLSECYFVFLFPHPTLAVHLRRE
ChiC_Module4 138 GIDPTVTSRHH-----EASEEVALDPAOLFHHFLRRLCGMIDEA
LkcC_Module2 136 GDSISMEIKP-----FRVYV-VPATVFEYPTVRLVEPTTE
DisA_Module2 166 GHDACQAVVYVSS-----NQALTSATAMDLRCCPLADVFHLLAS
MmpA_Module7 135 GDSVGTQWRG-----VSRRESHCLADATVTPPLKSLADSDWR
OnnB_Module2_ACP-ACP-ACP_1 176 GLDSIIAVSWIQ-----LNQAYGLSEATVYVYTNLDDLACHFPE
OnnB_Module 130 GLDYVMAGSWVQV-----LNQAYGLSEATVYVYTNLDDLACHLASE
VirA_Module5 132 GLDSILVVEVAF-----VNOYVGLDCACTLYDHPSLAALSRHVAGR
OnnI_Module7 124 GLDSILGVNFTQ-----ENDDLSEMNNTVIFHTSNALADHHSKT
RhiC_Module9 162 GLDSILGASVDE-----LNEALGHEISALFLFYPTLSTVYLDH
consensus 201 *

```

b

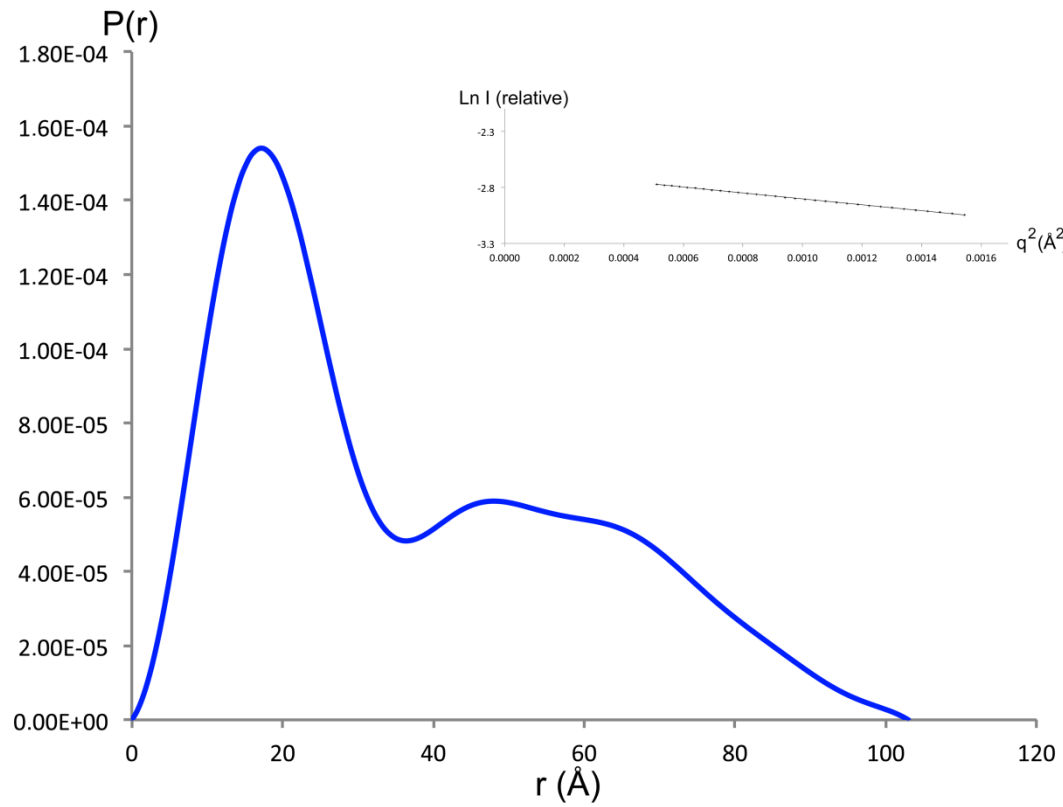
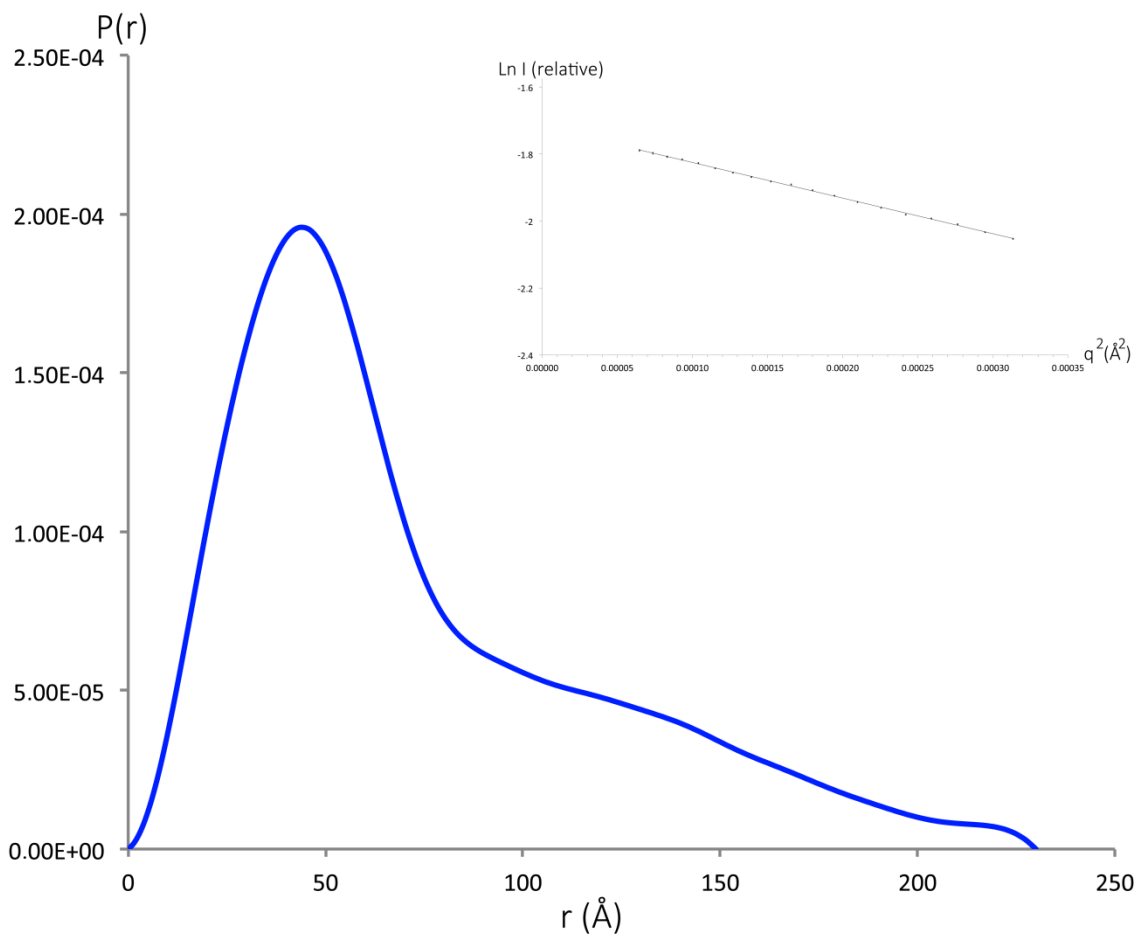
```

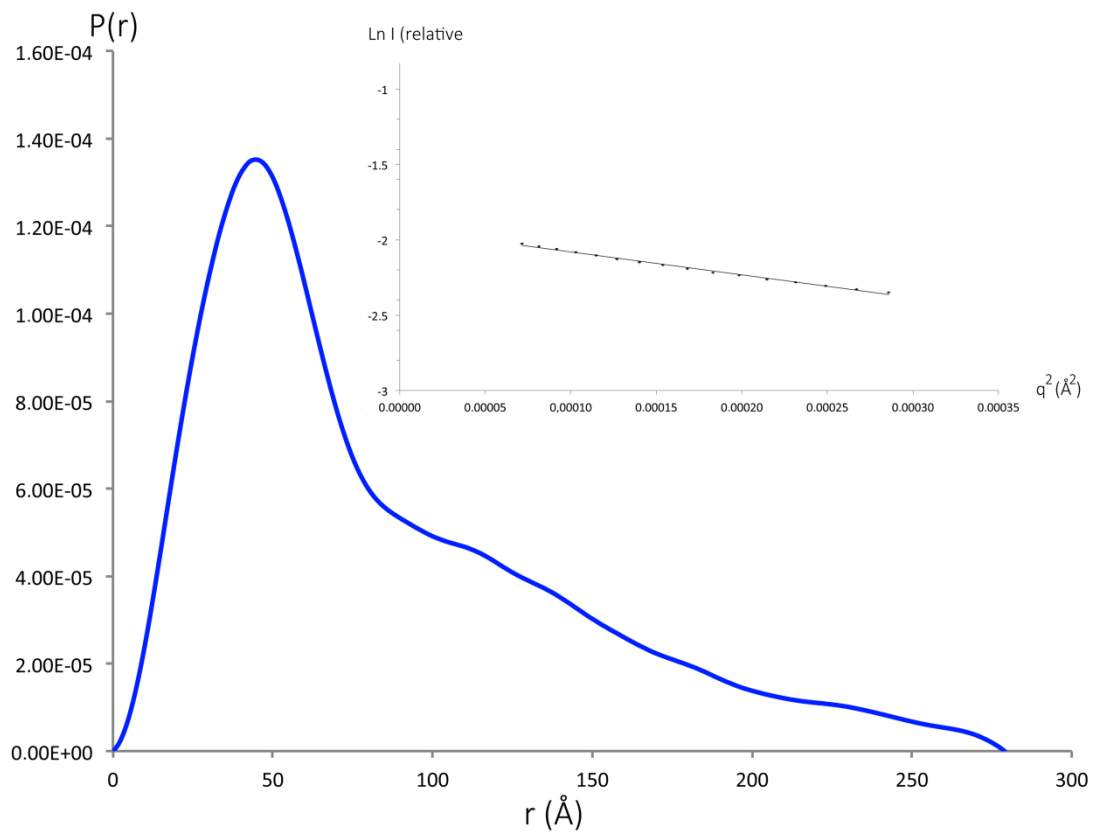
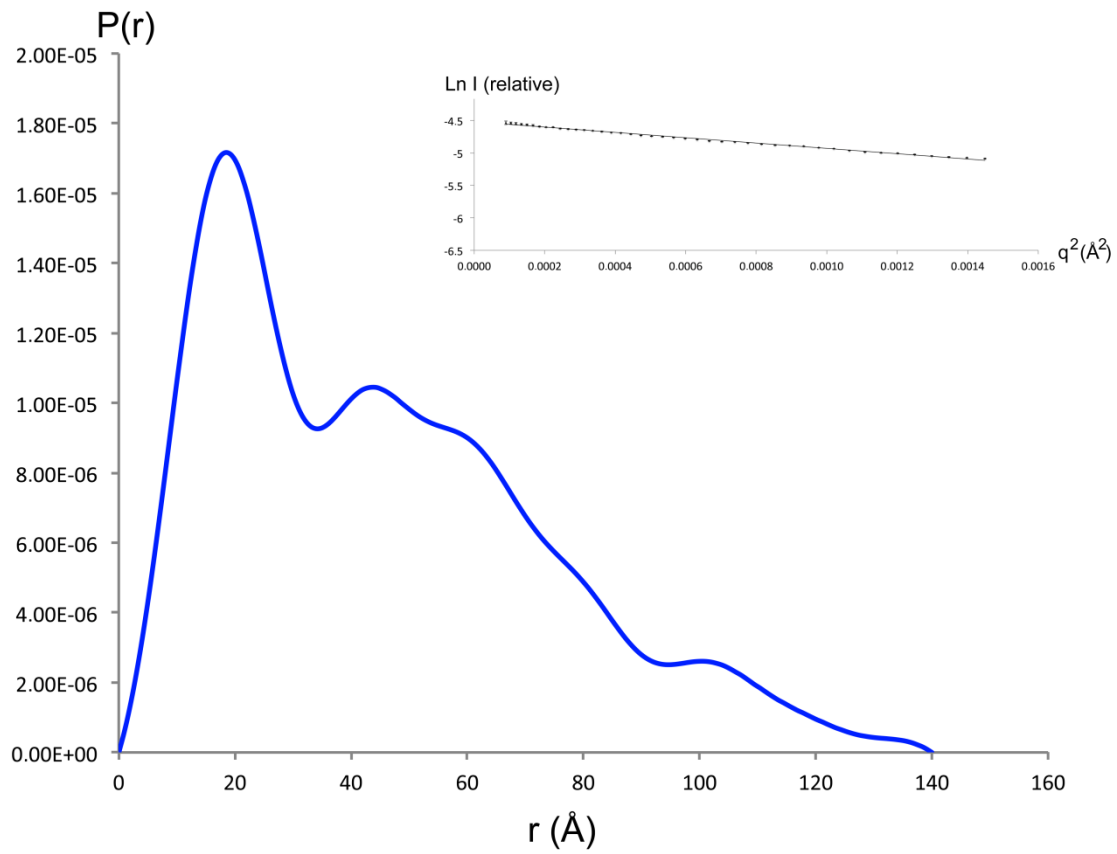
M LEVVAEKTGY PTEMLDLDMD MEADLGIDSI
1261 KRVEILGTVQ DQMPNLPELS PEDLAECRTL REIVTYMNSK MPAAAAASAPV TSASNGLDAA
1321 QVQSTMLEVV ADKTGYPTM LDLEMDMEAD LGIDSIKRV ILGTVQDQLP TLELSPEDL
1381 AECRTLGEIV SYMNSKLPAA SAVAAPVASA PVASANGLD AAQVQTM LD VVADKTGYPT
1441 EMLDLAMDME ADLGIDSIKR VEILGTVQDQ LPGLPELNPE DLAEACRTLGE IVDYMNSKLE
1501 AVSTQNVAIQ TAAPVASASN GLDAAQVQGT MLEVVADKTG YPAEMLD FAM DMEADLGIDS
1561 IKRVEILGTV QDQLPGLPEL NPEDLAECRT LGEIVDYMNS KLPAASTQNV AVQTVAAAPVA
1621 TAPATNGLDA AHVQNTMMNV VADKTGYPAE MLDLAMDMEA DLGIDSIKRV EILGTVQDQL
1681 PGLPELNPED LAECRTLGEI VAY

```

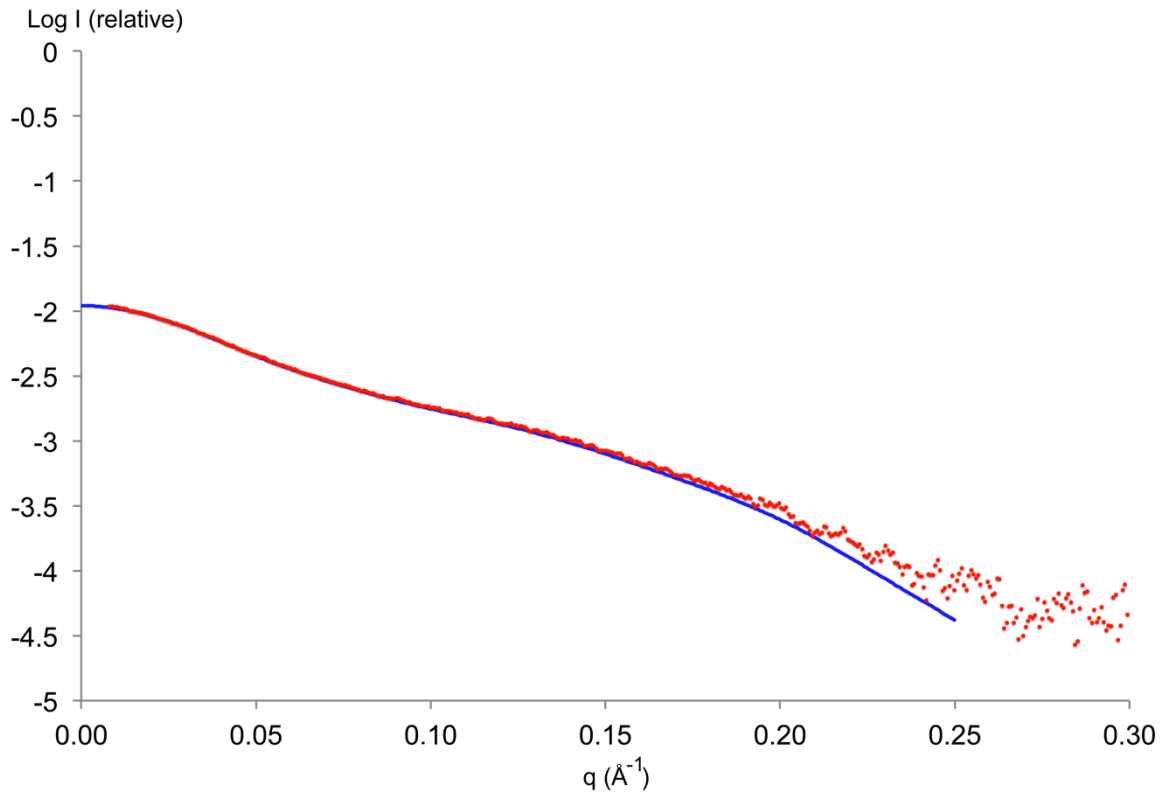
- Linker 1 = 24 residues (P+A = 10; 42%)
- Linker 2 = 29 residues (P+A = 14; 48%)
- Linker 3 = 31 residues (P+A = 10; 32%)
- Linker 4 = 34 residues (P+A = 13; 38%)

Fig. S4 Sequence analysis of ACP-ACP regions in *trans*-AT PKS and PfaA. (a) Sequence alignment of the ACP-ACP regions from a selection of *trans*-AT PKS modules. In each case, the subunit name and module number are indicated. The positions of the conserved ACP domains (based on the presence of secondary structure elements in the NMR structures of Vir ACP_{5a} and ACP_{5b}) are shown for the sequence of VirA. The ACP-ACP linkers of VirA and mupirocin are highlighted in green and red, respectively. Key: Mmp, mupirocin; Mln, macrolactin; Bae, bacillaene from *Bacillus amyloquefaciens*; Alb, albicidin from *Xanthomonas albilineans* Xa23R1; Xab, albicidin from *Xanthomonas albilineans* Xa13; Dif, difficidin; Chi, chivosazol; Dis, disorazol; Lkc, lankacidin; Onn, onnamide; Vir, virginiamycin; Rhi, rhizoxin. (b) Sequence of the tandem ACP region of the PUFA synthase, PfaA (GenBank: AAL01060.1). The positions of the four ACP-ACP linkers are indicated in red, while their respective lengths and content in Pro and Ala are given.

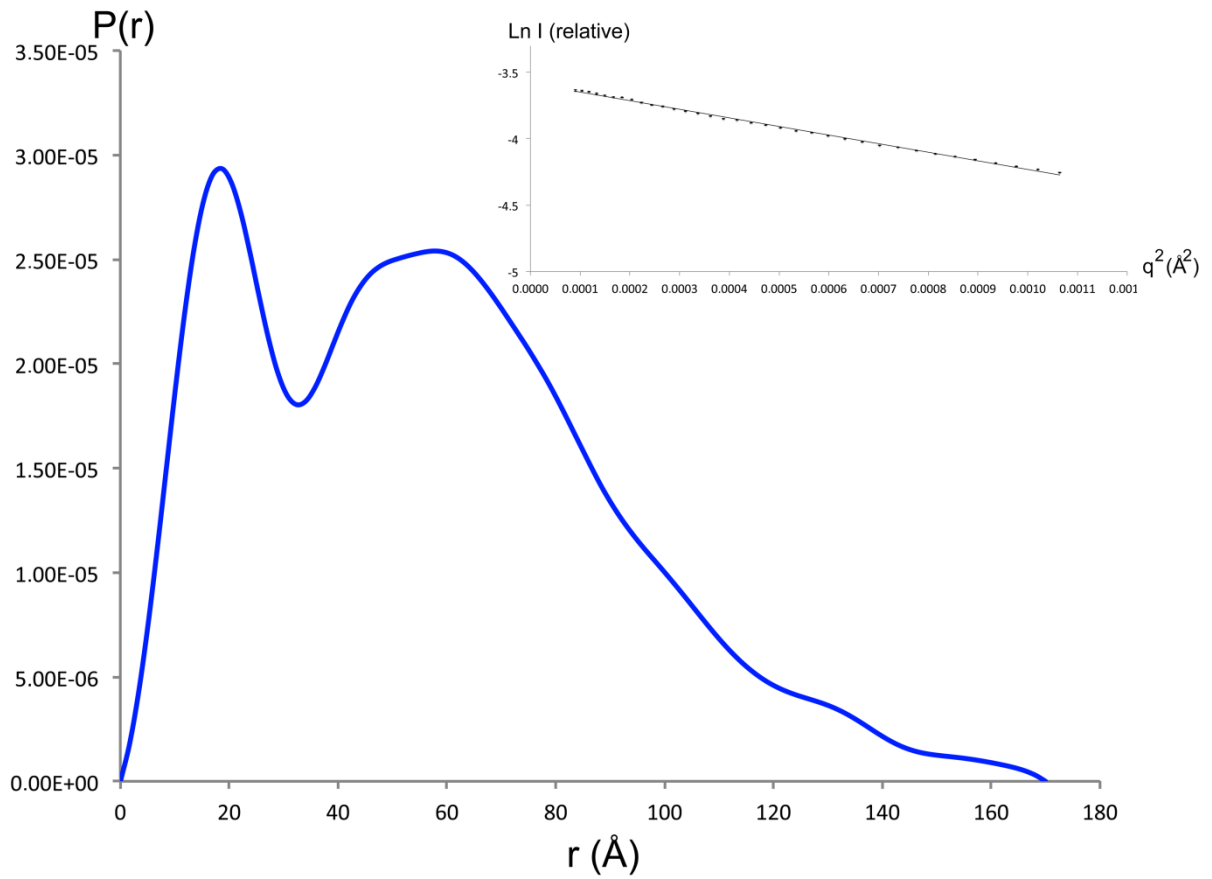
A**B**

C**D**

E



F



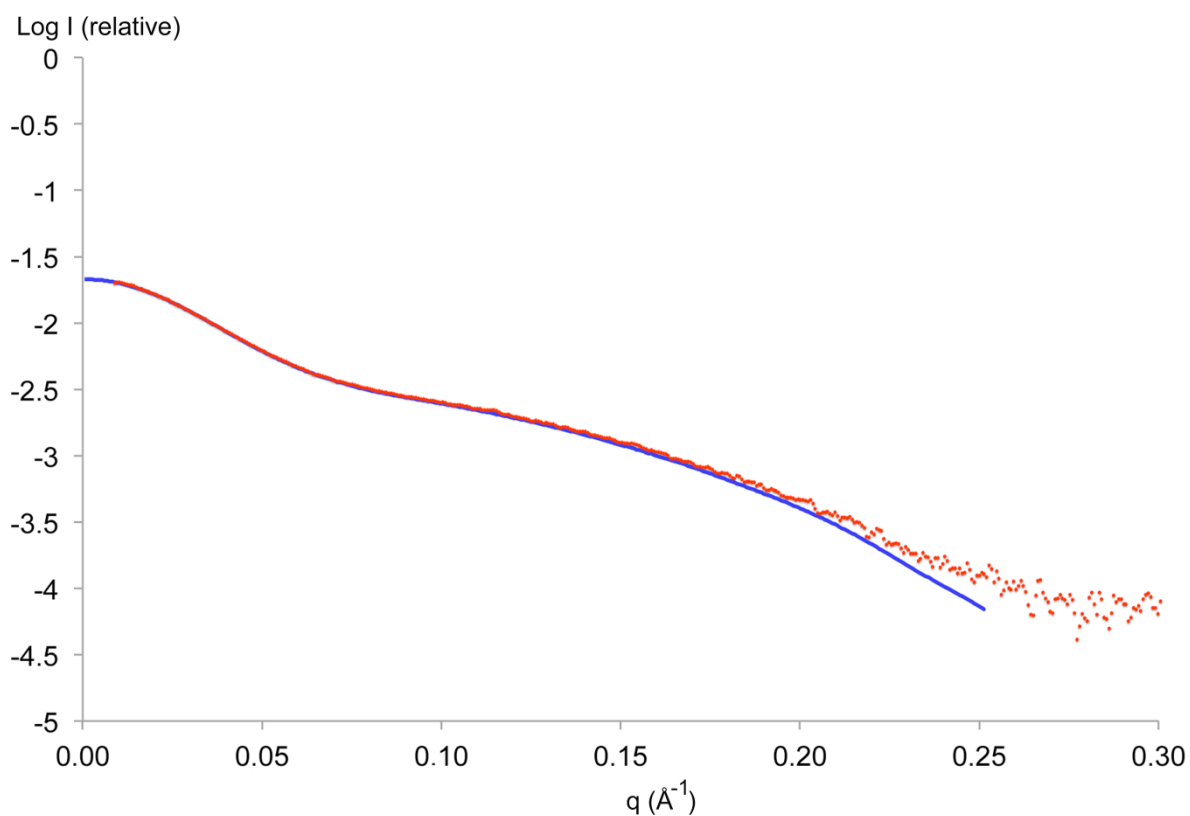
G

Fig. S5 SAXS analysis of the various constructs. (a) The distance distribution function derived for the ACP_{5a}-ACP_{5b} construct calculated with GNOM. Inset is the Guinier plot. (b) The distance distribution function derived for KS₅-ACP_{5a}-ACP_{5b} calculated with GNOM. Inset is the Guinier plot. (c) The distance distribution function derived for KS₅-ACP_{5a}-ACP_{5b}-DD calculated with GNOM. Inset is the Guinier plot. (d) The distance distribution function derived for the ACP_{5b}-DD calculated with GNOM. Inset is the Guinier plot. (e) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data acquired on the ACP_{5b}-DD (red dots). (f) The distance distribution function derived for the ACP_{5a}-ACP_{5b}-DD calculated with GNOM. Inset is the Guinier plot. (g) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data acquired on the ACP_{5a}-ACP_{5b}-DD (red dots).

a

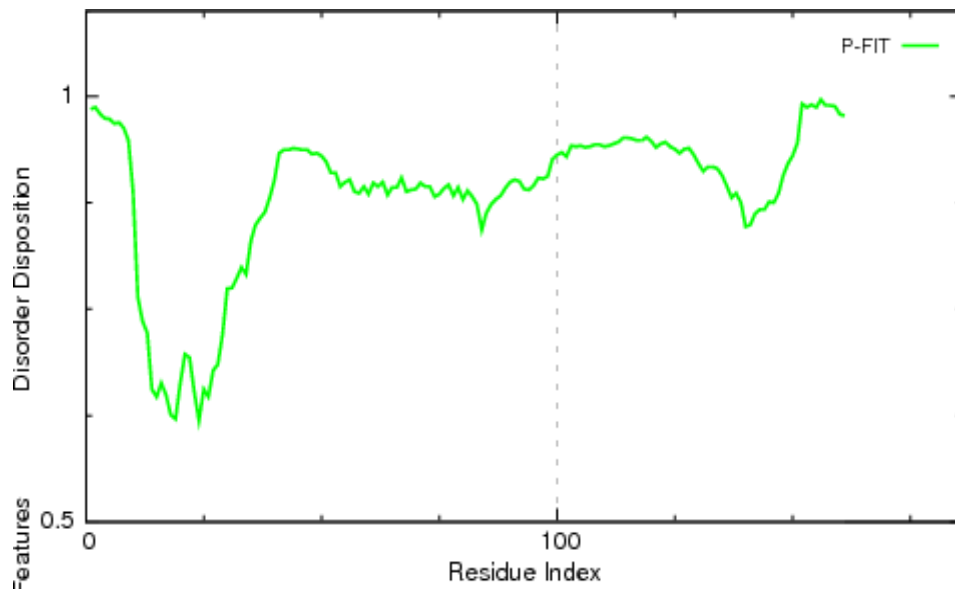
Amino acid bias of intrinsically disordered regions:⁴

Low content: C, V, L, I, M, Y, F and W

High content: Q, S, P, E, K, G and A

LQIAHLHRPGLGSAGPEPAFARDHLRDFGRAPSDRPPAMPGTAPAPAAPVPSAAEGAPVSAP
TGVPARAAVRPAVPAPTAVPTLTLRGAAEPARRPATDPMKPPMTSDRKISLRPLAPATAAT
RPAAPATPTPAAPAAPVAPPAAPAPAAVPAGRQE

b



C

```

BryC_Module11_ACP      1 ENYDRKKMMDHIVNEGCTMVSIDDEL---IEST---Q---SLNTDATSFIPDLNLFIEE-----KNETDIYHIESYD-----DSYDDL
PksM_Module11_ACP      1 -----ISSGTQQSEAV-----KQ-----HSQDMK
BryD_Module13_ACP      1 -----ISEQYGDVACQEDKVNIN---VNSEK-----SISQNEI
PedH_Module14_ACP      1 -----LGGNAGRUVVLKA-----
BryX_Module16_ACP      1 -----IVKDRSTHT-----HCD-----VMTSTKAISSK-QLPLGKNKEILKNNDSFVGGDDV
PksL_Module6_ACP-ACP   1 -----IDVNAKAE---KRT-----EPPFAPVQPVIKPPSVDRASGKPANIT-LQPLMTN-----QD-
BaeL_Module6_ACP-ACP   1 -----IDTSA-DAAVKRAETQSAPIAPAAAPEPVIEVTTLHEETVKE---KPSNIT-LQTLGSAKPKRTVQ-----ERV
OzmQ_Module1_ACP       1 -----VEPEHRT-----ARRTPEPDTL-TDRLGAAVGH-----ERV
OnnI_Module4_ACP       1 -----FSSQTQDVDLK-SQSIETPVTETAPQP-VLQQDNV
DifL_Module14_ACP-ACP  1 -----AADVNSSLPGQAMKSETQTVLPSSPQPEHEGLISLSPLTEIEMTAFEPETLQKPIV-LQLDCTVSVQQPEK-AVSEDNR
PedF_Module4_ACP       1 -----
VirH_Module9_ACP       1 -----
ChiC_Module9_ACP       1 -----
MmpA_Module7_ACP-ACP   1 -----APSEVAVSVE-----
MmpA_Module5_ACP       1 -----
LnmI_Module2_ACP       1 -----LAAPA-----
DszC_Module11_ACP      1 -----L
AlbI_Module2_ACP       1 -----L
ChiF_Module19_ACP      1 -----IGQGEAPFWSQPSQ-----QT-----RRNGEAEHAEPPP-RL-----
LnmJ_Module5_ACP       1 -----IGG---WRS-----AQETGESREPVETA-----
EtnE_Module6_ACP-ACP   1 -----
EtnD_Module1_ACP-ACP-ACP 1 -----
KirAII_Module5_ACP     1 -----ITPGP-RFDQAQA
OnnB_Module3_ACP       1 -----VARRKETGMLTEQNHEQREPTSEKADHEFFWKIEELIEE---WER-----QQYGGHLTPETLTK-VFAYEVV
LnmJ_Module8_ACP-ACP   1 -----
ChiD_Module11_ACP      1 -----V
DszB_Module8_ACP       1 -----
SnaE2_Module5_ACP-ACP  1 -----YDLQIEHL-----HH-----KGLG---TATAEP-ALSRPHL
DszC_Module10_ACP      1 -----
KirAIV_Module9_ACP     1 -----
OzmJ_Module10_ACP      1 -----
VirA_Module5_ACP-ACP   1 -----YDLQIAHL-----HR-----QGLG---SAGQEP-AFARDHL
EtnI_Module20_ACP      1 -----
consensus               1 -----

BryC_Module11_ACP      71 KDNDNMSEM---SDDI-----
PksM_Module11_ACP      20 TEIDEPNGKT-----
BryD_Module13_ACP      31 TQQ-----
PedH_Module14_ACP      13 -----
BryX_Module16_ACP      47 AETTLMVD-----
PksL_Module6_ACP-ACP   50 -RLERVPSDTETETIT---AE-----
BaeL_Module6_ACP-ACP   63 -----PASRPAVSAE-----
OzmQ_Module1_ACP       31 D-----
OnnI_Module4_ACP       34 LEKKPLPETRSTETS-----SCGSPREQQV-----
DifL_Module14_ACP-ACP  80 -----
PedF_Module4_ACP       1 -----
VirH_Module9_ACP       1 -----LPVPAASARP-----EKEQQPD-----
ChiC_Module9_ACP       1 -----ITQAPVSAPAPEAPPVTD-----ANGTA-----ASPSTLRAPQRGTSL
MmpA_Module7_ACP-ACP   11 -----SVASVAQGGKIRL
MmpA_Module5_ACP       1 -----
LnmI_Module2_ACP       6 -----ARR---P-----DGAAGSAP-----AAPESGQSAPP-----ASP--QVQDDRAD--
DszC_Module11_ACP      2 DDEALPHGAWSATAAPPAQTAWSATAAPPARAA-----
AlbI_Module2_ACP       2 RDD-----AAHAAEPAP---ADG-----AAEDASADAPNAANAPTDPV-----ATL
ChiF_Module19_ACP      34 -----
LnmJ_Module5_ACP       21 -----HAPTAVPAAPV---DQADLPEVREERFGLDPQ-----
EtnE_Module6_ACP-ACP   1 -----GFWPAPSAAPPATAPSRPASPPTSIALAGA-----DAAVRSDAASPPAIARSAPKERAERIRADAAPISL
EtnD_Module1_ACP-ACP-ACP 1 -----
KirAII_Module5_ACP     13 EAAGLA-----
OnnB_Module3_ACP       65 KNLGPEQIHR---VDQLL--AATPKPATAPPPVQAAA-----VVPEPKLVRSED-----
LnmJ_Module8_ACP-ACP   1 -----FPTDAEP-----TAPAVPARKEA-----TMTARS-----EAQ--PTRRGPKIRL
ChiD_Module11_ACP      2 REISRASGAA---PAPAAAAPAPALAAAPAVAA--AFAAAPAPA--P-----
DszB_Module8_ACP       1 -----
SnaE2_Module5_ACP-ACP  28 RDFGRPHTPA---P--AAAPAPSPAADPAPTAPRRAPEAAPAPAPVPRRVEPAPALVPRPAPASAPAPGAVTPATDRDPQETP--FMRDRKISL
DszC_Module10_ACP      1 -----LGRPAGDAAAP-AVARGETAEEAPSRGET-----AEEAP--SRG-----
KirAIV_Module9_ACP     1 -----KLFTQPAAR---P--VAQPAA-----EPAPAPRTEPVAEP--LAGAGAG--
OzmJ_Module10_ACP      28 RDFGRAPSDR---P--QAMPGTAPAPAAVPSAAEGAQVS--APTGVPARAAVRP-AVPAPTAVPPLTLRGAEPARRPATDPMEKP--QMTSDRKISL
VirA_Module5_ACP-ACP   1 -----
EtnI_Module20_ACP      1 -----
consensus               201 -----

```

```

BryC_Module11_ACP -----
PksM_Module11_ACP -----
BryD_Module13_ACP -----
PedH_Module14_ACP 13 -----SEHS
BryX_Module16_ACP -----
PksL_Module6_ACP-ACP -----
BaeL_Module6_ACP-ACP -----
OzmQ_Module1_ACP -----
OnnI_Module4_ACP 59 -----LAE-----
DifL_Module14_ACP-ACP -----
PedF_Module4_ACP 1 -----VDSAVVEAERAPN-----SKAPASMLS
VirH_Module9_ACP 19 -----GPT-----GTGPT-----PF-----
ChiC_Module9_ACP 40 D-----
MmpA_Module7_ACP-ACP 24 -----RALDITLPVADSAPIVPPAPRILSQPAPLAAKAAFPV-----PVADD-----
MmpA_Module5_ACP 1 -----YDAPAPMAAYPAPGTSRT-----GRRAWSPD-----
LnmI_Module2_ACP -----
DszC_Module11_ACP 37 -----DPGGA-----APPEG-----PGGAPPGGAARQ-----
AlbI_Module2_ACP 40 VRRITVAQVLGYF-----
ChiF_Module19_ACP 34 -----
LnmJ_Module5_ACP -----
EtnE_Module6_ACP-ACP 65 VALVPPAAMDTAASSAAR-----SAP-----VAISLSPLRAEPASAAY-ARRADEPPASGGE-----
EtnD_Module1_ACP-ACP-ACP 1 VTR-----RQVAPAPEVSPTVAP-PGPAAIKPKSGIALR-APGEERFAAPA--VSSGVEHTVLSLSPVGAAVNQAAVAVGSAAGAAAPYGGGSPAGGAASSAIA
KirAII_Module5_ACP 19 -----GTAQDEVTTWPEP-----AAAVP-----APPAESAPAPAATAAGGDPEP-----
OnnB_Module3_ACP -----
LnmJ_Module8_ACP-ACP 39 AAP-----RSAGGASATAPTREPVPAP-----PAPAPAP-----A-TTAPQPVRAVPPALPEAAAAA-----
ChiD_Module11_ACP 43 -----ALAAPAVAA-----AFAAAPA--PAPVTPAGSQ-----AEDGLLD-----
DszB_Module8_ACP 1 -----LDPAPS-----DEASPSFAA--PPPEAPR--PAAAP-PAPPSAEAR-----
SnaE2_Module5_ACP-ACP 120 -----RPLGTAPTAPAPAPAPAFVSVAAAPV-----PATPT--PAAAP-AAPVAAPSVAAPVAGAVD-----PAVA
DszC_Module10_ACP 37 -----ETAEEAPSRE-TAE--EAPAAAL--PATADP-----ALRK-----
KirAIV_Module9_ACP 2 -----RSAAPRAATPTPTRP-AGSSGP-----
OzmJ_Module10_ACP 38 -----VAEAAAAGAEFY-----APERLA--A-----
VirA_Module5_ACP-ACP 117 -----RPLAPATAATRPAAP-ATPATPAAPAAV-----APPAAPA--PAAVP-AGRQE-----
EtnI_Module20_ACP -----
consensus 201 -----

```

```

BryC_Module11_ACP -----
PksM_Module11_ACP -----
BryD_Module13_ACP -----
PedH_Module14_ACP 17 ERDAVEFEVERNGSA-----
BryX_Module16_ACP -----
PksL_Module6_ACP-ACP -----
BaeL_Module6_ACP-ACP -----
OzmQ_Module1_ACP -----
OnnI_Module4_ACP -----
DifL_Module14_ACP-ACP -----
PedF_Module4_ACP 23 G-----ER-----
VirH_Module9_ACP -----
ChiC_Module9_ACP -----
MmpA_Module7_ACP-ACP -----
MmpA_Module5_ACP -----
LnmI_Module2_ACP -----
DszC_Module11_ACP -----
AlbI_Module2_ACP -----
ChiF_Module19_ACP 34 -----NGTINGAHAD-----
LnmJ_Module5_ACP -----
EtnE_Module6_ACP-ACP -----
EtnD_Module1_ACP-ACP-ACP 94 GVERVTPRATEGGEGGHMPGTTATARDRRASPVDNAAGGRSLA
KirAII_Module5_ACP -----
OnnB_Module3_ACP -----
LnmJ_Module8_ACP-ACP -----
ChiD_Module11_ACP -----
DszB_Module8_ACP -----
SnaE2_Module5_ACP-ACP 179 D-----
DszC_Module10_ACP -----
KirAIV_Module9_ACP -----
OzmJ_Module10_ACP -----
VirA_Module5_ACP-ACP -----
EtnI_Module20_ACP -----
consensus 301 -----

```

| PKS Module | Linker length |
|--------------------------------|---------------|
| VirA_Module 5 (KS-ACP-ACP) | 165 |
| SnaE2_Module 5 (KS-ACP-ACP) | 179 |
| PksM_Module 11 (KS-ACP) | 28 |
| EtnI_Module 20 (KS-ACP) | 10 |
| OnnB_Module 3 (KS-ACP) | 110 |
| DifL_Module 14 (KS-ACP-ACP) | 79 |
| OzmJ_Module 10 (KS-ACP) | 56 |
| PedF_Module 4 (KS-ACP) | 25 |
| OnnI_Module 4 (KS-ACP) | 61 |
| MmpA_Module 7 (KS-ACP-ACP) | 64 |
| MmpA_Module 5 (KS-ACP) | 26 |
| OzmQ_Module 1 (KS-ACP) | 31 |
| KirAll_Module 5 (KS-ACP) | 58 |
| LnMJ_Module 5 (KS-ACP-ACP) | 91 |
| EtnD_Module 1 (KS-ACP-ACP-ACP) | 136 |
| BaeL_Module 6 (KS-ACP-ACP) | 72 |
| PksL_Module 6 (KS-ACP-ACP) | 66 |
| BryX_Module 16 (KS-ACP) | 54 |
| DszB_Module 8 (KS-ACP) | 36 |
| EntE_Module 6 (KS-ACP-ACP) | 116 |
| LnMJ_Module 8 (KS-ACP-ACP) | 49 |
| ChiF_Module 19 (KS-ACP) | 42 |
| KirAIV_Module 9 (KS-ACP) | 22 |
| AlbI_Module 2 (KS-ACP) | 51 |
| DszC_Module 10 (KS-ACP) | 65 |
| ChiC_Module 9 (KS-ACP) | 40 |
| BryC_Module 11 (KS-ACP) | 84 |
| BryD_Module 13 (KS-ACP) | 33 |
| PedH_Module 14 (KS-ACP) | 31 |
| Lnml_Module 2 (KS-ACP) | 39 |
| ChiD_Module 11 (KS-ACP) | 57 |
| DszC_Module 11 (KS-ACP) | 98 |
| VirH_Module 9 (KS-ACP) | 28 |

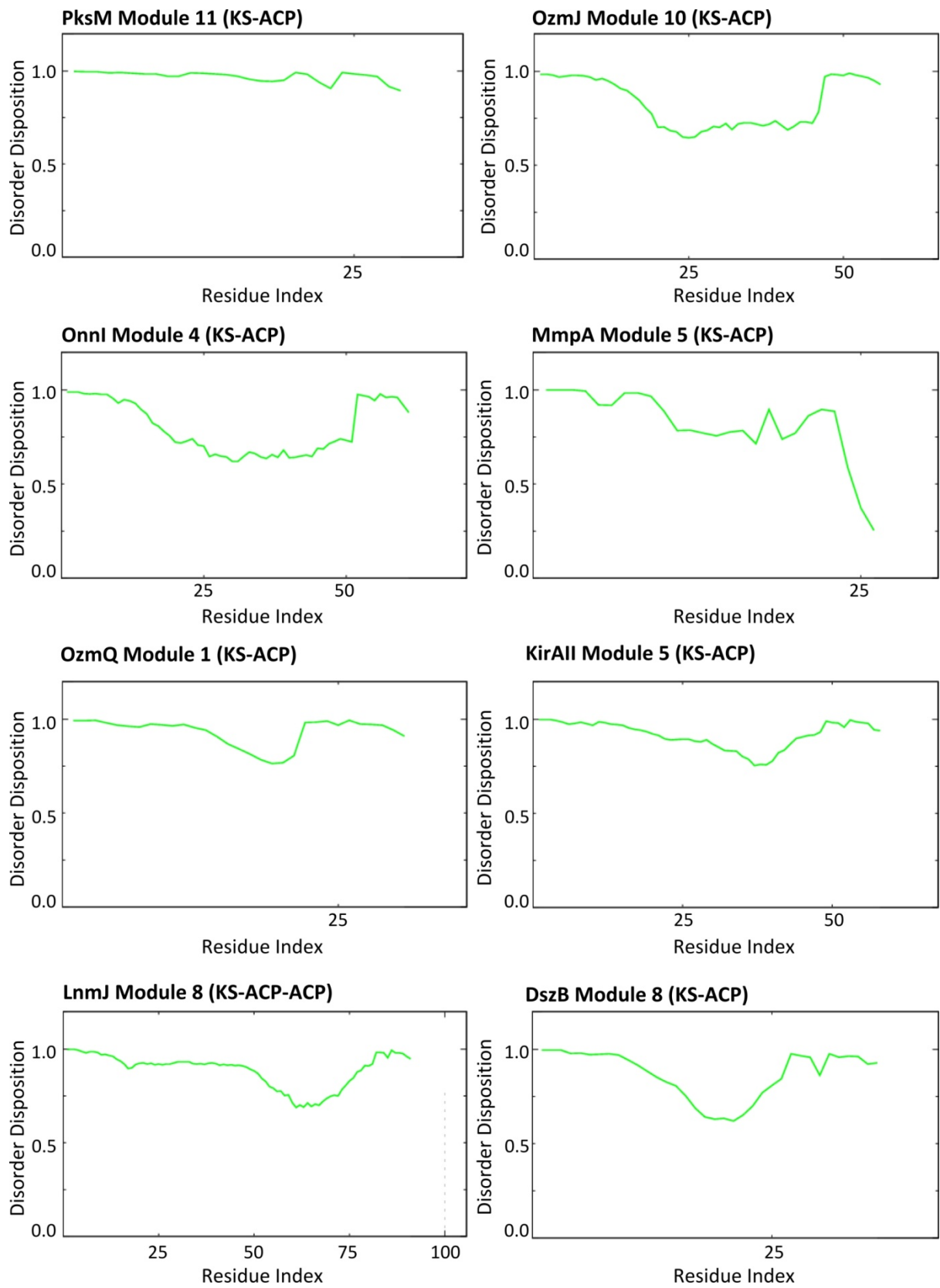
d

```

VirA_Module5_ACP-ACP      1 YDLQIEHLHSCGLGSAQQBPAPARQHLRDFGRAPDRQAMPETAPAPAAVVSBAEGAQVS--AFTGVFARRAVRPA-VVEAFVAVPTLTLRGAAPFARR
SnaE2_Module5_ACP-ACP    1 YDLQIEHLHCKGLGATAPALSRQHLRDFGRPHPAABAPPSPADPAPTPAPRRAPPAAPAPAPVRRVVEPAPALVEREAPASAPAAPCAVTEPAT
consensus                 1 ***** ** . *** * ** * ***** . * * * . ** * * * . . ** ** * ** ** *
VirA_Module5_ACP-ACP      98 PATDEMEKQDMISDRKISLRPLDPTAATRPAPAPATPAAPAAVPAEPAAPPAAVPAR-----Q-----
SnaE2_Module5_ACP-ACP    101 DRPDQETDPMIRDRKISLRPLDPTTPAPAPAPAPAVSVAAPAVPAT---TPAAAPAPVVAAPSVAPVAGAVPAVAD
consensus                 101 ** * * * ***** * ***** * . * * * . . . . . * * * * *

```


e



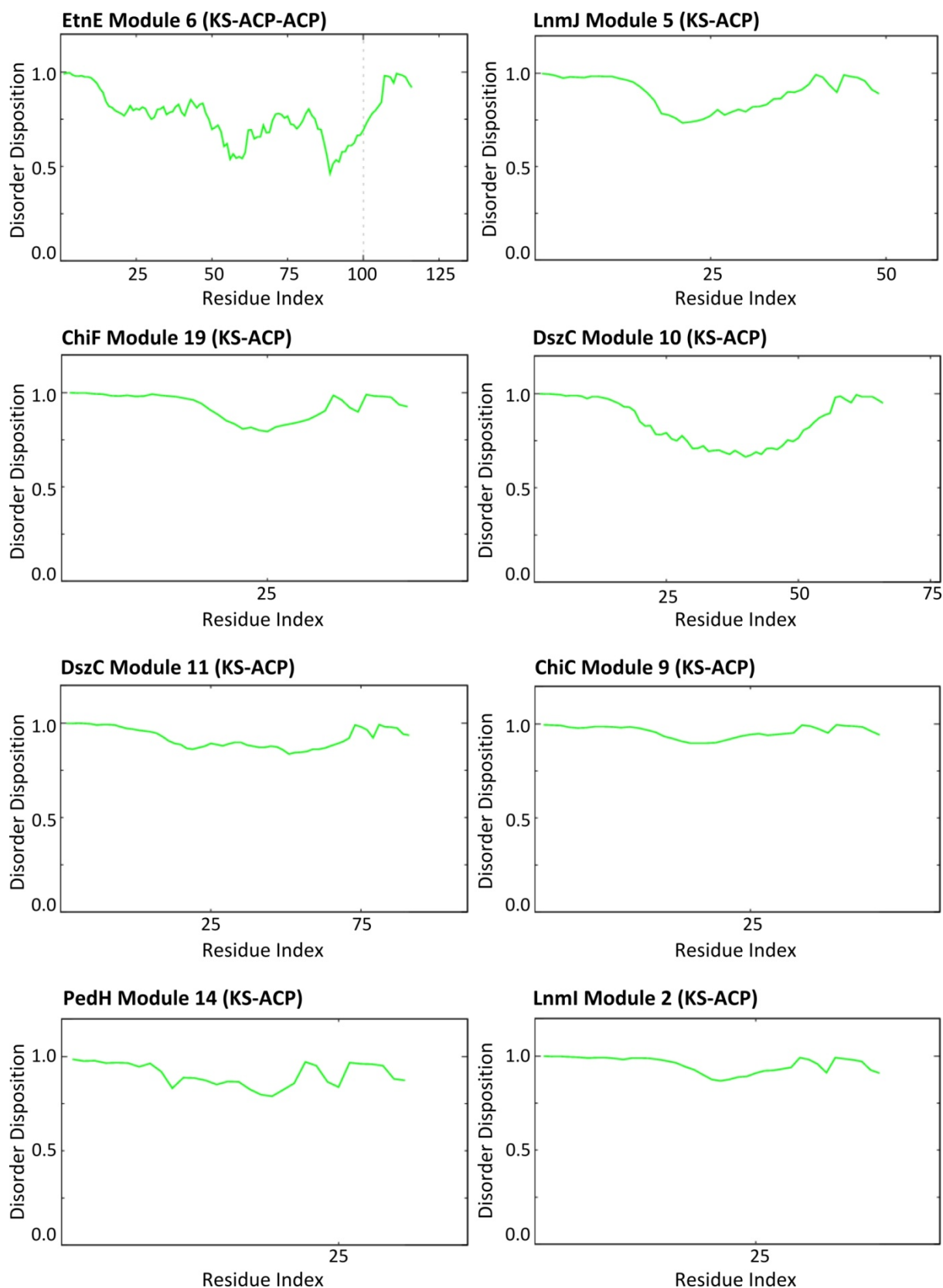
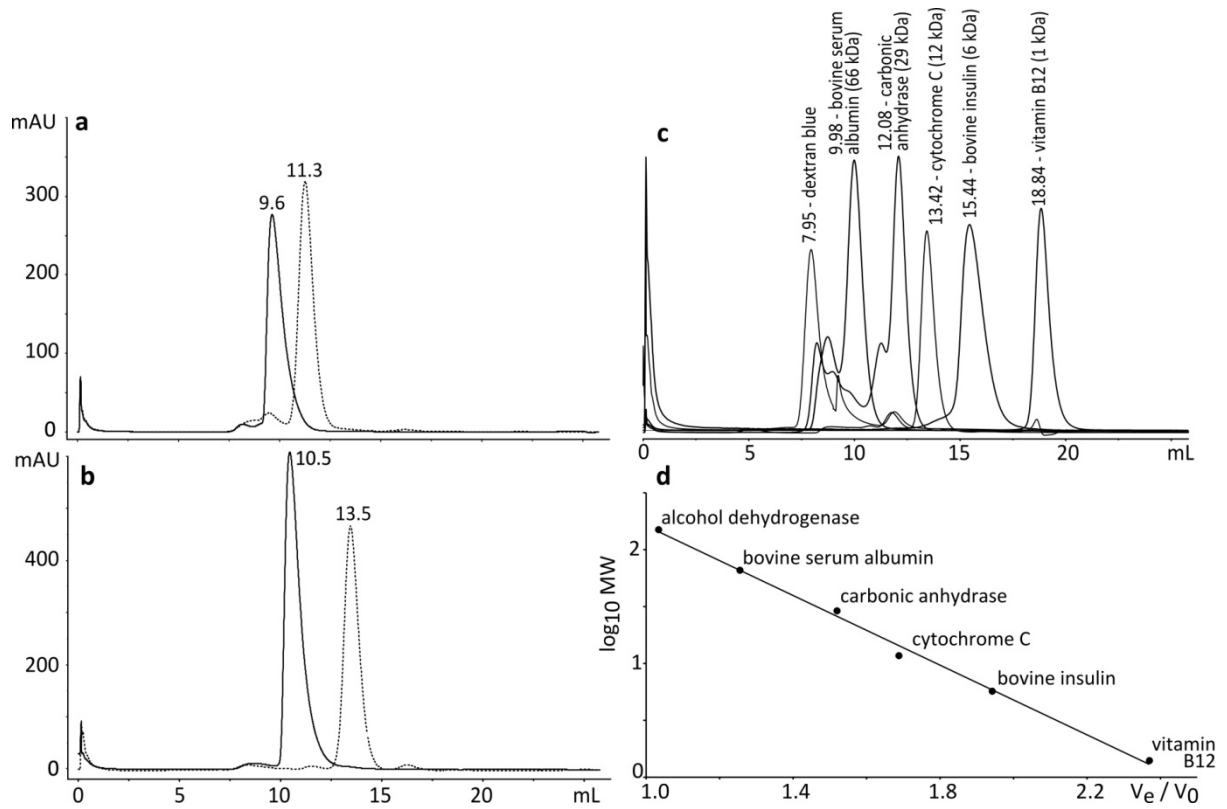
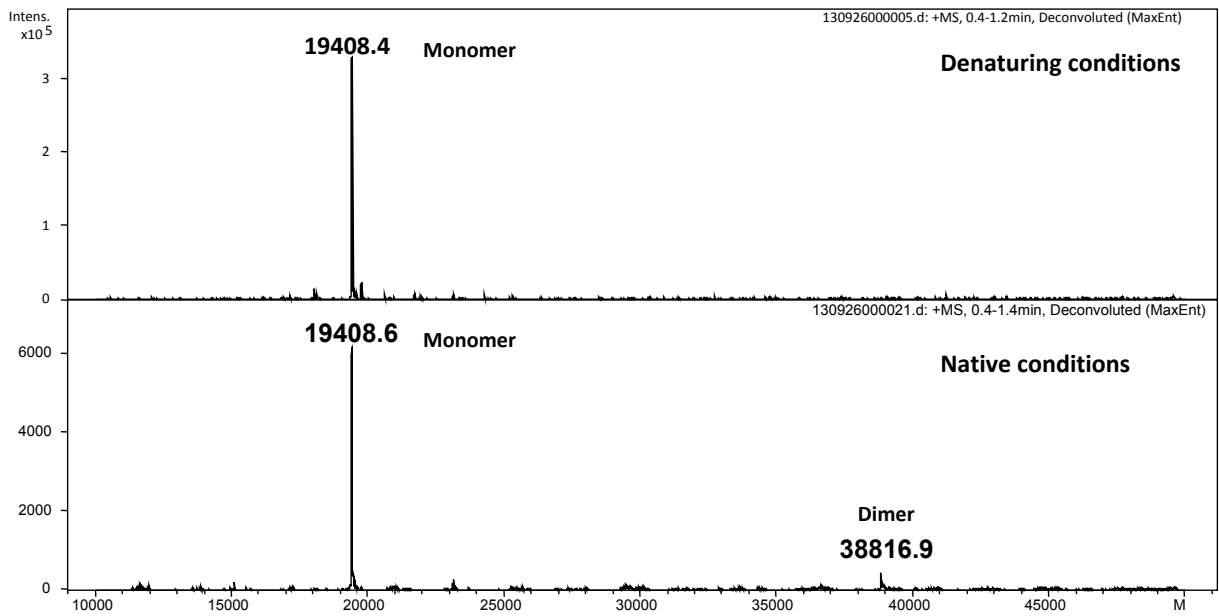


Fig. S6 Sequence analysis of the linker region following the ‘post-AT Linker’ (referred to as the ‘post-post-AT-linker’ (Fig. 3)) in various *trans*-AT PKS. (a) Analysis of the amino acid composition of the 159 residue post-post-AT linker of VirA module 5 reveals a sequence bias consistent with an intrinsically disorder region.⁴ (b) Disorder propensity analysis by PONDR-FIT⁵ of the post-post-AT linker from VirA module 5 (<http://www.disprot.org/metapredictor.php>). (c) Alignment of the post-post-AT linker

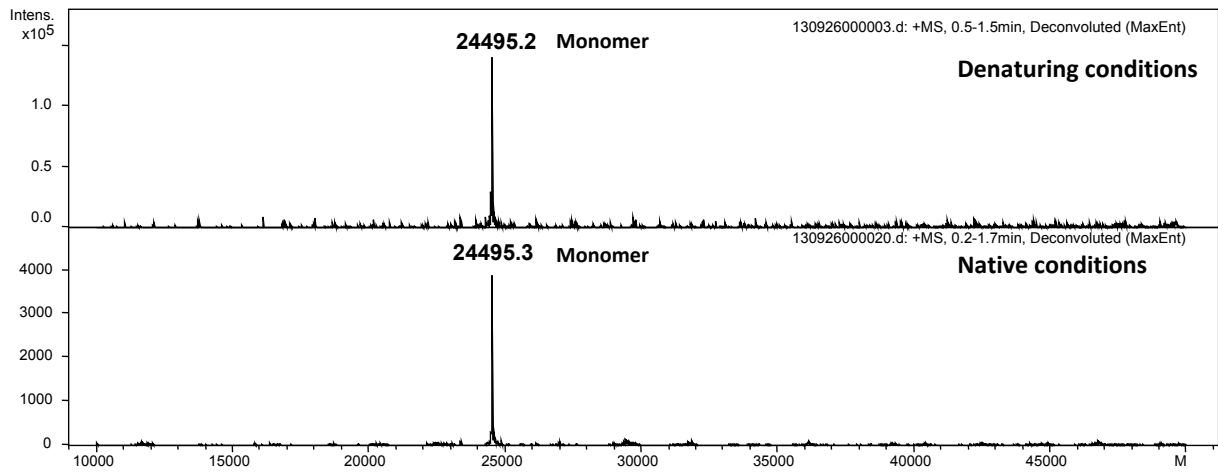
from a selection of KS-(ACP)_n modules from *trans*-AT PKS. The subunit name, module number and its ACP composition are indicated. The linker lengths are given in the table, with the two extremes highlighted in pink. (d) Comparison of the linker regions following the post-AT linker in virginiamycin and pristinamycin modules 5. (e) PONDR-FIT analysis of the linker regions in (c) which show the highest disorder propensity. Key: Vir, virginiamycin; Sna, pristinamycin; Pks, bacillaene from *Bacillus subtilis*; Etn, etnangien; Onn, onnamide; Dif, difficidin; Ozm, oxaxolomycin; Ped, pederin; Mmp, mupirocin; Kir, kirromycin; Lnm, leinamycin; Bae, bacillaene from *B. amyloliquefaciens*; Bry, bryostatin; Dsz, disorazol; chi, chivosazol; Alb, albicidin.



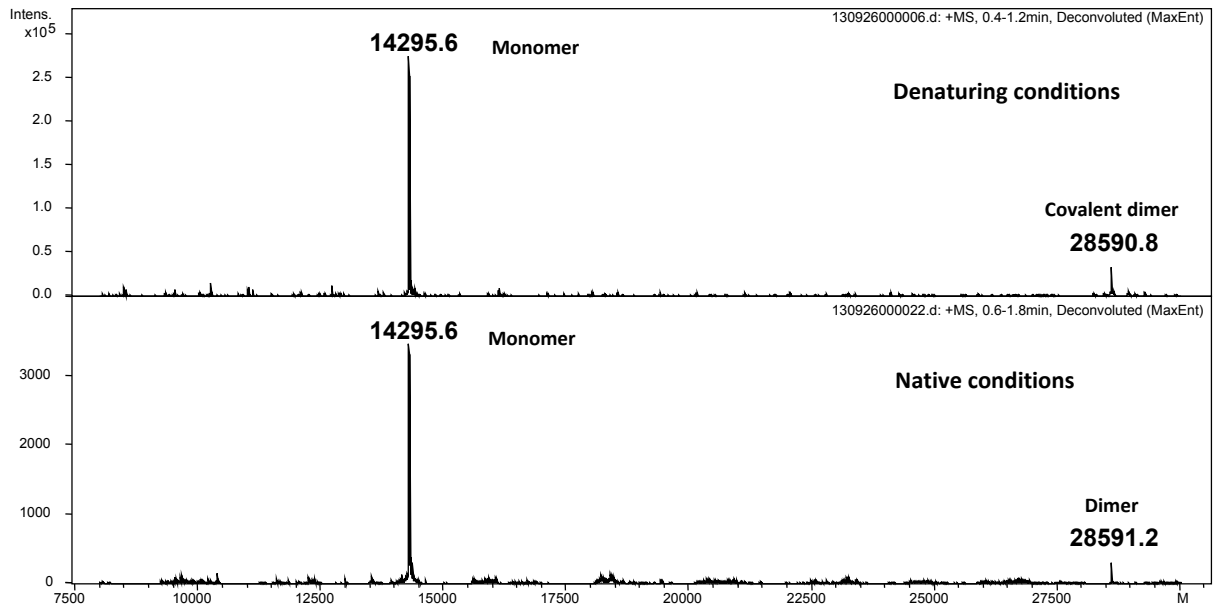
e



f



g



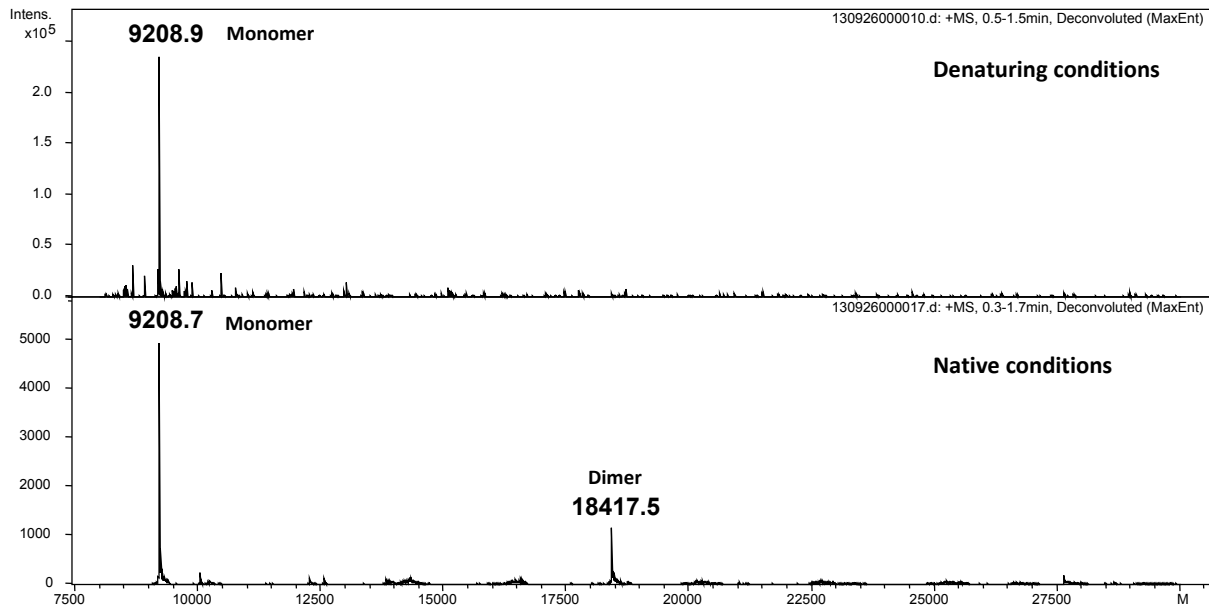
h

Fig. S7 Analysis of the effect on the oligomerization state of various constructs of the presence of the putative dimerization domain (DD). FPLC traces of (a) ACP_{5a}-ACP_{5b}-DD (solid line) and ACP_{5a}-ACP_{5b} (dotted line), (b) ACP_{5b}-DD (solid line) and ACP_{5b} (dotted line), and (c) the calibration standards (Sigma) dextran blue, bovine serum albumin, carbonic anhydrase, cytochrome C, bovine insulin, vitamin B12 (alcohol dehydrogenase omitted for clarity). All samples were analyzed on a Superdex 75 10/30 column (GE) equilibrated with 50 mM sodium phosphate (pH 7.5), 250 mM NaCl. (d) The calibration curve plotting the ratio of elution volume (V_e) to column void volume (V_0) against log molecular weight. The curve was described by the formula $y = -1.53x + 3.74$, with $R^2 = 0.996$. The calculated molecular weights are shown in the **Table** below. (e) Analysis by mass spectrometry of ACP_{5a}-ACP_{5b} under native and denaturing conditions shows that the construct is almost exclusively monomeric, with a small proportion of dimer (estimated at 5%) present. (f) Analysis by mass spectrometry of ACP_{5a}-ACP_{5b}-DD shows that the construct is exclusively monomeric under all conditions. (g) Analysis by mass spectrometry of ACP_{5b} shows an estimated 18% of dimer under native conditions. (h) Analysis by mass spectrometry of ACP_{5b}-DD reveals a small proportion (estimated at 9–11%) of dimer under both native and denaturing conditions, suggesting the presence of a covalent link between the monomers (disulfide bridge?). Panels e–h show the deconvoluted neutral mass spectra obtained using maximum entropy analysis.

| Construct | Expected MW / kDa | Estimated MW / kDa | Ratio of MW in presence/absence of the DD |
|--|-------------------|--------------------|---|
| ACP _{5b} | 9.2 | 14.1 | |
| ACP _{5b} -DD | 14.3 | 53.5 | 3.8 |
| ACP _{5a} -ACP _{5b} | 19.4 | 37.4 | |
| ACP _{5a} -ACP _{5b} -DD | 24.5 | 77.0 | 2.1 |

The calculated molecular weights of monomeric constructs with and without the putative dimerization domain, compared to the estimated molecular weights derived from calibrated gel filtration (**Fig. S7**).

Commentary on the table

The presence of the dimerization domain (DD) on ACP_{5b} gives a molecular weight approximately four-fold greater than the ACP_{5b} alone, while that of the ACP_{5a}-ACP_{5b} is more than doubled by addition of the DD. It could be argued that the addition of the DD makes the proteins substantially more elongated, thus explaining the apparent increases in molecular weight; indeed, an elongated protein can elute at twice the volume of a

globular protein of the same molecular weight.⁶ However, the fact that the ACPs even in the absence of the DD give estimated molecular weights in great excess of their calculated weights, shows that the constructs are already in an extended conformation. Thus, our analysis instead supports a change in oligomerization state in the presence of the putative dimerization element.

MDAKEILTRFKDGGDLDRAAAQALLAGRTPAA**APRPSEPAAPTRPAVPAVPAVEPAAGTTVA**AEGTAGRPEPVA
VIGYSARFPGAADADTFWQRILDGDDLVEVPPERWRTEEFYDPPAAEGRSVSRWGAYVADADRFDADFFRMT
REAELTDPQARLFLQEAWRALAHAGRDARSLAGTRCGVYAGVMLNDYQDLVERESPYKRLPQVMQGNNSILAA
RIAYHLDLKGPAVTVDTACSSSLTALHLACQSLWLGETDLTVVGGVTLYLTELPHVFMSAAGMLSPNGRCRPFDA
ADGIVPGEGCAVVVVKPLSKALADGDPVHAVIRASGLNQDGRNTGITAPSARSQTALVRDTLQRFVAVDPAGIDYVE
CHGTGTPLGDPPIEVTALNEAFAGAGLAPASVPIGSVKNIGHTSAAAGLAGLLKAAGVVRTGLVPPSLHYARANPQI
PFDQGPFTVAGERRELGRPDDGRRRATVSSFGSGTNAVYVVVEQAPEQAARPAGDDGAPPVLPVLSGRRADAPAA
HAADLARWLRGPGSDASPADVHTLAVARTHHTYRFALLVSGRDELLESLDLLAAGSADPRRTDTAPDAAAPDAA
VRRQAALLARLVERAGENPGPVELAALAKLYTQGHVTDWAAVSPPARHRRLSLPAYPFAPHRHYVERPAPVTPP
TPAATSALAEASHGTPAEPAGARTPQEALADVPLTEPVLHRQEWIPAPLPAAPAPTAPVLVHDPAGDLAAALAAT
GLDVHRLGGDRTAEEALRATGAQVVTVLRLTPPADGAPATAVDLAAFEAARAALAVLRTQQLTLLAVTADPTR
ADAAGALVQTLRQENPRLSGRAVLTADGTPDARRLLAEIAAPAAEHGHLADLRSARRLRRVLPVPLPGTRPFVRQ
DGVYLVSGGAGGIGLALARHLTDRPGTRVVLVCGRTAWDDLAEATRSVSGSDRLRYARADVTDPEAAAAALVAEV
VRTEGALHGVFHAAGVVRDGYLVRKDPADVADVLAPKILGARALDEATAALPLDAFVLFSSVAAVTGNLQSDY
AFANGFLDGFATRRAGQVARGERHGRTLSVQWPLWDVPGMSIPEPVLEVVAQHTGMAPLPAAVGLAALERLLAA
DGPEVVSFLFHGDAATWRAHLAALHLERPAAGQVTAAPAAPAPSDTPVVASADPAAADDGRRAELADRVSRTV
ADTIGRPAGSIGGHTSLESMGLDSVMIRALASRLSAEVAPVGPPEMLFGLRDLDELVDHLVAARPVPVAEPATPAVPP
APAATAATVFAPVAPADPAVVPAAAPALPVPAAPASASRTLPAPSADDRFAIIGISGRYPQAPDLDAFWQNLNGK
DTASDLPTDRWPDAAGVNRGHFLEGVDADFPTFFGLSAHDGTLDDPQERLFLEVAWEALEDAGYTGSRHLDLVA
PDGERRSVGVFAGITSSDYKLLGAERWAAGHREMPSGHYWVSLPNRLSYLLDLRGPSPVDTACSSSLVALHLALDA
LRRGECAAALVGGVNLVYVHPSRFRMLRRSGFLAEDGLCRSFGAGGAGFGPGEGGAVVVKPLAVALADGDTVHA
VVRGSAVAHGGRNGFTAPSPWAQARVLRALRNSGTDPETVNVIEAHGTGTELGDVVELAGLQDAYGSGRVPCS
LGSVKSAVGHGESAAGIAALTKVVLQLRHGELVPTLHADPVNPLRLLEDTRFVLQHTPGRWERLTGADGSPLPRA
GISSFGAGGVNAHVIVEEYLPEPHGRAAAAEPGRPELVLLSAPTREHLAATADRLARRLEGPEAPADLRAVAYSRT
GRSAMDCLAVVATDTAGLAAALGAFARSADAGEGESAVRYADLRDGRRAHRDLDTVPETTAFLADLWRNRRHL
QLGELWLSGLDIERAAPREPGRVPLPPTSFLRRRLWITDPVTADTTPTDPLPTDPLPVPLSAPAPAPAPAPQAP
TTTPAPAAAPAAAPARSGEAVTEQLSRLVAGFVDAAGSVDPDRITLLEHGIDSINLMNLRFEITERFGRTLPLQLLSE
STVPVLAHLSADRAHDRA*

Fig. S8 Corrected sequence of VirFG. The predicted natively unstructured region is shown in bold.

a

```
Ta-1_ACP-KS      1 -----TIVEIWKNVLQV---NEVGVDFRFFVGGDSVLAAVLVEEMNRRFDTRLAVTDLFKVYVNTIR
BonA_ACP-KS_2   1 -----AEEMQIPR-EAISAHTNLAEYGFDSIALTEFARRLARHFSHELEPTLFYSHPSLIG
Bat2_ACP-KS     1 -----QLKEQARELLKLPY-DRIGQEVNLVDFGDSIALVTFAKRLSQCFDFVLPSPVFFSHSTLIG
VirA_ACP5b      1 -----VVARALREELARLYCEP-CDID-EASFNILGLDSILGVE-VAFVNQTYGLDEKAGLLYD-PSLA
OocJ_ACP-KS     1 -----LAKK-VDIDETKAFSEYGLDSFAAINLAVLNAAFSIKLEPCTVLFDDYGCVS
Ta-1_ACP-KS_M   1 -----IREELATSIAQALYIDR-AQVNAESTFVEIGLDSIVGVEWTHAINQOYGLSLPVTVVYDHPNLS
EtnE_ACP-KS_M   1 -----APRPVEELVAGLTQTLAAALYAA-SDIDEDRLFVHGVDVIVSVEWVRINATFHIDLRATLFLFDHPTIH
BaeL_ACP-KS_M   1 -----HDELSKALADVLYMER-HEVDIDEAFILGLMDSITGLFWTKAVNKRYCTDCNVTKVYDYPTIR
OocL_ACP-KS     1 -----DEIQNDRAFAALGLDSILGAEWTHLNLNLTIGTELATRLYDYPTIQ
BonA_ACP-KS_M   1 -----IDVDRFVALGLDSILGVEWMAINRRRHGLALNATLVYFHPVTR
BonA_ACP-KS     1 -----RAVIRRESIGQALKIGE-SQLDDEAFSNYGLDSITGVAVETINARIGLIDLEPTVLFDDYVSE
PedI_ACP-KS_M   1 -----VEVGLDSTAVVWIKALINQRYGLSITGATRVVYMSCLA
OnnB_ACP-KS_M   1 -----DVFVEMGLDYVMAGSWVQKLNQAYGLSLBATVIYVYTNLL
PsyA_ACP-KS_M   1 -----YKILRETLARELHMV-EDIDDRPFLMGLDSVIGVTVWRKLNREHGLSITVTKVYAHETVC
EtnD_ACP-KS_M   1 PRDARVGAARSSELDALQVEIAKSLADALYMEP-CDVDVDRPFVILGLDSIVGVEWTHAINQAYGMSLEPATRLYDHSIVR
Bat2_ACP-KS_M   1 -----ARSPAMSHEELSAELVQGLAKVLYLEH-DDIDRDTAFSELGLDSITGVWVRLNRRYQTSLEGASRLYDYSTIK
Bat3_ACP-KS_M   1 -----ADLRMEIAESLAKVLYLEN-DDVDMDTAFSELGLDSVIGVWIRBELNQRYSLSLQASQLYDYPTIK
OocN_ACP-KS     1 -----TGRILGNGADDETCQQAFTEILGLDSVVAAFVSRVLAAPVWKRRSDFVFSHTTPA
PedF_ACP-KS_2   1 -----ARRQDVAARVRELLDLSLAQALSIGR-EQIEQDIFPSDYGLDSILGVGFVQLNDELGLSLNTLLFFDYTTVQ
PsyD_ACP-KS_2   1 -----VPQPASEDQIASHVVDALTRVLRMEN-GRIPHVAFSDYGLDSILGAFIDRVNNAALGLALNTLLFEYTTTA
OnnI_ACP-KS_2   1 -----LEVDPHGELVHRLVSHLAKVLDIAE-STIEGDVPSDYGLDSILGVNFIQTQINDDLGLENTTIVFCHTSVN
BonD_ACP-KS_M   1 -----AIDATPEFTEILGLDSITGVWVRKINRHGLSLIAMQMYEHPETLR
EtnE_ACP-KS     1 -----REEALRQVKKILAPVTKIPV-ERIDAAASFETYGLISVMVAVELTDLLEA-VFGPLSKTLLFEAKTVR
EtnF_ACP-KS     1 -----AREGASLLHATLARLELLAVVTTIEA-ERIDGAEFPPEAYGLDSIMITQINQLEA-IFGAIKTLFFEHRTIE
EtnD_ACP-KS     1 -----HADVTKIPV-SQIDPEESLELYGLDSILIRLNALS-VEFGDISKTLFFEVQTLA
PedI_ACP-KS     1 -----LKKIVARILRMS-GELAVREPLETYGLDSILIVQNTNALRE-VFPEIPSTLLFEQCTID
PsyD_ACP-KS     1 -----DQLRRLEADVMRISV-DDVIVQAPLESEGLDSIVVVTYLNQRLRD-AFGEVSAITLFEYRTIA
OnnB_ACP-KS     1 -----YFQTLIGKVLRIE-ORIDPAESLEAYGLDSILIVQIIEALQE-IFEDVSSITLFFEVQTV
PedF_ACP-KS     1 -----QLKRIILAQVIGRAV-ERISCEPMDRYGLDSIAITQINRLEE-QFGGLSKTLFEYQVTE
OnnI_ACP-KS     1 -----AEIDDQELATRTLQEMKTLGSGVIGV-DEIDAQKPLENYGLDSIAITQINRLEDG-VFADLSKTLFEYQTV
VirA_KS         1 -----
BaeM_ACP-KS     1 -----LKKILADKLGQPW-EHIVLAGVYILGLDSSSLEHIVQDLSKKTICADAPITLLFFEYTNLK
Bat3_ACP-KS     1 -----EQGLIADKLGAP-EQIDRDSGLYELGWSAGLLELVDDIEQKIDASLPPITLLFFEYVTTG
OocL_ACP-KS_M   1 -----EVLAIISVILKIPV-EKIDSEKISYGVDSIITEHMGRIAGALGMSLAAIFFEARNIE
PsyD_ACP-KS_3   1 -----QTTAADGLLKEVGAHVELVAVVQKIQP-EKIRLNRELAHYGFDSISFTTLANALNAYDLSLMPITLFFETENEA
Bat3_ACP-KS_2   1 -----ALMQALSEBOLKVKV-ENVDVAEISEEFGDSISLTTIGNHLNRYGLELSPITLFFEYSTLR
Ta-1_ACP-KS_2   1 -----AVLEGVSEELKVPV-EEEDADTKLSYGFDSITFTTEFANLNRQLSLGLSPVIFFEETIAN
TaO_ACP-KS     1 -----QVLMVSVSKVLPF-EEEDADAELKMGHDSISLTLALQOLAGEVRLDLAPTIFFEHETIN
BaeJ_ACP-KS     1 -----DADKSLQTHHTALTAMVSGHLKVR-EDIEETDTELSHYGDSVSTVYTNQLEAYQLELAPTIFFEHGTIS
BaeN_ACP-KS     1 -----DADSLLEKVKHLRQQTASLKVNI-EKIDPHEEMRYGLDSISMTFTNQLNRYRLILLPITLFFHPTIH
```

```

Ta-1_ACP-KS      59  DVARHVEGATQARTGATE-----
BonA_ACP-KS_2   55  QFAAFLLASHGERLAAFYDAGGVAAALS-----TAPASRTS-----VALPDEPRVDV-GAAAPLV
Bat2_ACP-KS     61  KLTSHFVVEHADLRYRQQEHVHGART-----RVTVRGE-----VSMP---AVDIPGLAREAV
VirA_ACP5b      65  ALSHVAG-----
OocJ_ACP-KS     51  DLRRYLLENYRHEVARMQPAQRQDAASA-----GNAAPRAE-----T
Ta-1_ACP-KS_M   64  LFAEYLARQLPTAPPPAPP-----QVEAKT-----VRVAPVVSSEPP-----S
EtnE_ACP-KS_M   71  QLAHHLATLDATAVRAAVDGISSAERAGAHESTARREREAAEREAARRAREDAERQVAARHHEEASFEVEARRASDA
BaeL_ACP-KS_M   63  QFAEFLRQPSVVRGKKQ--VP-----VR-----PKPLTQOHAP-----QEKTPAE--ERLTI-----
OocL_ACP-KS     47  VLAEYLTQSHGATPR-----PKPNAAVSEPEA-----PC-SQDL-----K
BonA_ACP-KS_M   45  RMAARLALALAEATFAAAMPAPATESAPAPA-----PAPAPASVPAQA-----PAVARDLPAA--GPLPAAAPASATDH
BonA_ACP-KS     63  QLSREAGAYASRFRVPGTEPMPMPRSFVP-----P-----E---AAAAPAVRV-----A
PedI_ACP-KS_M   38  EFSQYVRSLLAAQFAAVSRAEASDRAESLLLRRFEPDVTNQPRE-----EQEE-ASTQDEPIVRL-----L
OnnB_ACP-KS_M   42  DLAGEHLASEMSHRLATA-RPLAKLERERPRRT-----P-----RADA-----KPVVS-----P
PsyA_ACP-KS_M   64  AMCHHLQEEVSRKLLA-DAGEPSRVEMAAS-----IVEV-K-----TSDD-----A
EtnD_ACP-KS_M   80  RLAHLSELDVRAAG-RPVAETERERPALG-----GDIAA-----R
Bat2_ACP-KS_M   74  MLASHLLSQEAFRSRDA-SMNPAPGPTS-GWMATATASDKRNEPVS-----IAFV-KTLAEPQSWAP-----K
Bat3_ACP-KS_M   66  VLAHLLTRGSLRSREA-SEPAQALVATPLATLIQVPECNERVISTA-----ITTL-ETLALPLPLAP-----K
OocN_ACP-KS     57  ALARAVARMQTDAPAKEGRQNA-----
PedF_ACP-KS_2   73  RLAHHIVVEYCHTLDVPAALPCPE-----LSV-SEPAMDIP-----L
PsyD_ACP-KS_2   73  NLTQHQQRIEETRLRQMSGDA-----
OnnI_ACP-KS_2   73  ALADHISKTFQGDNLNIREKVKETGDLK-----KGITDEPKWEPD-----P
Bond_ACP-KS_M   46  ALAKLLASTGERGHAGAPTAEA---PPA-----PT-----RPADPEP-----SRSPQ
EtnE_ACP-KS     66  BLAYFVEHHAATLSLLCGATAPAAPATAVSA-----G-A-----GPA-----RRRAR
EtnF_ACP-KS     73  GLACHLAEEHREACQRWAGCPDGDAPDD-VARA-E--P-RG---PS-----SAAQ
EtnD_ACP-KS     55  ALAGHLSRHHAAACRWTCMDRLADAPVHANEP-----GGAM-----APLDAGA-----RRDEA
PedI_ACP-KS     59  ALSYLLREHQRITALTCLRPSAGLDQSGRS-----GER-----MV-----SS--
PsyD_ACP-KS     61  DLACHLANTGVEV-----R-----G-----P-----SPIV
OnnB_ACP-KS     60  ALVTHFLQHOREALNALMCIIDEFT---DSE-----P-----SEEV
PedF_ACP-KS     60  ALAYYLVNLTVSCRAWTGLRDESVLVADAAR-----RGLP-----LP-----ET--
OnnI_ACP-KS     74  ELVSLLDLDCPHACTRWVGRFEAPPVREPEPET-E--TVNGLC---EP-----ASQV
VirA_KS         1  -----
BaeM_ACP-KS     60  BLAYYLKKTVSFECE-----RAAEPS-----RSEA
Bat3_ACP-KS     61  KLAAYLVERYPSELALGLKIDDTQRAKQPAV-----VS-----RTTS
OocL_ACP-KS_M   61  ELTEIITITRFQQAARYDPSHRSDGRQPPASAT--KPENGP-----AAGEKSA-----RSIQA
PsyD_ACP-KS_3   75  ALAHLVEYHQTQQMCAGEPEVCD-EFKVE-----A-----
Bat3_ACP-KS_2   61  DFAYYLVEKHGEAFVCFE---PPVQTSGSKE---QP-----VPRAEVP-----VSV--
Ta-1_ACP-KS_2   62  ALAGYLCESHGPRLLDQLCAEAAAFEPASPPS-----AP-----MPVEE
TaO_ACP-KS     62  ALVGRLVSEHRPLAERHRGVGGPVADTSP-E-----DT-----LPTPEVS
BaeJ_ACP-KS     72  GLAGYLAKHE---PGRFGEKKESPKK---E-----QP-----KAQKEKM
BaeN_ACP-KS     72  EFAYHLLISEYEEEFAGRAVAVNKTQVQSARPVK-----ET-----QITREAV

```

Ta-1_ACP-KS 78 PAR-----E-----DTSERDYEGSLAVIGLSCQLPGAADPWFVKNLREGRDGVVAYRHHELRE--LGV-----PE
 BonA_ACP-KS_2 109 PAP-----ADAPRAR--DAVVRDDEPIAIVGMSGRFPGARNVDFEWTILRDGLDMWRPAPAQGP-----AD
 Bat2_ACP-KS 110 PAD-----SQASTVA--RRGLAPOEFMAIIGMSGRFPGARDVDFEWHVLEGRDAVQOEIPLTRDWRNYYCAPDTSVGAQ
 VirA_ACP5b
 OocJ_ACP-KS 89 PFP-----SPRREQAAEDKREHDGDIIVIGMSGRFPGMANVDFEWMVLAISGDCITSEVPAADRWDAAYYD-----ED
 Ta-1_ACP-KS_M 104 VQA-----VQAPAIVKKEVAPAGTVRIAVIGMSGRFPGANLRQWVNDLRDGVSCISELPIYTRWDVATA-----K
 EtnE_ACP-KS_M 151 VAC-----VAVRERVIAAGPGEAACRIAVIGMSGRFPGDLDEWVSNLARGVDAVAEVPASRWSVERYYD-----PA
 BaeL_ACP-KS_M 109 -OK-----EPEFVQAKREPKEDEIAIAIIVGMSGRFPGAPLPLKTVWVNDLRAKNAIRDIPLSRWVWVNYD-----PA
 OocL_ACP-KS 82 PAA-----SDAPAEAT--AASHRRDEKIALIGMAGRFPDANLDQWVNDLARGNSVREWFRERWVLDQYED-----AD
 BonA_ACP-KS_M 113 ARP-----AASPVQGA--TPPHREHPIAIVIGMAGRFPDANLDQWVNDLARGNSVREWFRERWVLDQYED-----AD
 BonA_ACP-KS 106 BIA-----P-----VPPSSIEPIAIVGMSGRFPGAGNRELVAAALAGGEDLIE--PVTRWVWVNYD-----D
 PedI_ACP-KS_M 101 PCA-----PP-----VRQPTDIPAIIVIGMSGRFPGAPLPLKTVWVNDLRAKNAIRDIPLSRWVWVNYD-----S
 OnnB_ACP-KS_M 84 PAS-----QPLTPVSKRWLPEQKTGRIIVIGMSGRFPGASTLDQWVNDLARGNSVREWFRERWVLDQYED-----AD
 PsyA_ACP-KS_M 104 EIS-----DVEGDMPAIAIVIGMSGRFPGANNVVALWVNDLARGNSVREWFRERWVLDQYED-----PT
 EtnD_ACP-KS_M 116 PAR-----EPEARAP--RPAPEHPIAIVIGMAGRFPKARDLDEFWENLARGVDCVTEIPDRWVAEAYD-----PR
 Bat2_ACP-KS_M 135 PAS-----PPEELVS--GTQVLLIGIIVIGMAGRFPKARNAQVFWENLAQGLDVCSEIPSSRWSVEEYD-----PN
 Bat3_ACP-KS_M 128 PAS-----PPEESVP--GTQVLLIGIIVIGMAGRFPKARNAQVFWENLAQGLDVCSEIPSSRWSVEEYD-----PN
 OocN_ACP-KS 80 -----APG-----PLPGSGDIAIIGLDINVAQADNAAEFWQLREGRSSVGVQVPOGRQREDLS-----M
 PedF_ACP-KS_2 109 PAV-----QAVPSSLPREAVVQTIGIIVIGMAGRFPGADSVDAIWNVAVGNVPTSELYLPYHAYS-----P
 PsyD_ACP-KS_2 96 -I-----AEVD-----MSNPSEHPIAIVIGMAGRFPGADNVDVFWENLVQGVDAVGEIPARYLSREAYS-----S
 OnnI_ACP-KS_2 113 PA-----DMSVEFGEKDFGSSPRIAIVIGMSGRFPGAKDMESFWQNLISGDDTIGELPPHYLPEESH-----P
 Bond_ACP-KS_M 88 EV-----AASAAPSSLGAPADFEPIAIVIGMSGRFPDANLDQWVNDLARGNSVREWFRERWVLDQYED-----ED
 EtnE_ACP-KS 109 RLPS-----RGRAGEAREAAIAIIGLSGRYPQANDLAEFWENLRSKDCITEIPDRWVWVNYD-----AE
 EtnF_ACP-KS 115 BARS-----GERAPAEALPAQELDVAIIGMAGRYPHARTLDEFWENLRAKDCITEIPDRWVWVNYD-----GA
 EtnD_ACP-KS 104 ELPR-----SEEAAGARPARPEPIAIVIGLSGRYPQANDLAEFWENLRAKDCITEIPDRWVWVNYD-----ED
 PedI_ACP-KS 99 -----SISAPSSSEHPIAIVIGMSGRFASADLDEWQVLAAGESCISEVPAARWVWVNYD-----ED
 PsyD_ACP-KS 81 TAPR-----LA-VASQLPSRAEVEPIAIVIGLSGRYPGSPDLDAFWENLRAKDCITEIPDRWVWVNYD-----AS
 OnnB_ACP-KS 92 VAPA-----AS-VRGAPSNRLREPIAIVIGMSGRFPDANLDQWVNDLARGNSVREWFRERWVLDQYED-----AD
 PedF_ACP-KS 100 -APV-----VERNVLVPG--NAVQEPPIAIVIGLSGRYPQANDLAEFWENLRAKDCITEIPDRWVWVNYD-----ED
 OnnI_ACP-KS 120 BOPH-----ARSAKKSCNSVQOEPPIAIVIGLSGRYPQANDLAEFWENLRAKDCITEIPDRWVWVNYD-----AD
 VirA_KS 1 -IAVIGVGRFPADLDRWVNLREGRDITEVPELDRWDVAYYD-----PD
 BaeM_ACP-KS 86 BAPD-----TPEHHKNTGDIIVIGMAGRFPKAKSVAEFWENLRAKDCITEIPDRWVWVNYD-----NV
 Bat3_ACP-KS 99 SDPE---AGVNAQRTLLSTGEPSOPIAIVIGMAGRFPAAQNVDFEWNLRKAGKDCITEIPASRWDWRRE---D-----SL
 OocL_ACP-KS_M 114 LLAKSHQIRQKRQPVVVPANGRYEPIAIVIGMSGRFPAQSSVVALQRHLYQGRDCISEVFNARWVWVNYD-----NP
 PsyD_ACP-KS_3 107 -----LDAQTPVEPVRTGPEPIAIVIGMSGRFPGSPDLETFWRHLEANADLITEIPADRWVWVNYD-----DP
 Bat3_ACP-KS_2 101 DAVHRRSRLNDLPHVPVLRQVPEPIAIVIGLSGRFPGARDLDEFWNTLQCKDCISEIPDRWVWVNYD-----DP
 Ta-1_ACP-KS_2 101 --ARSE---E----RVAPQPRAREPIAIVIGMAGRFPDADLDAFWGNLVAGKDCIREVFNRSRWDWRVYD-----DP
 TaO_ACP-KS 102 --RRSGRRSPGFSAAV--KVPQGRPEPIAIVIGMSGRFPGAPDLAAFWRNLAAGRDCISEIPADRWVWVNYD-----DP
 BaeJ_ACP-KS 106 --QRKKRFATVMNASAAATQEPRRFPVAIVIGLSGRFPGAKDLEEFWRNLKECKDSEITLIPKDRWVWVNYD-----DP
 BaeN_ACP-KS 113 --KRRRTLPETL--PQTVQOQGPPEPIAIVIGLSGFPEMAEDLEAVWQNLKECKDCITEIPKDRWVWVNYD-----DP

| PKS System | Subunit name | Length of linker following β -modification module | Length of linker following non-modifying module |
|----------------|--------------|---|---|
| Myxovirescin | Ta-1 | 51 | 24 |
| | Ta-1 | | 46 |
| | TaO | | 54 |
| Bonkrekic Acid | BonA | 78 | 63 |
| | BonA | | 46 |
| | BonD | | 52 |
| Kalimanticin | Bat2 | 70 | 58 |
| | Bat3 | 71 | 56 |
| | Bat3 | | 51 |
| Oocydin | OocJ | | 49 |
| | OocL | 69 | 45 |
| | OocN | | 24 |
| Etnangien | EtnE | 91 | 48 |
| | EtnD | 44 | 56 |
| | EtnF | | 52 |
| Bacillaene | BaeL | 56 | |
| | BaeJ | | 48 |
| | BaeN | | 53 |
| | BaeM | | 30 |
| Pederin | PedI | 67 | 41 |
| | PedF | | 47 |
| | PedF | | 48 |
| Onnamide | OnnB | 53 | 41 |
| | OnnI | | 49 |
| | | | 56 |
| Psymberin | PsyA | 41 | |
| | PsyD | | 26 |
| | PsyD | | 29 |
| | PsyD | | 38 |
| | | Average: 62.8 | Average: 45.6 |

b

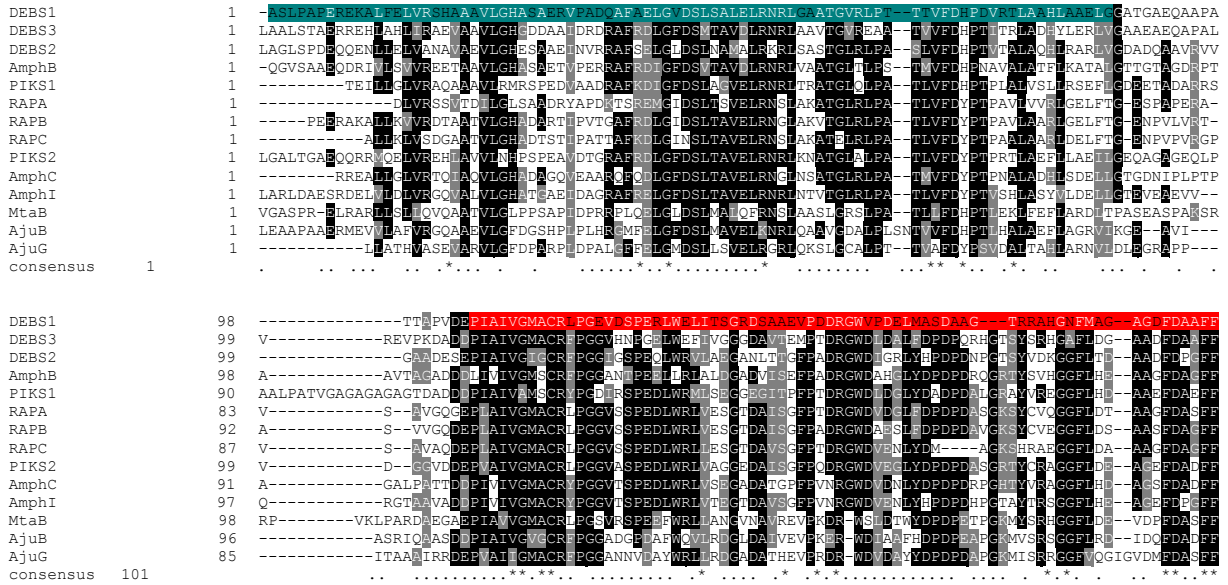


Fig. S9 Analysis of ACP-KS linkers. (a) Portion of a multiple alignment of the ACP-KS regions (junction between two modules) from a selection of *trans*-AT PKS subunits, to determine the lengths (shown in the table) of the linkers intervening between the conserved ACP and KS domains. The C-terminal and N-terminal boundaries of the ACP and KS domains, respectively, were determined by alignment with VirA module 5 KS and ACP_{5b}. Linkers which follow modules where β-modification occurs are indicated by M (2 indicates a second junction from the same subunit). Although only 11 sequences of such linkers are available in the database (that is, both the gene sequence and the product structure are known, validating the β-modification), these linkers are generally longer than comparable regions from the same systems, and the average length of all these linkers (ca. 63 residues) is significantly greater than the average length of linkers separating non-modifying modules (ca. 46). Key: Ta, myxovirescin; Bon, bonkrekic acid; Bat, kalimantacin; Vir, virginiamycin; Ooc, oocydin; Etn, etnangien; Bae, bacillaene from *Bacillus amyloliquefaciens*; Ped, pederin; Onn, onnamide; Psy, psymberin. (b) Multiple alignment of the ACP-KS regions from a selection of *cis*-AT PKS from both *Streptomyces* and myxobacteria. The C-terminal domain boundary of the ACPs was based on the NMR solution structure of DEBS ACP₂.⁷ The average linker length is 20 residues. Key: DEBS, erythromycin; Amph, amphotericin; PIKS, pikromycin; RAP, rapamycin; Mta, myxothiazol; Aju, ajudazol.

Supplementary Tables

Table S1 Summary of SAXS Data

| Construct | R_g (Guinier) (Å) ^a | R_g (GNOM) (Å) ^a | D_{max} (GNOM) (Å) ^b | χ^2 (DAMMIN) ^c | χ^2 (SASREF) ^c | χ^2 (MONSA) ^c |
|---|----------------------------------|-------------------------------|-----------------------------------|--------------------------------|--------------------------------|-------------------------------|
| KS | 33.1 ± 0.5 | 32.6 | 105 | 1.689 | 1.333 | 1.421 |
| ACP _{5a} -ACP _{5b} | 29.0 ± 0.2 | 30.7 | 103 | 1.785 | 24.981 | nd |
| KS-ACP _{5a} -ACP _{5b} | 56.5 ± 0.7 | 60.9 | 230 | 2.813 | nd | 1.885 |
| KS-ACP _{5a} -ACP _{5b} -DD | 69 ± 1 | 72.5 | 279 | 2.204 | nd | nd |
| ACP _{5b} -DD | 37.0 ± 0.3 | 37.7 | 140 | 1.231 | nd | nd |
| ACP _{5a} -ACP _{5b} -DD | 43.5 ± 0.2 | 46.5 | 170 | 1.366 | nd | nd |

^a R_g is the radius of gyration derived from the atomic models using the program PRIMUS. For experimental data, R_g is given by the Guinier approximation⁸

^b D_{max} is the maximal particle diameter derived from the distance distribution function ($P(r)$) using the program GNOM⁹

^c χ^2 is the discrepancy between the experimental SAXS profile and the theoretical calculated curve using the programs DAMMIN,⁹ SASREF¹⁰ or MONSA¹⁰

nd: not determined

Table S2 NMR and Refinement Statistics for ACP_{5a}

| | Protein |
|--|-------------------|
| NMR distance and dihedral constraints | ACP _{5a} |
| Distance constraints | |
| Total NOE | 1968 |
| Intra-residue | 374 |
| Inter-residue | 1594 |
| Sequential ($ i - j = 1$) | 490 |
| Medium-range ($ i - j < 4$) | 566 |
| Long-range ($ i - j > 5$) | 538 |
| Intermolecular | n/a |
| Hydrogen bonds | 0 |
| Total dihedral angle restraints | 114 |
| ϕ | 57 |
| ψ | 57 |
| Structure statistics | |
| Violations (mean and s.d.) | |
| Distance constraints (Å) | 0.092 ± 0.012 |
| Dihedral angle constraints (°) | 0 ± 0 |
| Max. dihedral angle violation (°) | 0 |
| Max. distance constraint violation (Å) | 0.13 |
| Deviations from idealized geometry | |
| Bond lengths (Å) | 0.010 ± 0.001 |
| Bond angles (°) | 2.062 ± 0.018 |
| Impropers (°) | n/d |
| Average pairwise r.m.s. deviation** (Å) | 6734–6811 |
| Heavy | 0.94 ± 0.10 |
| Backbone | 0.27 ± 0.06 |

**Pairwise r.m.s. deviation was calculated using 20 refined structures. n/a = not available; n/d = not determined.

Table S3 NMR and Refinement Statistics for ACP_{5b}

| | Protein |
|--|-------------------|
| NMR distance and dihedral constraints | ACP _{5b} |
| Distance constraints | |
| Total NOE | 1846 |
| Intra-residue | 371 |
| Inter-residue | 1475 |
| Sequential ($ i - j = 1$) | 500 |
| Medium-range ($ i - j < 4$) | 522 |
| Long-range ($ i - j > 5$) | 453 |
| Intermolecular | n/a |
| Hydrogen bonds | 0 |
| Total dihedral angle restraints | 110 |
| ϕ | 55 |
| ψ | 55 |
| Structure statistics | |
| Violations (mean and s.d.) | |
| Distance constraints (Å) | 0.096 ± 0.023 |
| Dihedral angle constraints (°) | 0 ± 0 |
| Max. dihedral angle violation (°) | 0 |
| Max. distance constraint violation (Å) | 0.17 |
| Deviations from idealized geometry | |
| Bond lengths (Å) | 0.010 ± 0.001 |
| Bond angles (°) | 2.054 ± 0.018 |
| Impropers (°) | n/d |
| Average pairwise r.m.s. deviation** (Å) | 6836–6909 |
| Heavy | 1.15 ± 0.15 |
| Backbone | 0.38 ± 0.07 |

**Pairwise r.m.s. deviation was calculated among 20 refined structures. n/a = not available; n/d = not determined.

Supplementary References

- 1 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W.H. Gerwick, D. H. Sherman, D.H., and Smith, J.L., *Chem. Biol.*, 2013, **20**, 1340–1351.
- 2 Y. Tang, C. -Y. Kim, I. I. Mathews, D. E. Cane, and C. Khosla, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 11124–11129.
- 3 T. Bretschneider, J. B. Heim, D. Heine, R. Winkler, B. Busch, B. Kusebauch, T. Stehle, G. Zocher, and C. Hertweck, *Nature*, 2013, **502**, 124–128.
- 4 A. Mohan, C. J. Oldfield, P. Radivojac, V. Vacic, M. S. Cortese, A. K. Dunker, and V. N. Uversky, *J. Mol. Biol.*, 2006, **362**, 1043–1059.
- 5 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky, *Biochem. Biophys. Acta*, 2010, **1804**, 996–1010.
- 6 H. P. Erickson, *Biol. Proced. Online*, 2009, **11**, 32–51.
- 7 V. Y. Alekseyev, C. W. Liu, D. E. Cane, J. D. Puglisi, and C. Khosla, *Protein Sci.*, 2007, **16**, 2093–2107.
- 8 P. V. Konarev, V. V. Volkov, A.V. Sokolova, M. H. J. Koch, and D. I. Svergun, *J. Appl. Crystallog.*, 2003, **36**, 1277–1282.
- 9 D. I. Svergun, *J. Appl. Crystallog.*, 1992, **25**, 495–503.
- 10 M. V. Petoukhov, and D. I. Svergun, *Biophys. J.*, 2005, **89**, 1237–1250.