Electronic Supplementary Material (ESI) for Chemical Science.

Supplementary Information

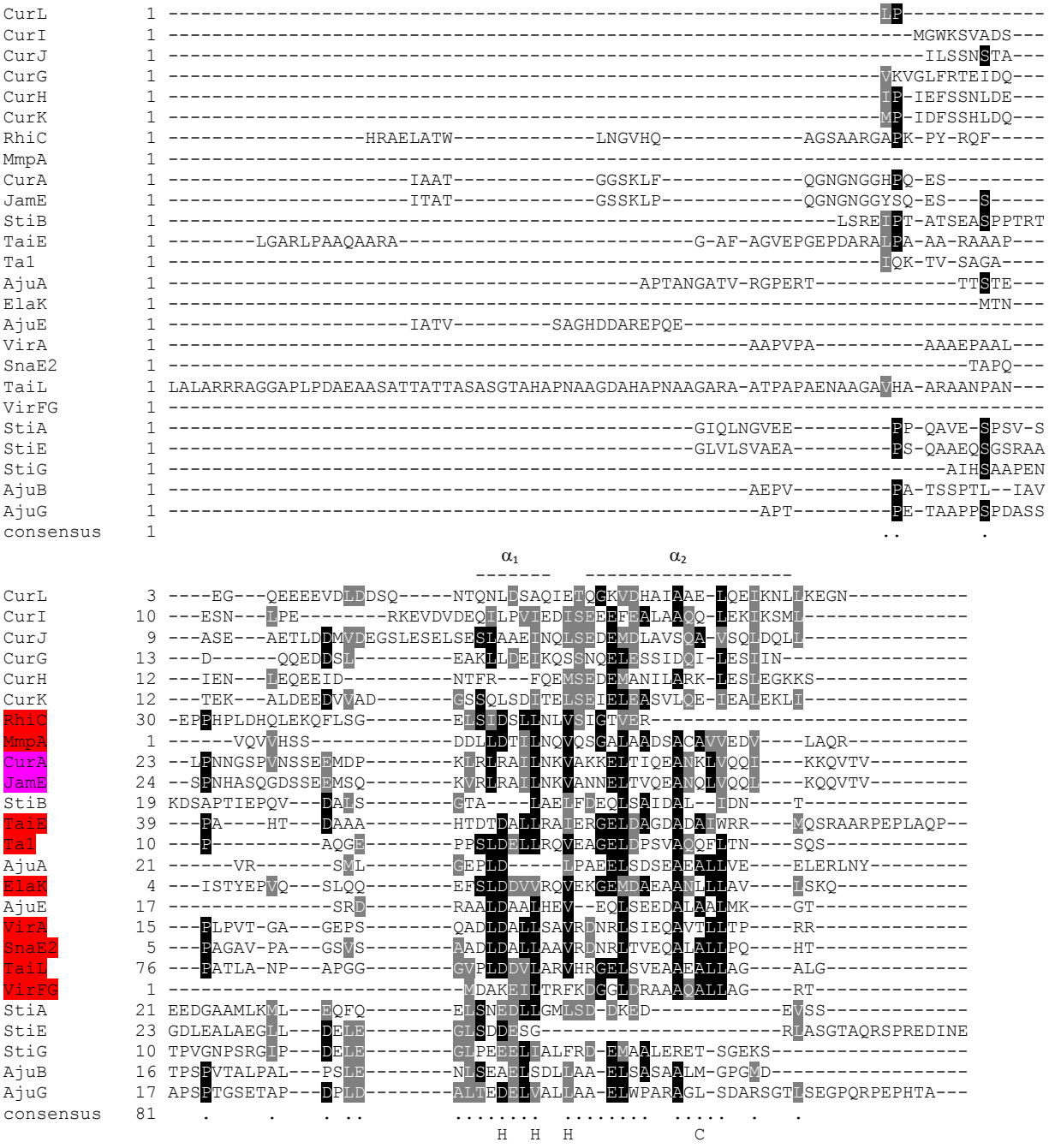
Insights into the function of *trans*-acyl transferase polyketide synthases from the SAXS structure of a complete module

Jack Davison[1], Jonathan Dorival[1], Hery Rabeharindranto[1,2], Hortense Mazon[1], Benjamin Chagot[1,*], Arnaud Gruez[1,*] & Kira J Weissman[1,*]

[1]Molecular 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.
[2]Present 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



b

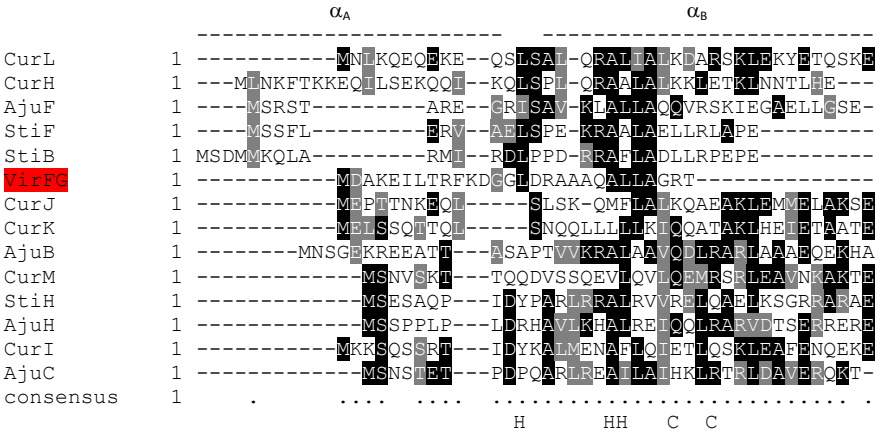


Fig. S1 Sequence analysis of putative dimerization/docking regions in various PKS. (a) Multiple alignment of the C-terminal regions of PKS subunits from *trans*-AT PKS (in red), with those of *cis*-AT PKS[1] (black, except for those of subunit CurA and the closely-related JamE (in pink)). The sequences 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, pristinamycin (*trans*-AT).

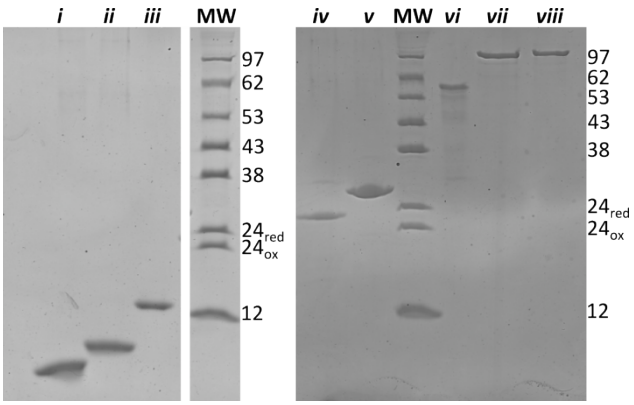
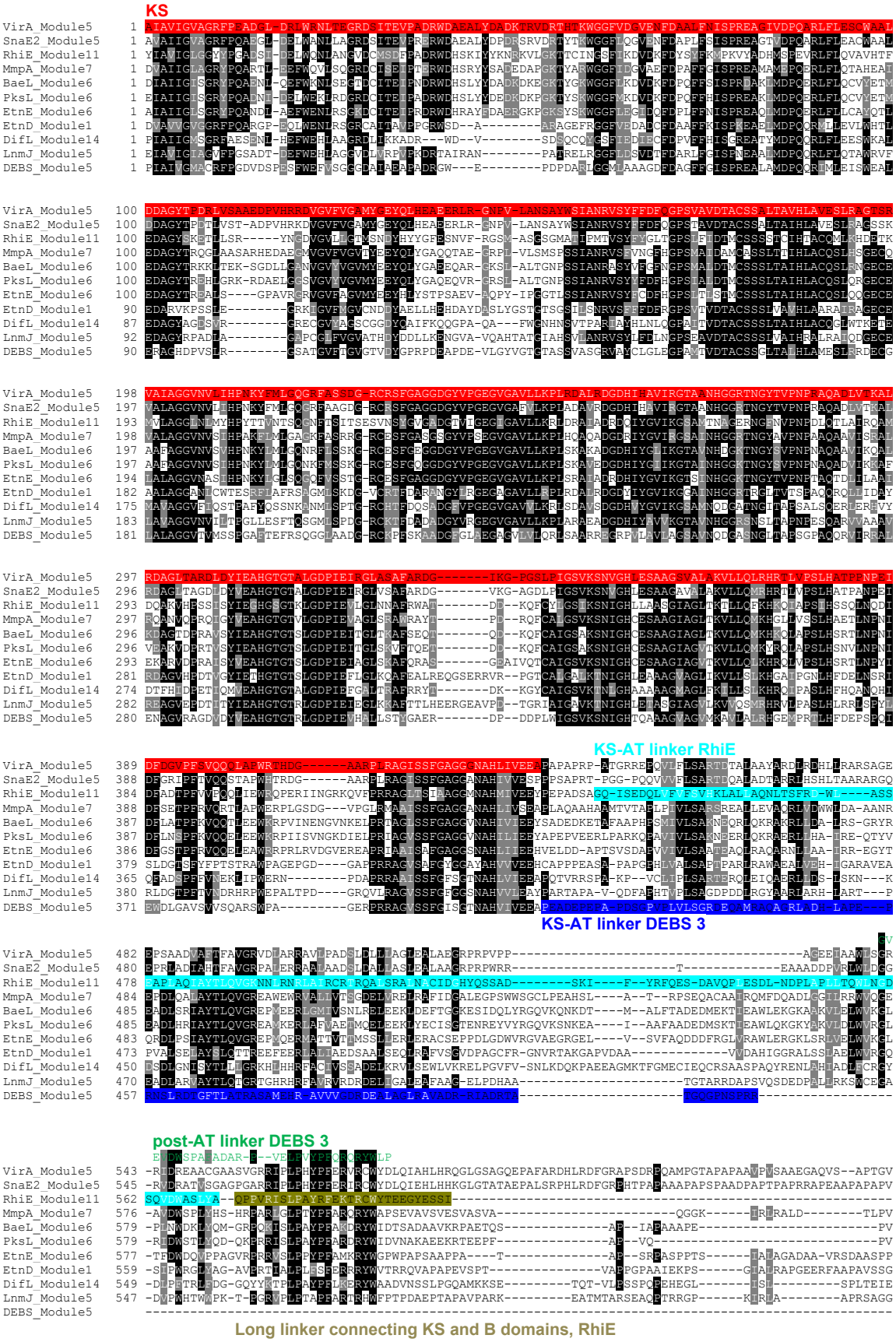


Fig. S2 SDS-PAGE analysis of protein preparations used in this study. Key to lanes: i) ACP_{5a} (calc'd: 8.9 kDa); ii) ACP_{5b} (calc'd: 9.2 kDa); iii) ACP_{5b}-DD (calc'd: 14.3 kDa); iv) ACP_{5a}-ACP_{5b} (calc'd: 19.4 kDa); v) ACP_{5a}-ACP_{5b}-DD (calc'd: 24.5 kDa); vi) KS_5-linker (calc'd: 62.2 kDa); vii) KS_5-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.


KS
KS-AT linker RhiE
KS-AT linker DEBS 3
post-AT linker DEBS 3
Long linker connecting KS and B domains, RhiE

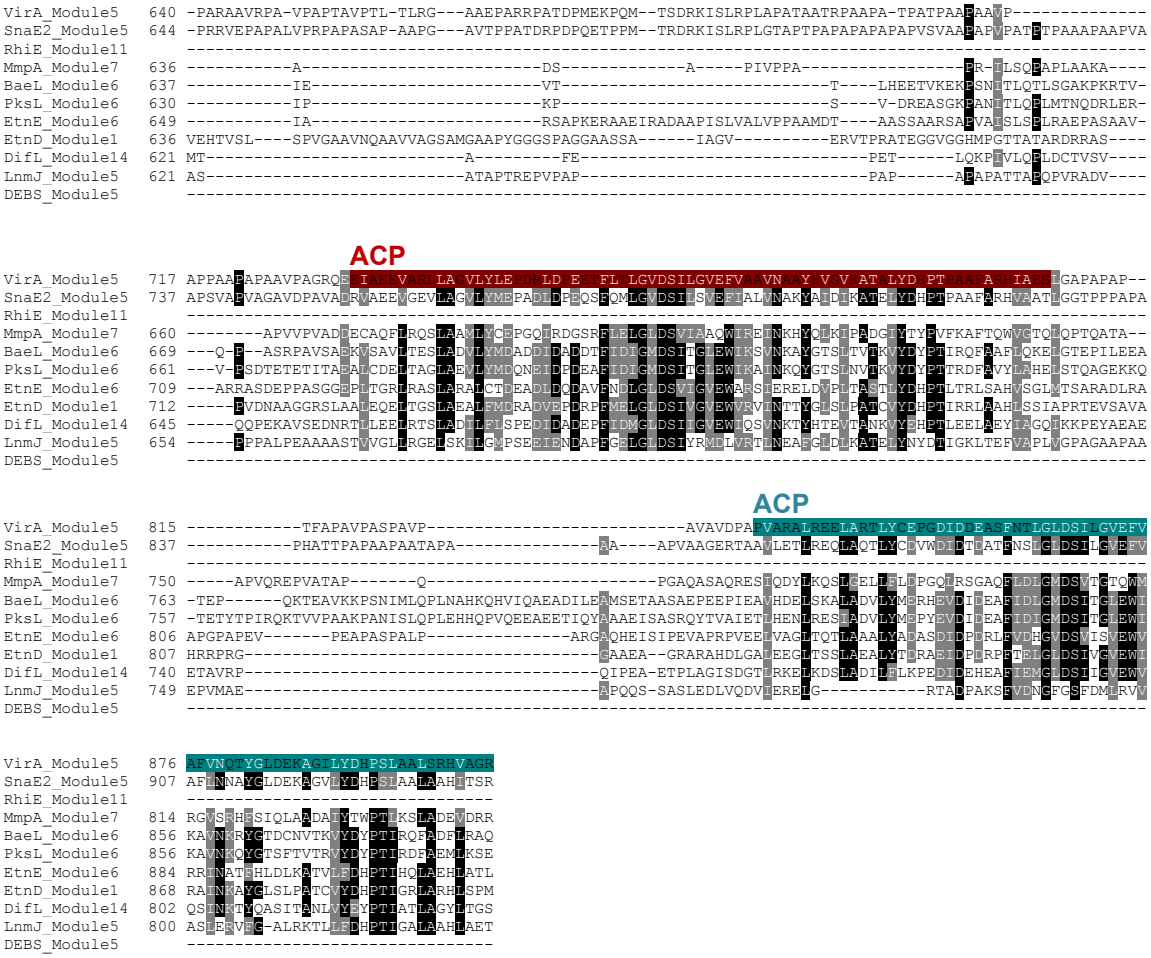
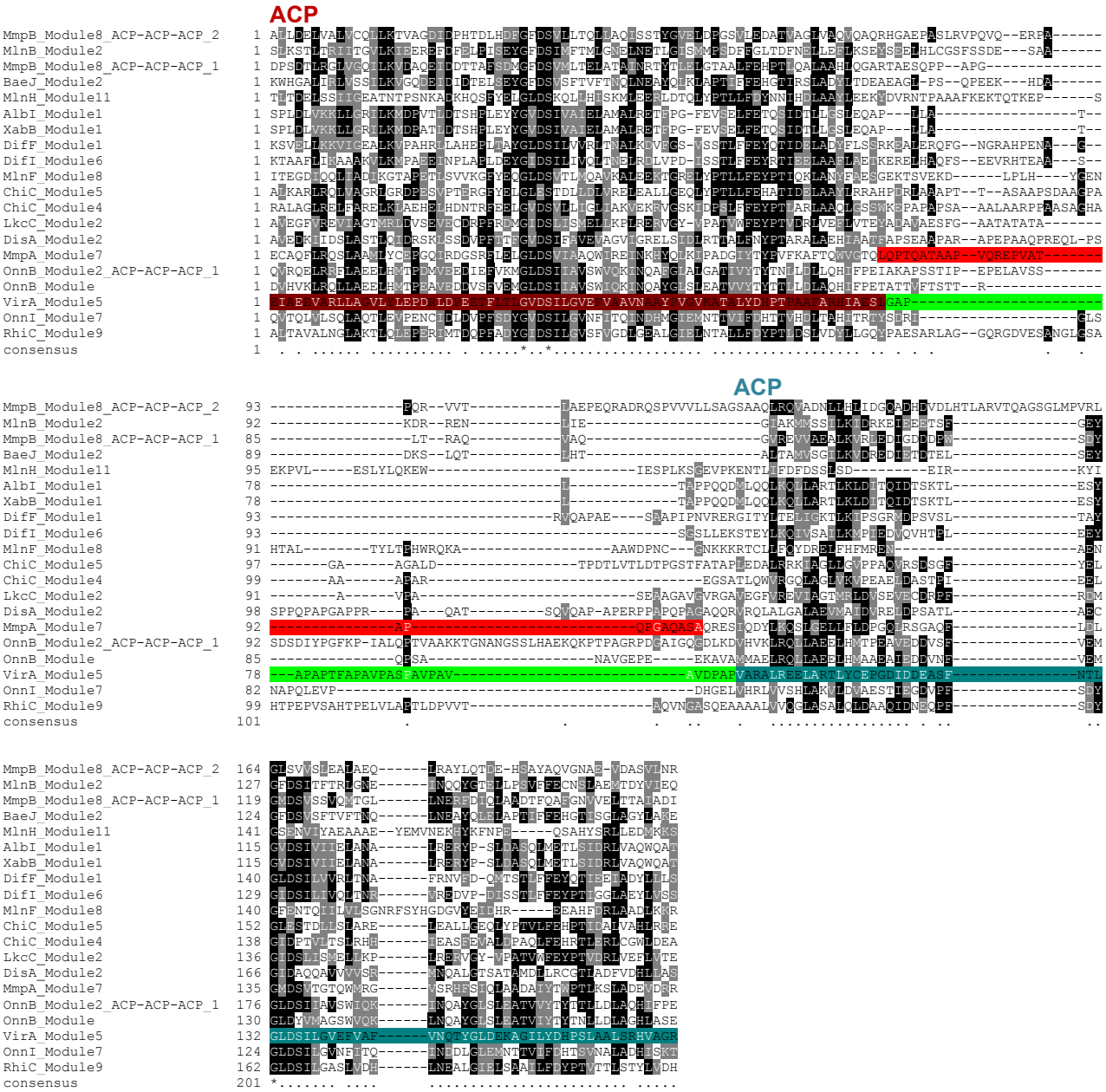


Fig. S3 Sequence analysis of a selection of modules (type KS-ACP-ACP and KS-ACP-ACP-ACP) from *trans*-AT PKS showing the homology to the KS-AT and post-AT linker regions of a typical *cis*-AT PKS module. The sequence of the KS-AT and post-AT linkers from module 5 of the erythromycin (DEBS) *cis*-AT PKS, 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 RhiE[3] 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 KS_5-AT_5 didomain of DEBS[2] and the NMR structures of VirA ACP_{5a} and ACP_{5b} (this work)) within the modules are also indicated in color. Key: Dif, difficidin; Vir, virginiamycin; Sna, pristinamycin; 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



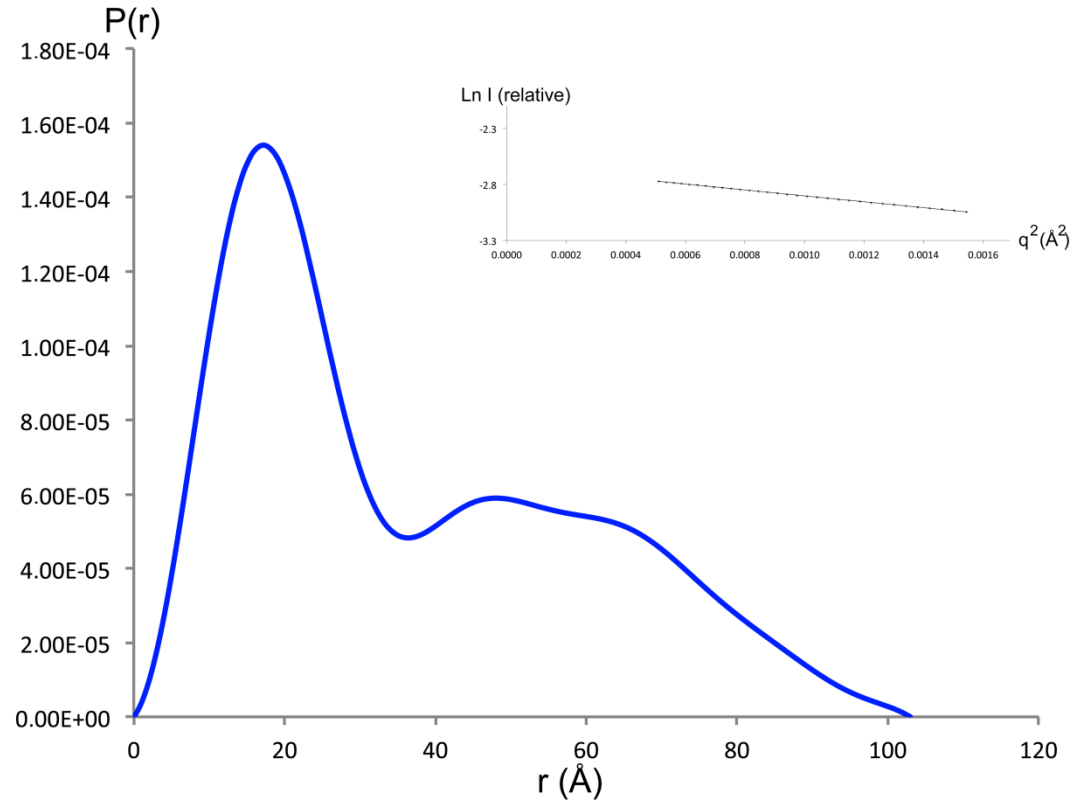
b

```
                                                    M LEVVAEKTGY PTEMLDLDMD MEADLGIDSI
1261 KRVEILGTVQ DQMPNLPELS PEDLAECRTL REIVTYMNSK MPAAAASAPV TSASNGLDAA
1321 QVQSTMLEVV ADKTGYPTEM LDLEMDMEAD LGIDSIKRVE ILGTVQDQLP TLPELSPEDL
1381 AECRTLGEIV SYMNSKLPAA SAVAAPVASA PVASASNGLD AAQVQTTMLD VVADKTGYPT
1441 EMLDLAMDME ADLGIDSIKR VEILGTVQDQ LPGLPELNPE DLAECRTLGE IVDYMNSKLP
1501 AVSTQNVAIQ TAAPVASASN GLDAAQVQGT MLEVVADKTG YPAEMLDFAM DMEADLGIDS
1561 IKRVEILGTV QDQLPGLPEL NPEDLAECRT LGEIVDYMNS KLPAASTQNV AVQTVAAPVA
1621 TAPATNGLDA AHVQNTMMNV VADKTGYPAE MLDLAMDMEA DLGIDSIKRV EILGTVQDQL
1681 PGLPELNPED LAECRTLGEI VAY
```

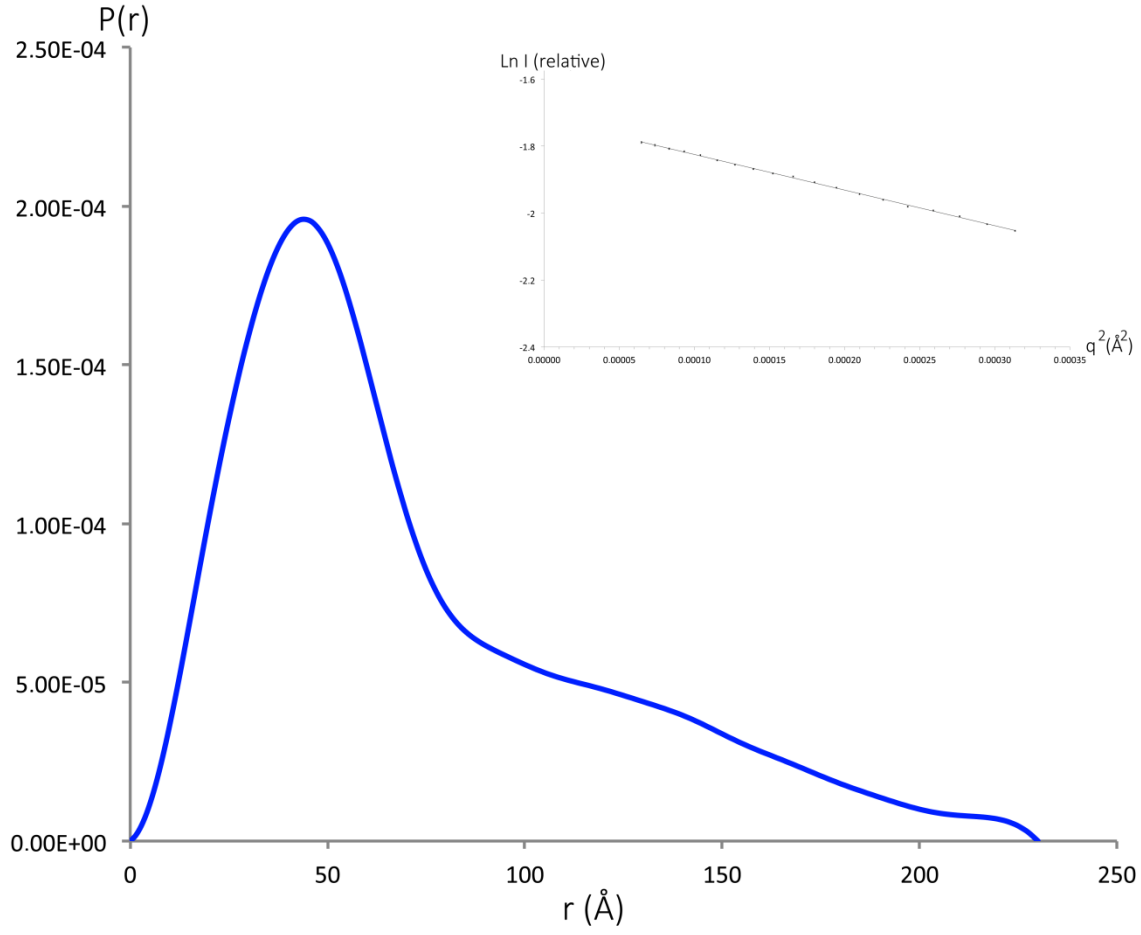
Linker 1 = 24 residues (P+A = 10; 42%)
Linker 2 = 29 residues (P+A = 14; 48%)
Linker 3 = 31 residues (P+A = 10; 32%)
Linker 4 = 34 residues (P+A = 13; 38%)

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.

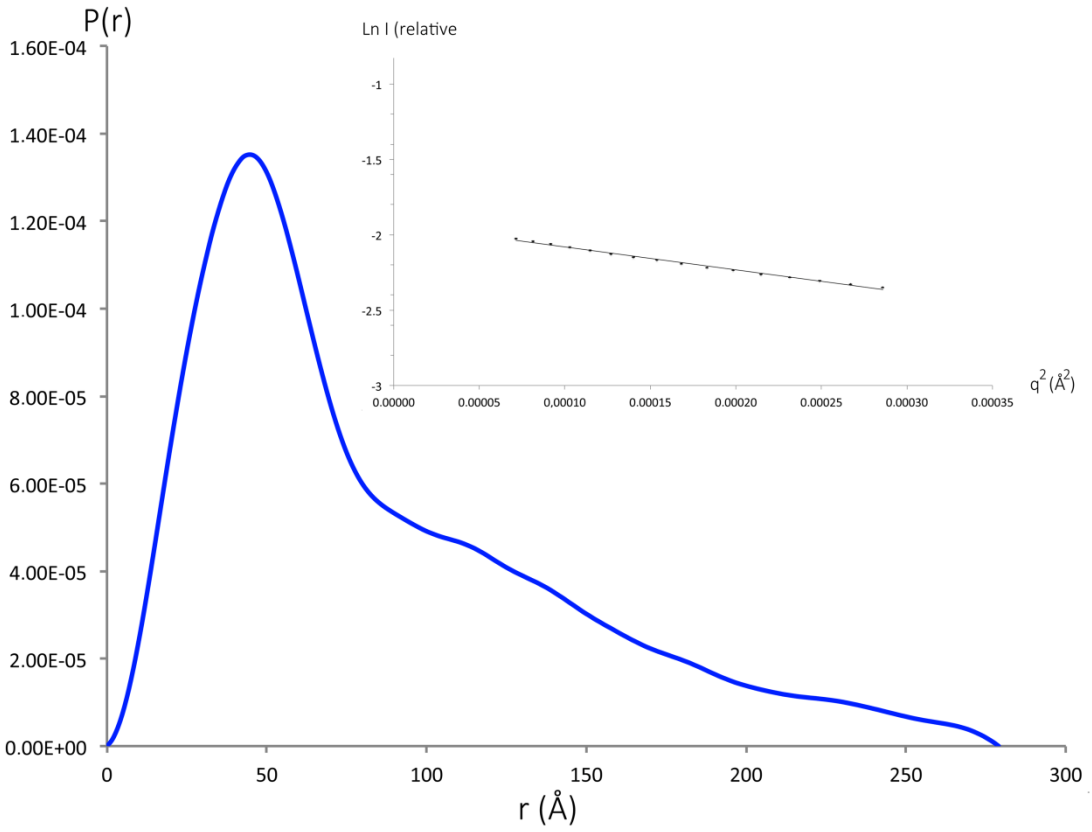
A


P(r)
1.80E-04
1.60E-04
1.40E-04
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
0
20
40
60
80
100
120
r (Å)
Ln I (relative)
-2.3
-2.8
-3.3
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
0.0014
0.0016
q2(Å2)

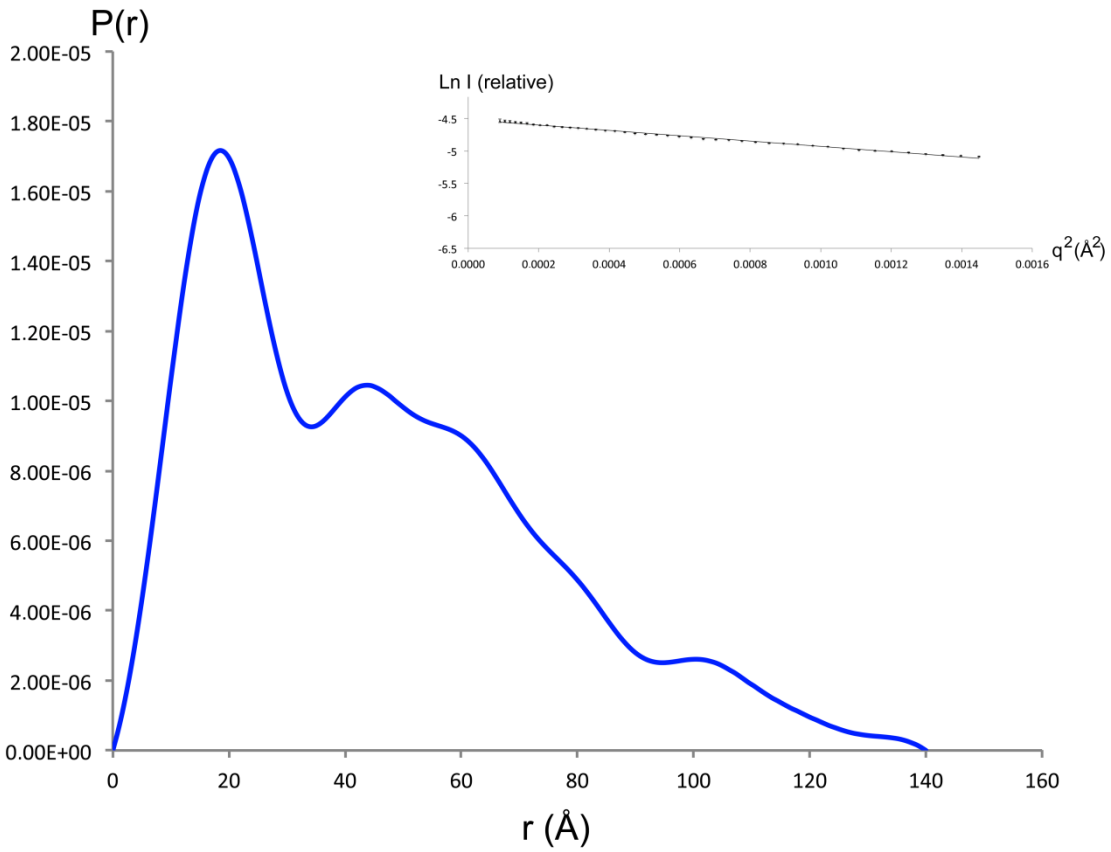
B


P(r)
2.50E-04
2.00E-04
1.50E-04
1.00E-04
5.00E-05
0.00E+00
0
50
100
150
200
250
r (Å)
Ln I (relative)
-1.6
-1.8
-2
-2.2
-2.4
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
0.00035
q2(Å2)

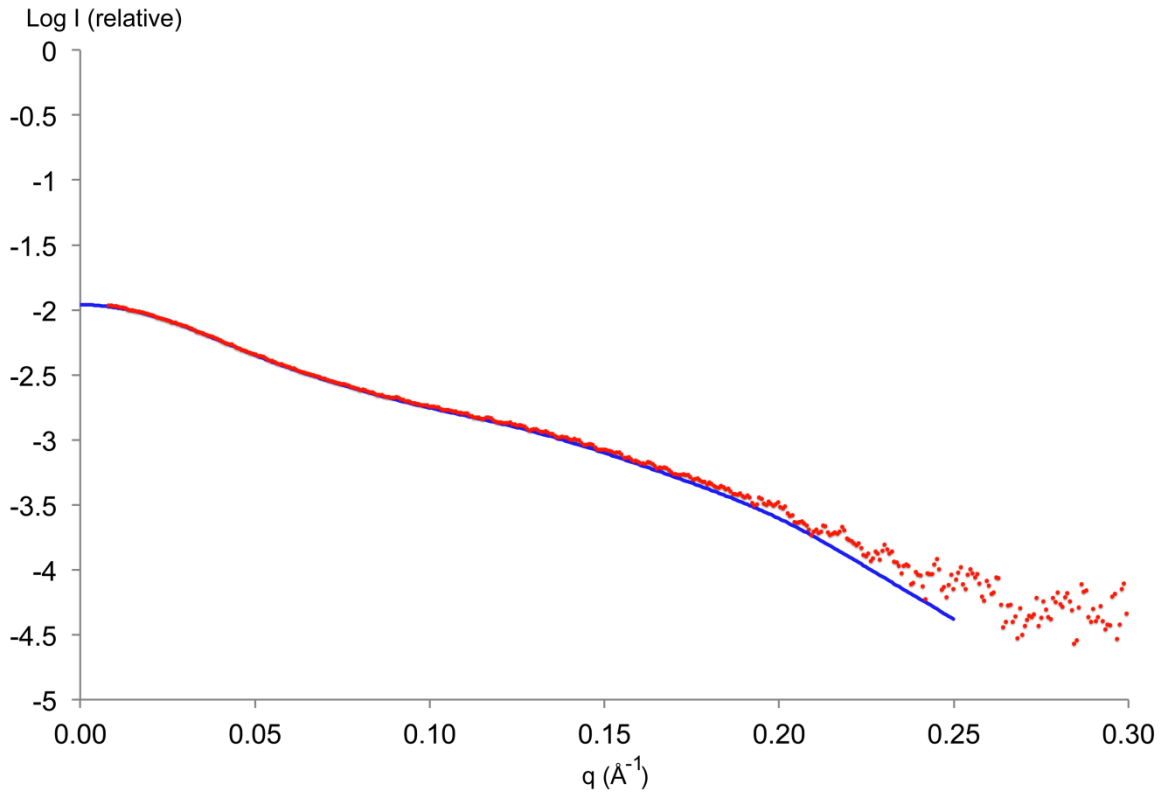
C


P(r)
1.60E-04
1.40E-04
1.20E-04
1.00E-04
8.00E-05
6.00E-05
4.00E-05
2.00E-05
0.00E+00
0
50
100
150
200
250
300
r (Å)
Ln I (relative
-1
-1.5
-2
-2.5
-3
0.00000
0.00005
0.00010
0.00015
0.00020
0.00025
0.00030
0.00035
q2 (Å2)

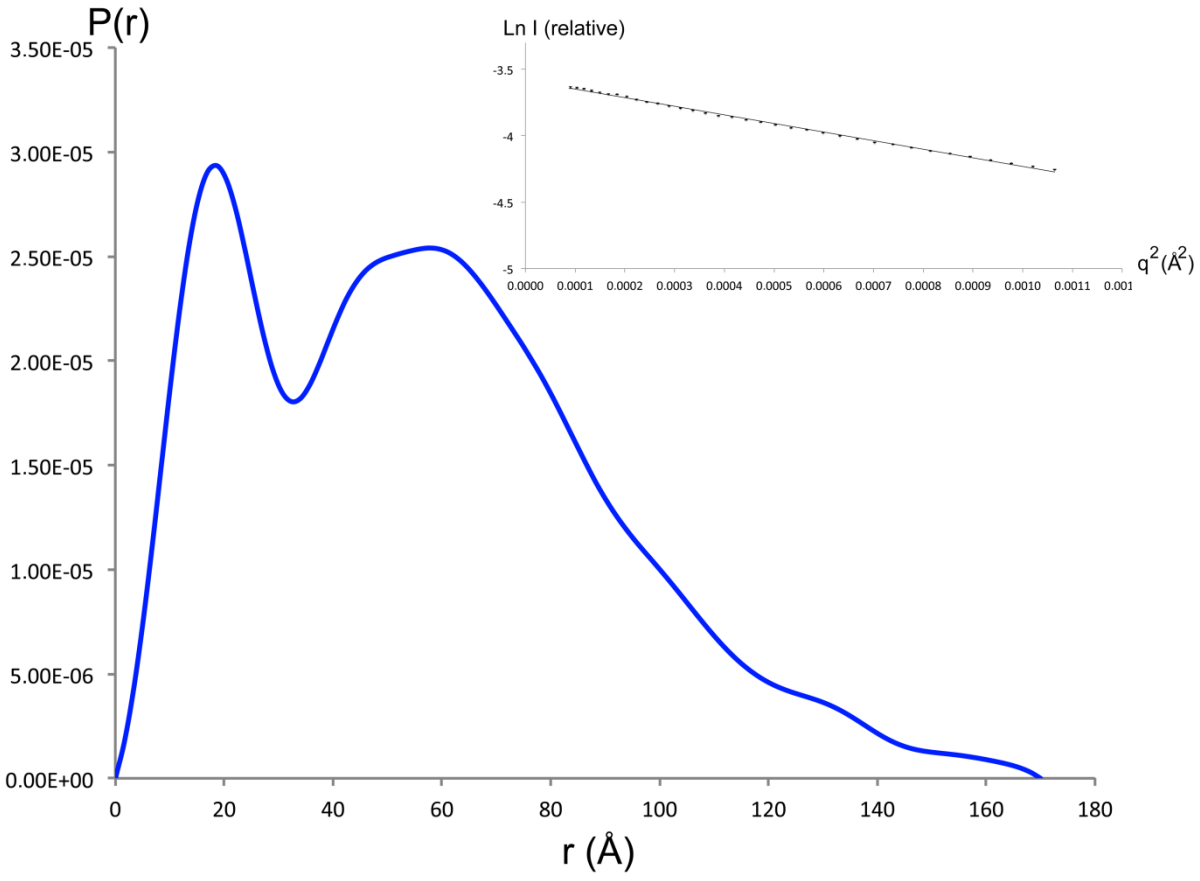
D


P(r)
2.00E-05
1.80E-05
1.60E-05
1.40E-05
1.20E-05
1.00E-05
8.00E-06
6.00E-06
4.00E-06
2.00E-06
0.00E+00
0
20
40
60
80
100
120
140
160
r (Å)
Ln I (relative)
-4.5
-5
-5.5
-6
-6.5
0.0000
0.0002
0.0004
0.0006
0.0008
0.0010
0.0012
0.0014
0.0016
q2(Å2)

E


Log I (relative)
0
-0.5
-1
-1.5
-2
-2.5
-3
-3.5
-4
-4.5
-5
0.00
0.05
0.10
0.15
0.20
0.25
0.30
q (Å-1)

F


P(r)
3.50E-05
3.00E-05
2.50E-05
2.00E-05
1.50E-05
1.00E-05
5.00E-06
0.00E+00
0
20
40
60
80
100
120
140
160
180
r (Å)
Ln I (relative)
-3.5
-4
-4.5
-5
0.0000
0.0001
0.0002
0.0003
0.0004
0.0005
0.0006
0.0007
0.0008
0.0009
0.0010
0.0011
0.001
q2(Å2)

G

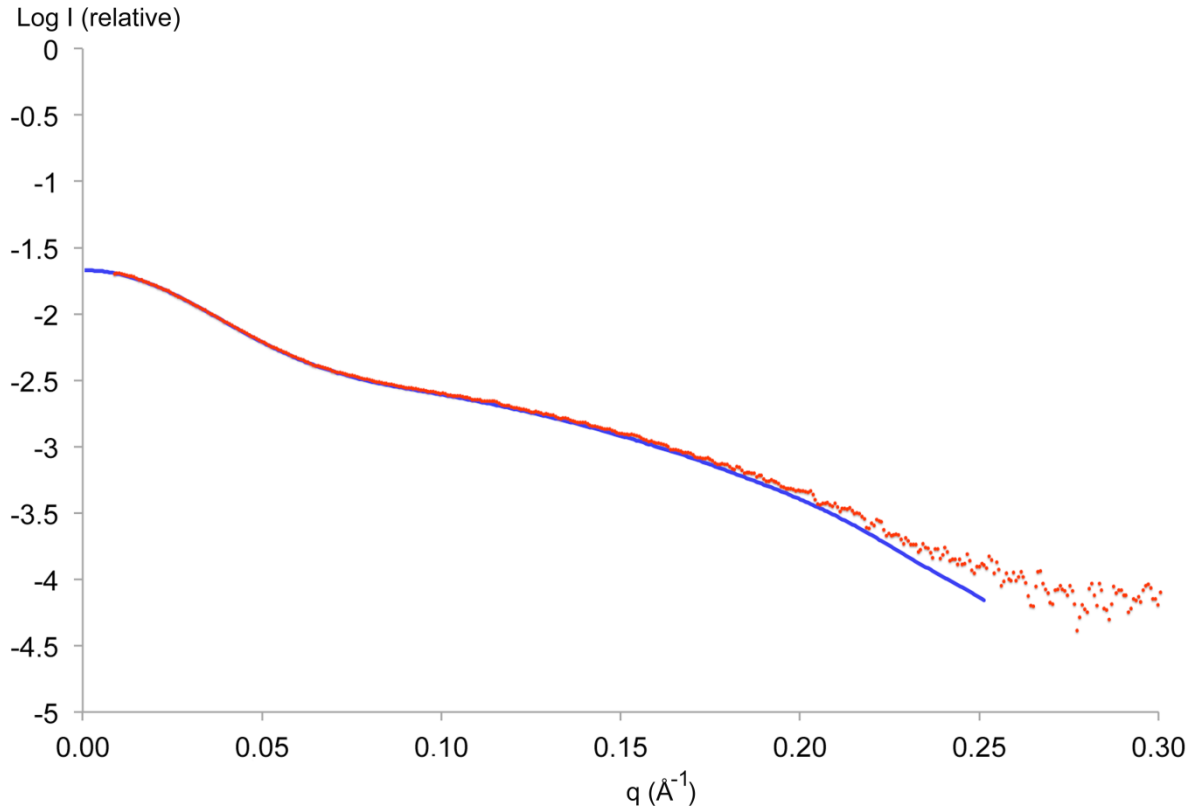


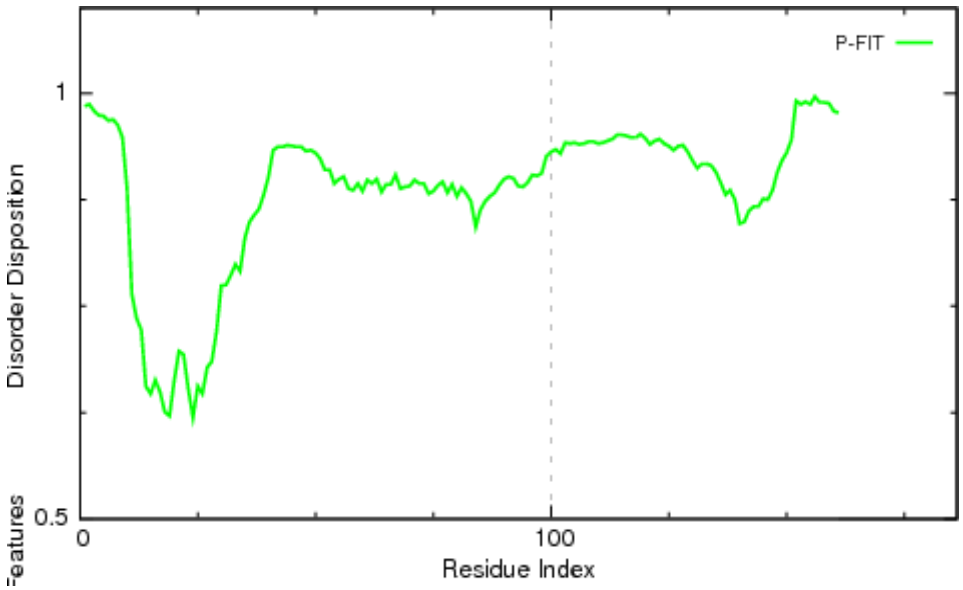
Fig. S5 SAXS analysis of the various constructs. (a) The distance distribution function derived for the ACP_{5a}-ACP_{5b} construct calculated with GNOM. Inset is the Guinier plot. (b) The distance distribution function derived for KS_5-ACP_{5a}-ACP_{5b} calculated with GNOM. Inset is the Guinier plot. (c) The distance distribution function derived for KS_5-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

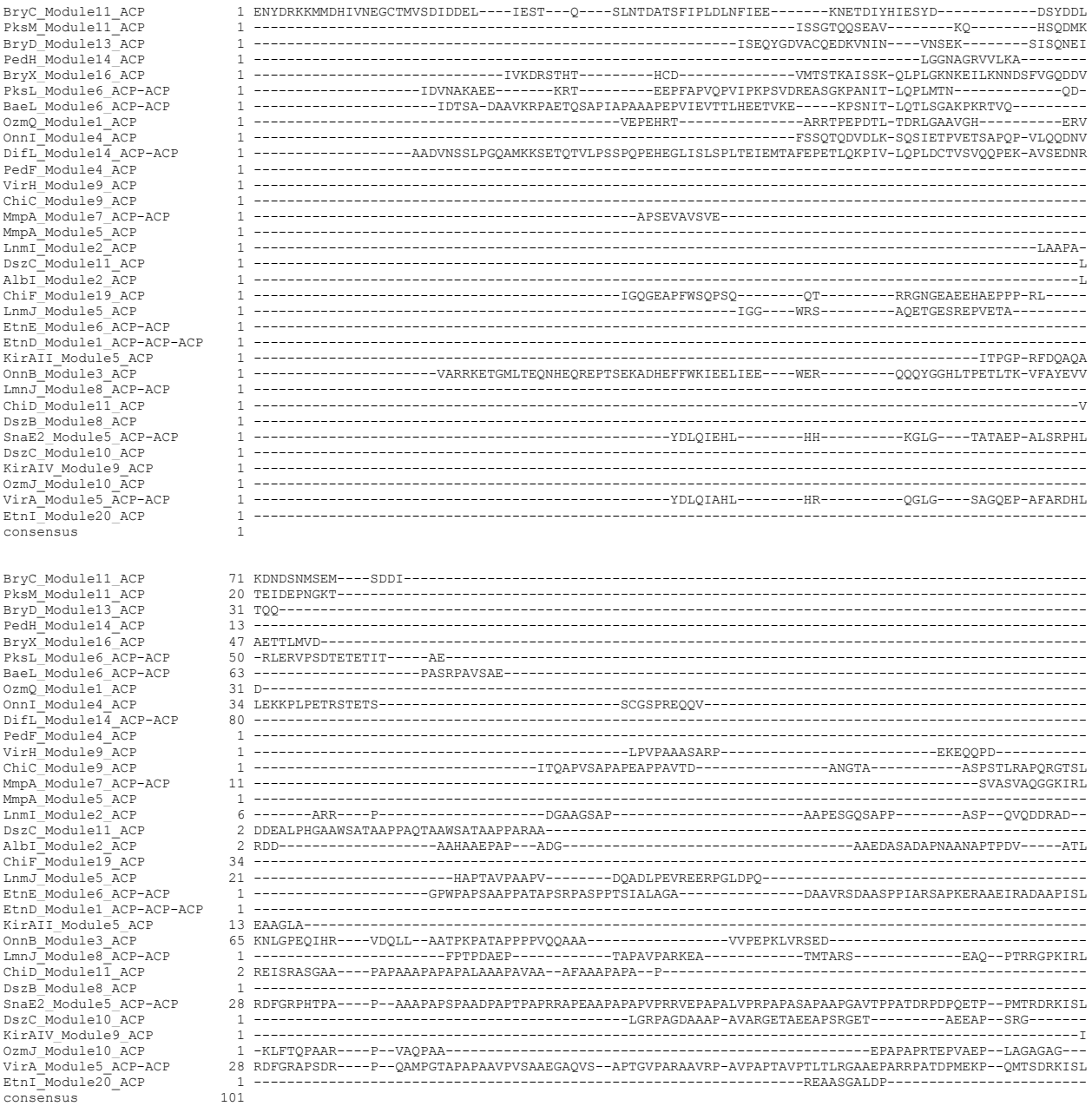
Amino acid bias of intrinsically disordered regions:[4]
Low content: C, V, L, I, M, Y, F and W
High content: Q, S, P, E, K, G and A

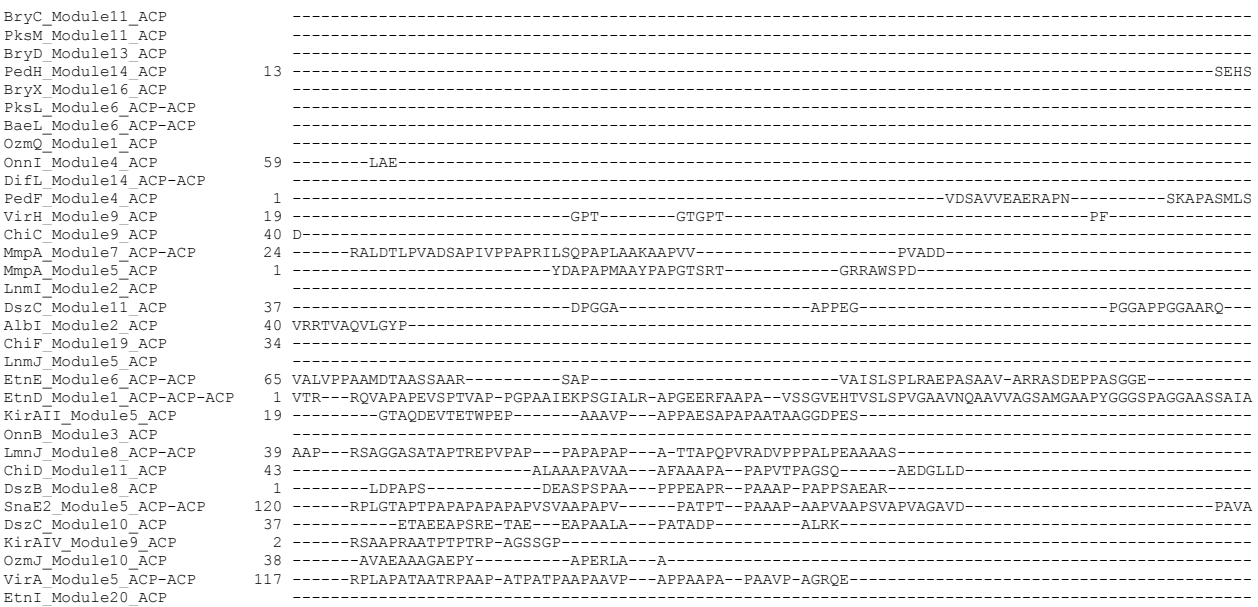
LQIAHLHRPGLGSAGPEPAFARDHLRDFGRAPSDRPPAMPGTAPAPAAVPVSAAEGAPVSAP
TGVPARAAVRPAVPAPTAVPTLTLRGAAEPARRPATDPMEKPPMTSDRKISLRPLAPATAAT
RPAAPATPTPAAPAAVPAPPAAPAPAAVPAGRQE

b



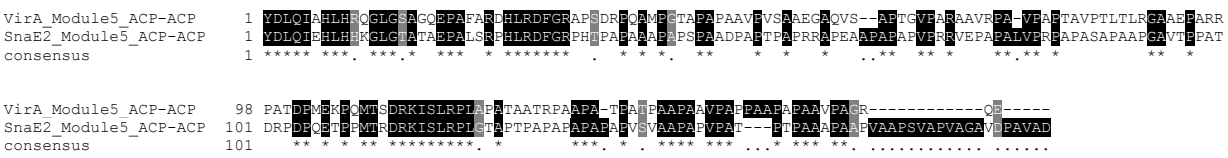
c


BryC_Module11_ACP 1 ENYDRKKMMDHIVNEGCTMVSDIDDEL----IEST---Q----SLNTDATSFIPLDLNFIEE-------KNETDIYHIESYD-----------DSYDDL
PksM_Module11_ACP 1 ----------ISSGTQQSEAV--------KQ--------HSQDMK
BryD_Module13_ACP 1 ----------ISEQYGDVACQEDKVNIN----VNSEK--------SISQNEI
PedH_Module14_ACP 1 ----------LGGNAGRVVLKA--------
BryX_Module16_ACP 1 ----------IVKDRSTHT---------HCD--------------VMTSTKAISSK-QLPLGKNKEILKNNDSFVGQDDV
PksL_Module6_ACP-ACP 1 ----------IDVNAKAEE-------KRT---------EEPFAPVQPVIPKPSVDREASGKPANIT-LQPLMTN------------QD-
BaeL_Module6_ACP-ACP 1 ----------IDTSA-DAAVKRPAETQSAPIAPAAAPEPVIEVTTLHEETVKE-----KPSNIT-LQTLSGAKPKRTVQ---------
OzmQ_Module1_ACP 1 ----------VEPEHRT---------------ARRTPEPDTL-TDRLGAAVGH----------ERV
OnnI_Module4_ACP 1 ----------FSSQTQDVDLK-SQSIETPVETSAPQP-VLQQDNV
DifL_Module14_ACP-ACP 1 ----------AADVNSSLPGQAMKKSETQTVLPSSPQPEHEGLISLSPLTEIEMTAFEPETLQKPIV-LQPLDCTVSVQQPEK-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---------RRGNGEAEEHAEPPP-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---------QQQYGGHLTPETLTK-VFAYEVV
LmnJ_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 KDNDSNMSEM----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 ----------LPVPAAASARP----------EKEQQPD----------
ChiC_Module9_ACP 1 ----------ITQAPVSAPAPEAPPAVTD----------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 DDEALPHGAAWSATAAPPAQTAAWSATAAPPARAA----------
AlbI_Module2_ACP 2 RDD----------AAHAAEPAP---ADG----------AAEDASADAPNAANAPTPDV-----ATL
ChiF_Module19_ACP 34 ----------
LnmJ_Module5_ACP 21 ----------HAPTAVPAAPV--------DQADLPEVREERPGLDPQ----------
EtnE_Module6_ACP-ACP 1 ----------GPWPAPSAAPPATAPSRPASPPTSIALAGA----------DAAVRSDAASPPIARSAPKERAAEIRADAAPISL
EtnD_Module1_ACP-ACP-ACP 1 ----------
KirAII_Module5_ACP 13 EAAGLA----------
OnnB_Module3_ACP 65 KNLGPEQIHR----VDQLL--AATPKPATAPPPPVQQAAA----------VVPEPKLVRSED----------
LmnJ_Module8_ACP-ACP 1 ----------FPTPDAEP----------TAPAVPARKEA----------TMTARS----------EAQ--PTRRGPKIRL
ChiD_Module11_ACP 2 REISRASGAA----PAPAAAPAPAPALAAAPAVAA--AFAAAPAPA--P----------
DszB_Module8_ACP 1 ----------
SnaE2_Module5_ACP-ACP 28 RDFGRPHTPA----P--AAAPAPSPAADPAPTPAPRRAPEAAPAPAPVPRRVEPAPALVPRPAPASAPAAPGAVTPPATDRPDPQETP--PMTRDRKISL
DszC_Module10_ACP 1 ----------LGRPAGDAAAP-AVARGETAEEAPSRGET---------AEEAP--SRG-------
KirAIV_Module9_ACP 1 ----------I
OzmJ_Module10_ACP 1 -KLFTQPAAR----P--VAQPAA----------EPAPAPRTEPVAEP--LAGAGAG---
VirA_Module5_ACP-ACP 28 RDFGRAPSDR----P--QAMPGTAPAPAAVPVSAAEGAQVS--APTGVPARAAVRP-AVPAPTAVPTLTLRGAAEPARRPATDPMEKP--QMTSDRKISL
EtnI_Module20_ACP 1 ----------REAASGALDP----------
consensus 101

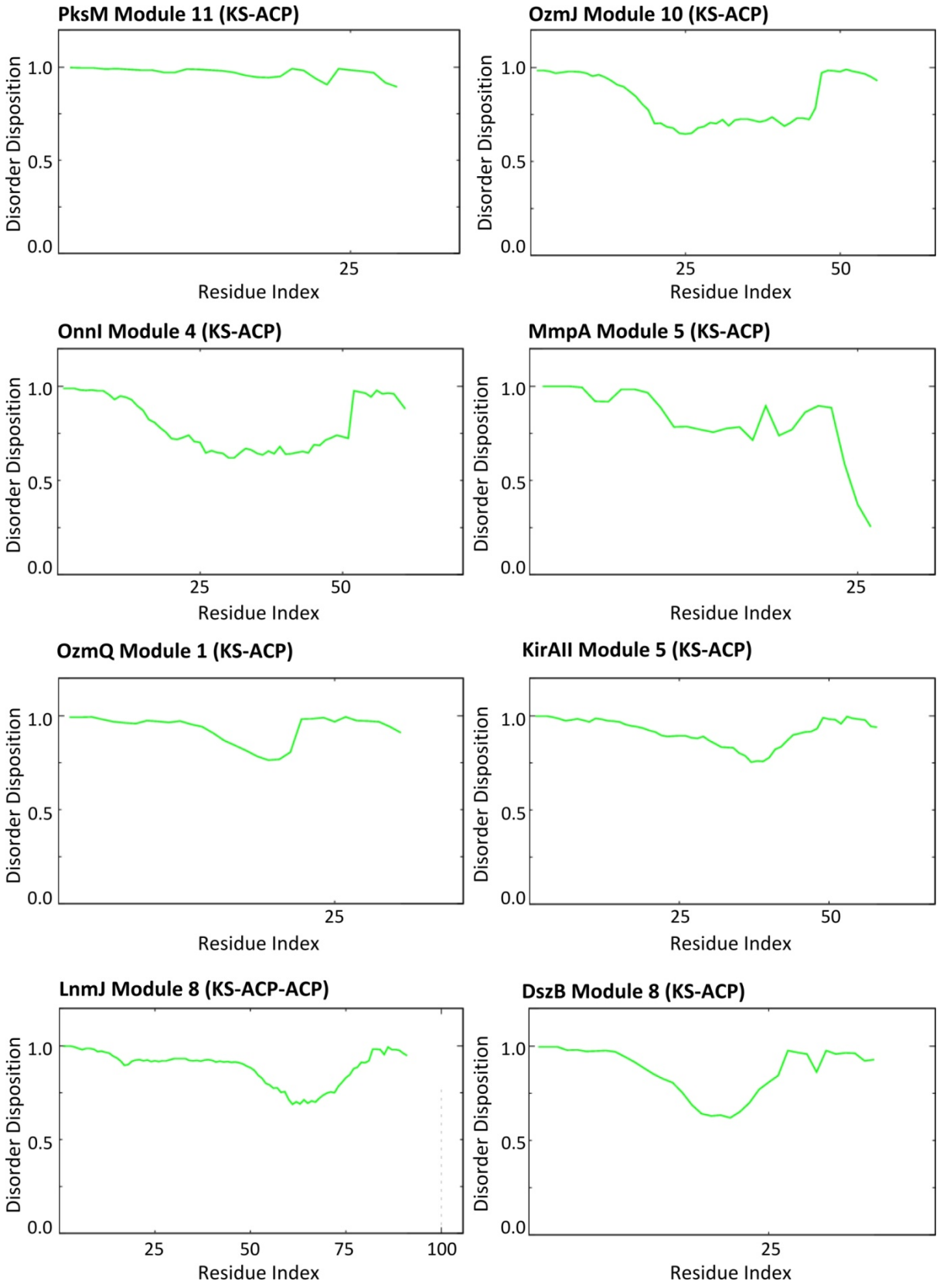

BryC_Module11_ACP
PksM_Module11_ACP
BryD_Module13_ACP
PedH_Module14_ACP 13
BryX_Module16_ACP
PksL_Module6_ACP-ACP
BaeL_Module6_ACP-ACP
OzmQ_Module1_ACP
OnnI_Module4_ACP 59
DifL_Module14_ACP-ACP
PedF_Module4_ACP 1
VirH_Module9_ACP 19
ChiC_Module9_ACP 40
MmpA_Module7_ACP-ACP 24
MmpA_Module5_ACP 1
LnmI_Module2_ACP
DszC_Module11_ACP 37
AlbI_Module2_ACP 40
ChiF_Module19_ACP 34
LnmJ_Module5_ACP
EtnE_Module6_ACP-ACP 65
EtnD_Module1_ACP-ACP-ACP 1
KirAII_Module5_ACP 19
OnnB_Module3_ACP
LmnJ_Module8_ACP-ACP 39
ChiD_Module11_ACP 43
DszB_Module8_ACP 1
SnaE2_Module5_ACP-ACP 120
DszC_Module10_ACP 37
KirAIV_Module9_ACP 2
OzmJ_Module10_ACP 38
VirA_Module5_ACP-ACP 117
EtnI_Module20_ACP
consensus 201
BryC_Module11_ACP
PksM_Module11_ACP
BryD_Module13_ACP
PedH_Module14_ACP 17 ERDAVEPEVERNGSA
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 NGTNGAHAD
LnmJ_Module5_ACP
EtnE_Module6_ACP-ACP
EtnD_Module1_ACP-ACP-ACP 94 GVERVTPRATEGGVGGHMPGTTATARDRRASPVDNAAGGRSLA
KirAII_Module5_ACP
OnnB_Module3_ACP
LmnJ_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
KirAII_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
LnmI_Module 2 (KS-ACP)	39
ChiD_Module 11 (KS-ACP)	57
DszC_Module 11 (KS-ACP)	98
VirH_Module 9 (KS-ACP)	28

d



e


PksM Module 11 (KS-ACP)
OzmJ Module 10 (KS-ACP)
OnnI Module 4 (KS-ACP)
MmpA Module 5 (KS-ACP)
OzmQ Module 1 (KS-ACP)
KirAII Module 5 (KS-ACP)
LnmJ Module 8 (KS-ACP-ACP)
DszB Module 8 (KS-ACP)
Disorder Disposition
Residue Index
0.0
0.5
1.0
25
50
75
100

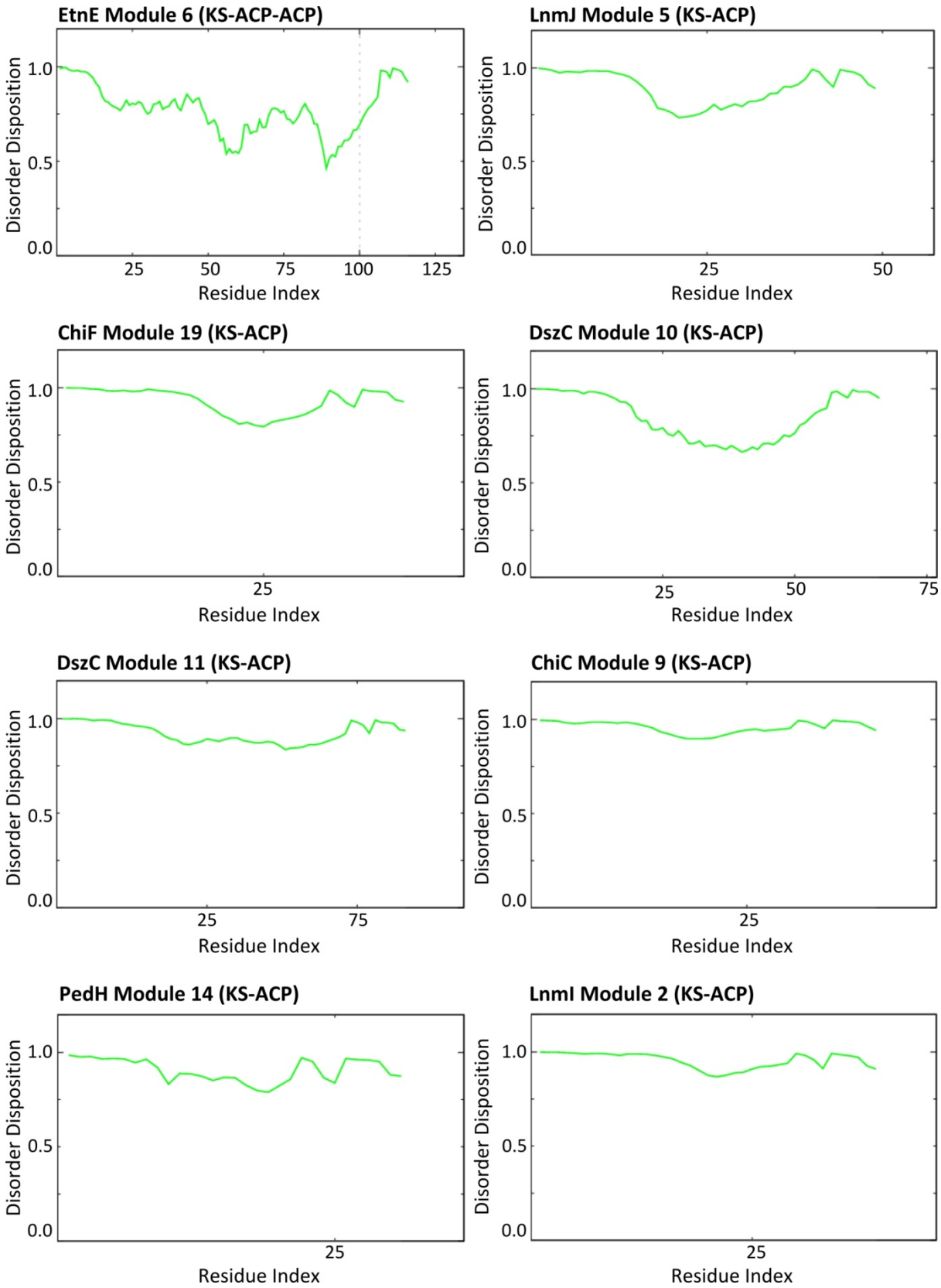
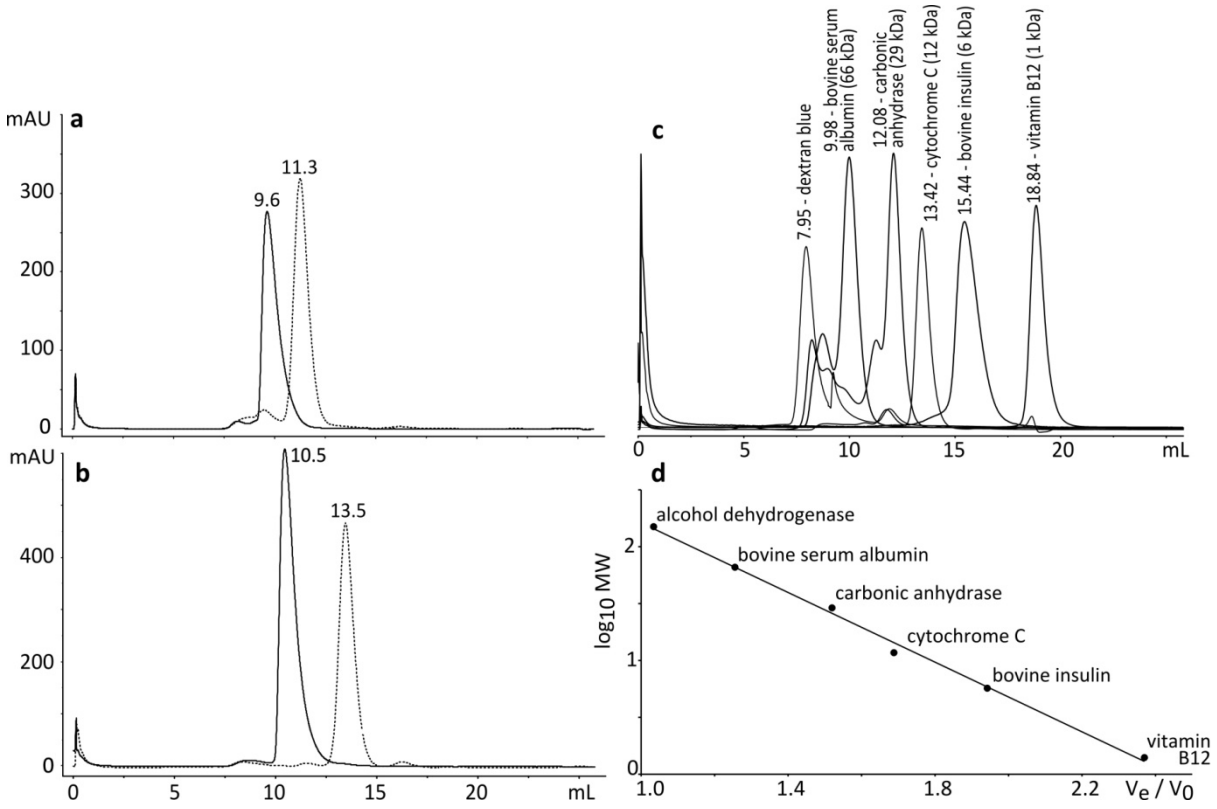
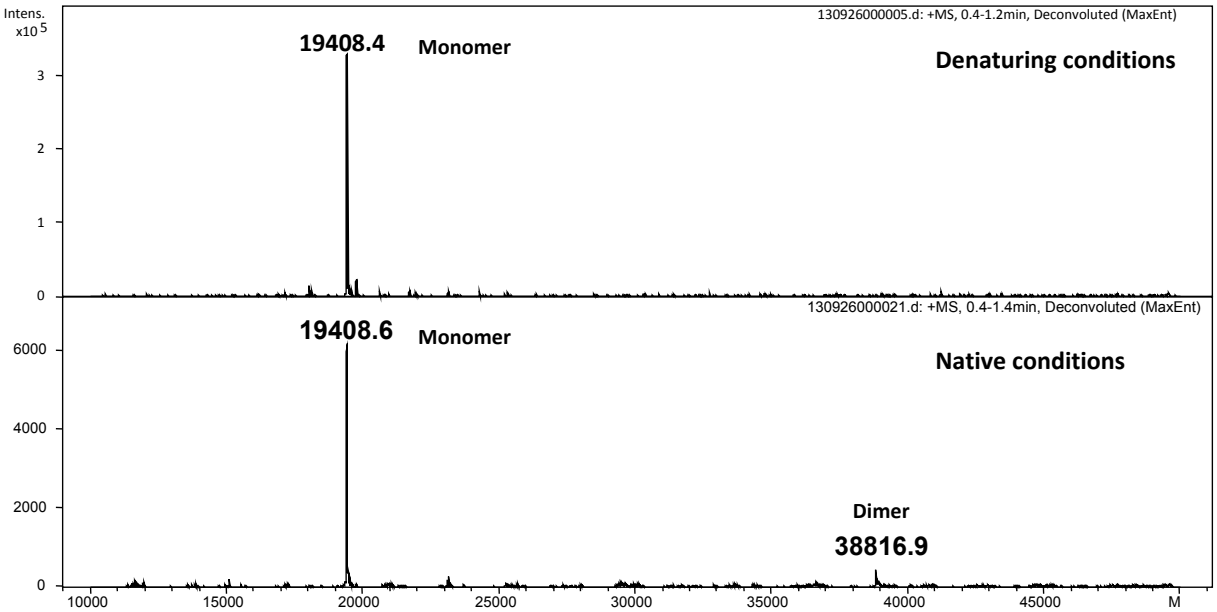


Fig. S6 Sequence analysis of the linker region following the 'post-AT Linker' (referred to as the 'post-post-AT-linker' (**Fig. 3**)) in various *trans*-AT PKS. (a) Analysis of the amino acid composition of the 159 residue post-post-AT linker of VirA module 5 reveals a sequence bias consistent with an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

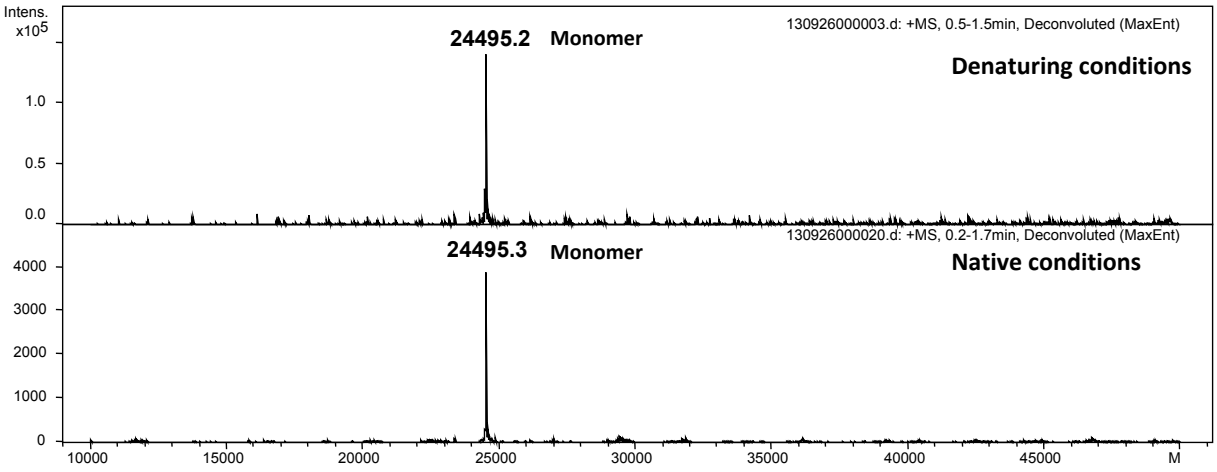
from a selection of KS-$(ACP)_n$ modules from *trans*-AT PKS. The subunit name, module number and its ACP composition are indicated. The linker lengths are given in the table, with the two extremes highlighted in pink. (d) Comparison of the linker regions following the post-AT linker in virginiamycin 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.


a
mAU
11.3
9.6
b
10.5
13.5
c
7.95 - dextran blue
9.98 - bovine serum albumin (66 kDa)
12.08 - carbonic anhydrase (29 kDa)
13.42 - cytochrome C (12 kDa)
15.44 - bovine insulin (6 kDa)
18.84 - vitamin B12 (1 kDa)
mL
d
alcohol dehydrogenase
bovine serum albumin
carbonic anhydrase
cytochrome C
bovine insulin
vitamin B12
log10 MW
Ve / V0

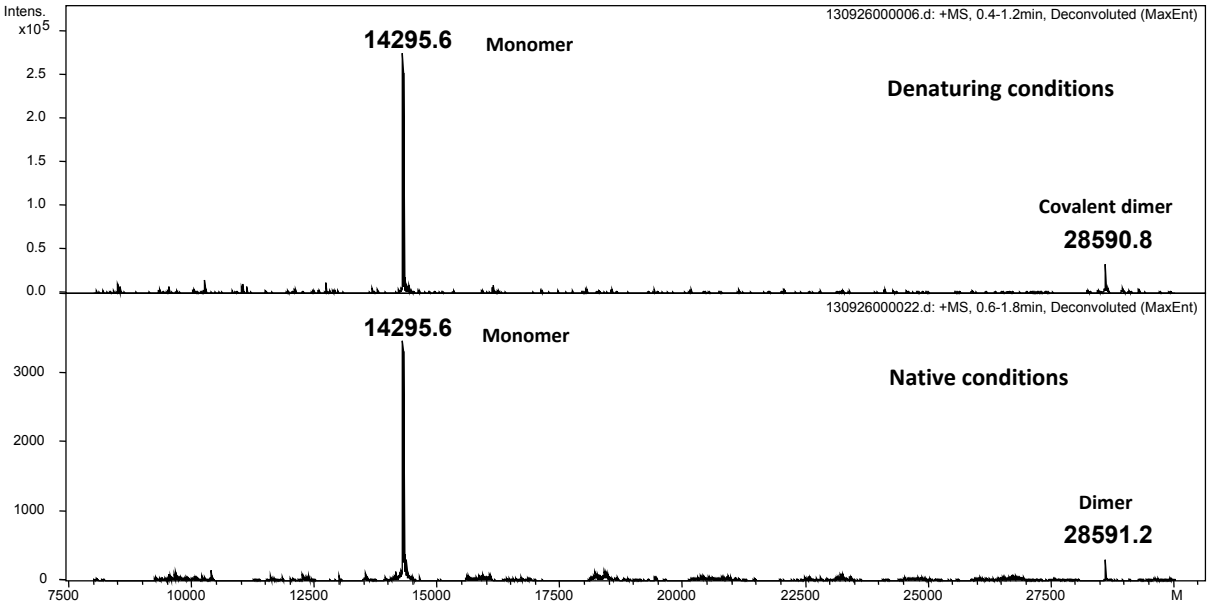
e


Intens.
x10 5
130926000005.d: +MS, 0.4-1.2min, Deconvoluted (MaxEnt)
19408.4
Monomer
Denaturing conditions
130926000021.d: +MS, 0.4-1.4min, Deconvoluted (MaxEnt)
19408.6
Monomer
Native conditions
Dimer
38816.9
M

f


Intens.
x10^5
130926000003.d: +MS, 0.5-1.5min, Deconvoluted (MaxEnt)
24495.2 Monomer
Denaturing conditions
1.0
0.5
0.0
130926000020.d: +MS, 0.2-1.7min, Deconvoluted (MaxEnt)
24495.3 Monomer
Native conditions
4000
3000
2000
1000
0
10000
15000
20000
25000
30000
35000
40000
45000
M

g


Intens.
x10^5
130926000006.d: +MS, 0.4-1.2min, Deconvoluted (MaxEnt)
14295.6 Monomer
Denaturing conditions
Covalent dimer
28590.8
2.5
2.0
1.5
1.0
0.5
0.0
130926000022.d: +MS, 0.6-1.8min, Deconvoluted (MaxEnt)
14295.6 Monomer
Native conditions
Dimer
28591.2
3000
2000
1000
0
7500
10000
12500
15000
17500
20000
22500
25000
27500
M

h

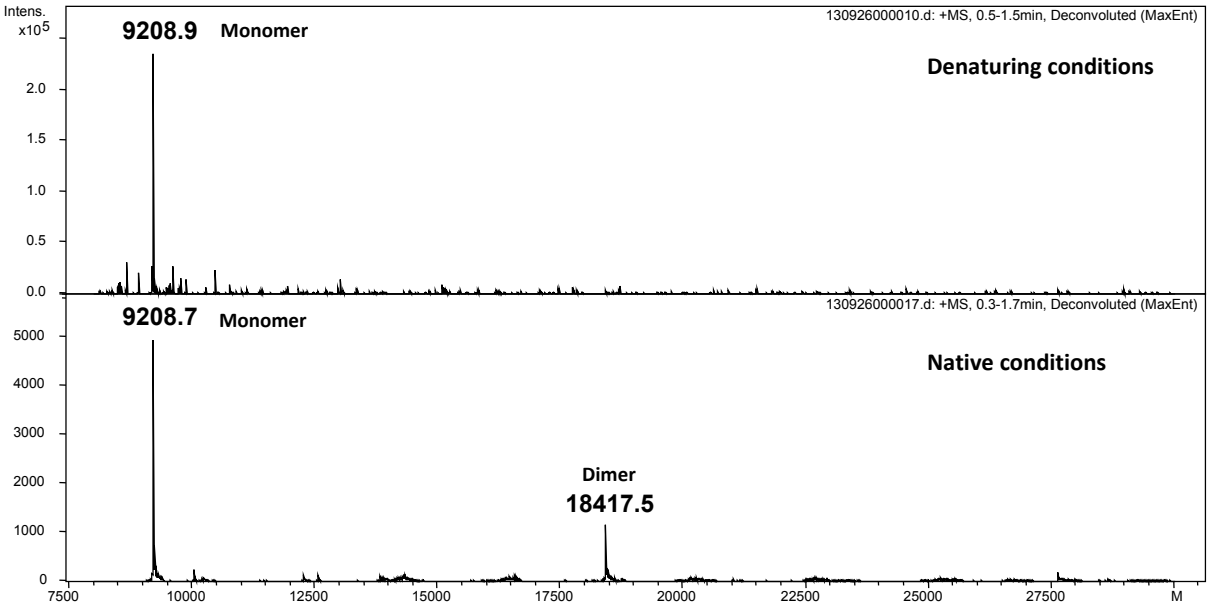


Fig. S7 Analysis of the effect on the oligomerization state of various constructs of the presence of the putative dimerization domain (DD). FPLC traces of (a) ACP_{5a}-ACP_{5b}-DD (solid line) and ACP_{5a}-ACP_{5b} (dotted line), (b) ACP_{5b}-DD (solid line) and ACP_{5b} (dotted line), and (c) the calibration 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

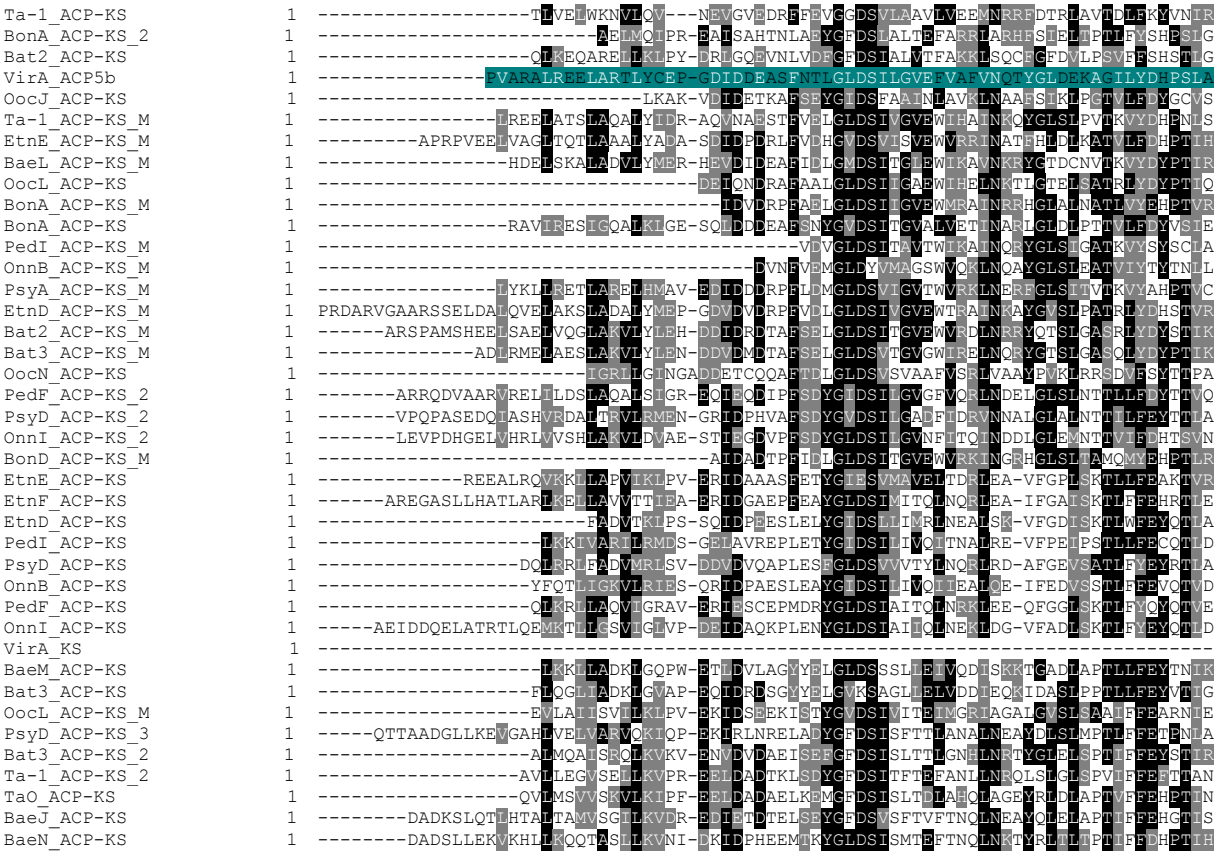
The presence of the dimerization domain (DD) on ACP_{5b} gives a molecular weight approximately four-fold greater than the ACP_{5b} alone, while that of the ACP_{5a}-ACP_{5b} is more than doubled by addition of the DD. It could be argued that the addition of the DD makes the proteins substantially more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

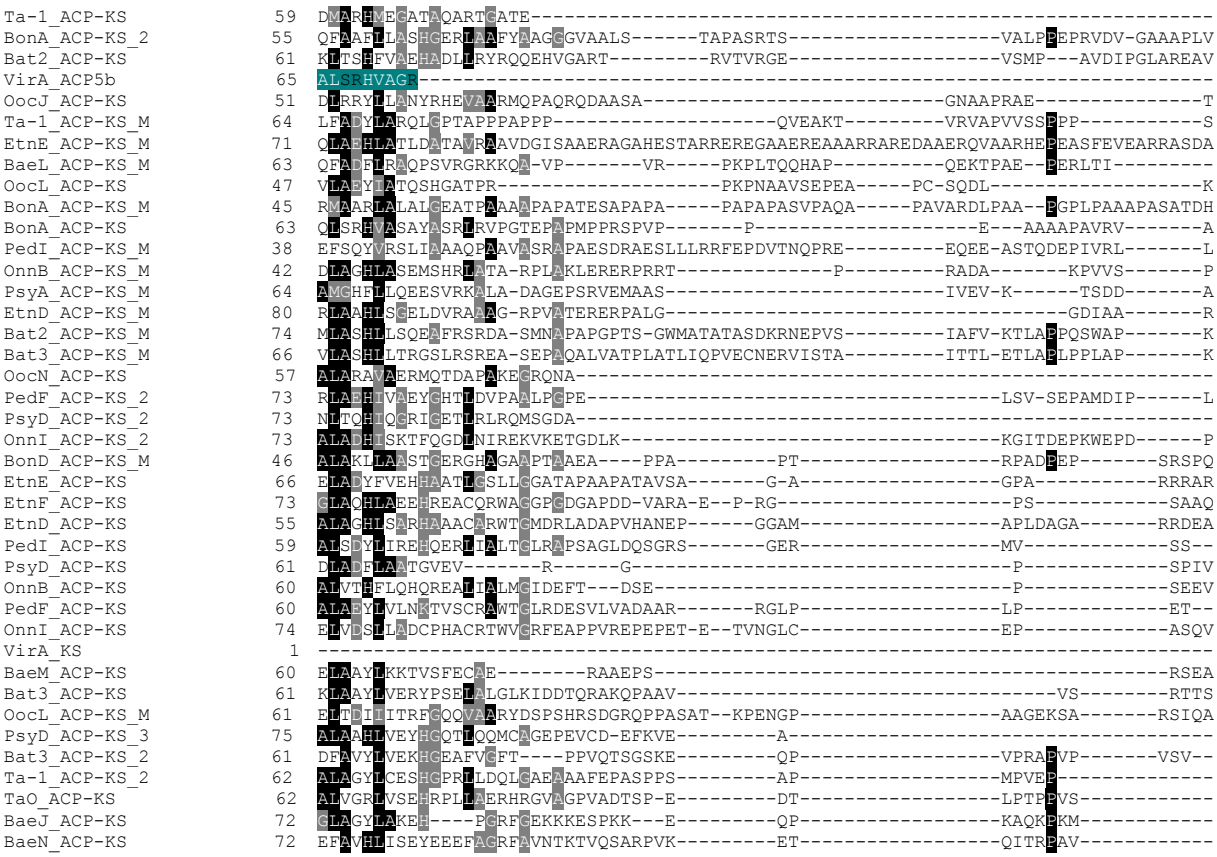
globular protein of the same molecular weight.[6] However, the fact that the ACPs even in the absence of the DD give estimated molecular weights in great excess of their calculated weights, shows that the constructs are already in an extended conformation. Thus, our analysis instead supports a change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

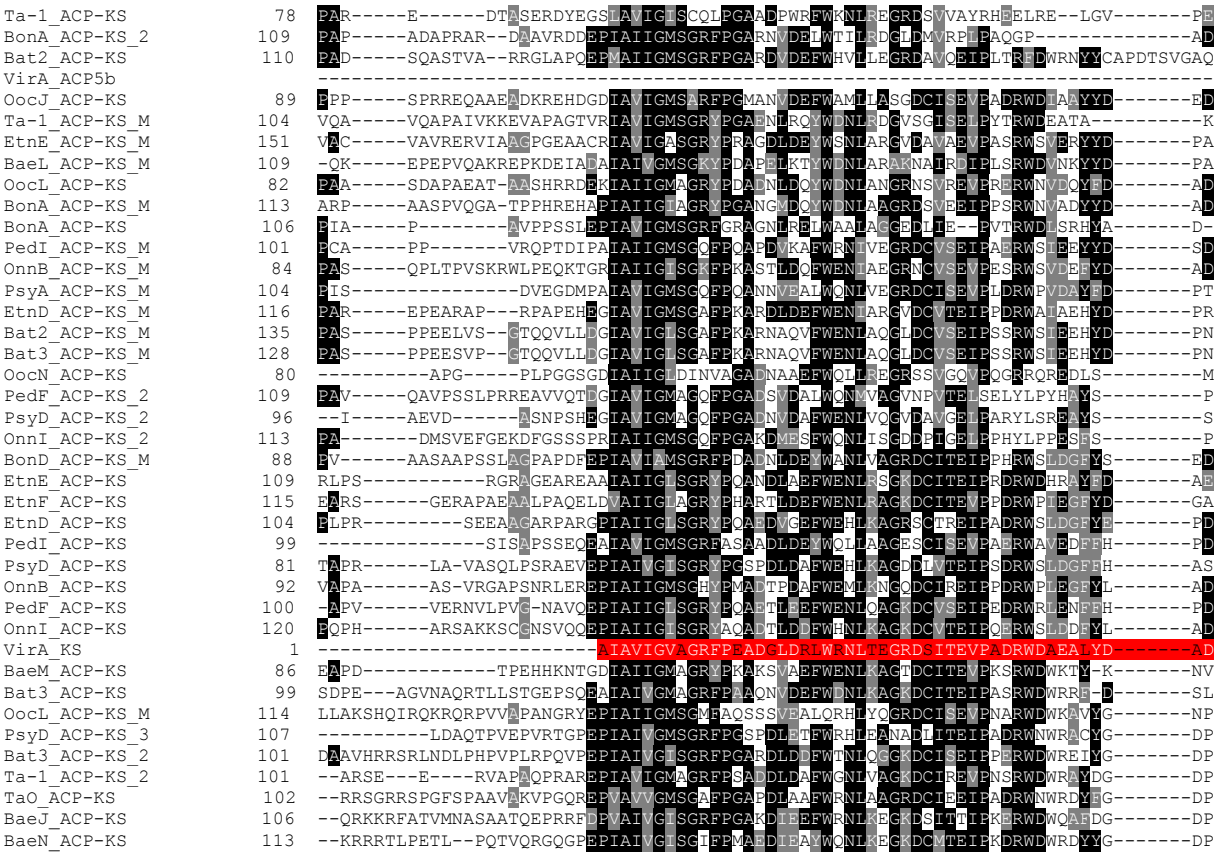
MDAKEILTRFKDGGLDRAAAQALLAGRT**PAAAPRPSEPAAAPTRPAVPAVPAVEPAAGTTVAAEGTAGRP**EPVA
VIGYSARFPGAADADTFWQRILDGDDLVTEVPPERWRTEEFYDPDPAAEGRSVSRWGAYVADADRFDADFFRMTP
REAELTDPQARLFLQEAWRALEHAGRDARSLAGTRCGVYAGVMLNDYQDLVERESPYKRLPQVMQGNSNSILAA
RIAYHLDLKGPAVTVDTACSSSLTALHLACQSLWLGETDLTVVGGVTLYLTELPHVFMSAAGMLSPNGRCRPFDAA
ADGIVPGEGCAVVVLKPLSKALADGDPVHAVIRASGLNQDGRTNGITAPSARSQTALVRDTLQRFAVDPAGIDYVE
CHGTGTPLGDPIEVTALNEAFAGAGLAPASVPIGSVKGNIGHTSAAAGLAGLLKAAGVVRTGLVPPSLHYARANPQI
PFDQGPFTVAGERRELGRPDGGRPRRATVSSFGFSGTNAYVVVEQAPEQAARPAGDDGAPPVLVPLSGRRADAPAA
HAADLARWLRGPGSDASPADVAHTLAVARTHHTYRFALLVSGRDELLESLDLLAAGSADPRRTDTAPDAAAPDAA
VRRAQAALLARLVERAGENPGPVELAALAKLYTQGHTVDWAAVSPPARHRRLSLPAYPFAPHRHYVERPAPVTPP
TPAATSALAEASHGTPAEPAGARTPQEALADVPLTEPVLHRQEWIPAPLPAAAPAPTAPVLVHDPAGDLAAALAAT
GLDVHRLGGDRTAAEALRATGAQVVTVVLRLTPPADGAPATAVDLAAFEAARAALAVLRTQQLTLLAVTADPTR
ADAAGALVQTLRQENPRLSGRAVLTADGTPDARRLLAEIAAPAAEHGHLADLRSARRLRRVLVPVPLPGTRPFVRQ
DGVYLVSGGAGGIGLALARHLTDRPGTRVVLCGRTAWDDLAEATRSAVSGSDRLRYARADVTDPEAAAALVAEV
VRTEGALHGVFHAAGVVRDGYLVRKDPRDVADVLAPKILGARALDEATAALPLDAFVLFSSVAAVTGNLGQSDY
AFANGFLDGFATRRAGQVARGERHGRTLSVQWPLWDVPGMSIPEPVLEVVAQHTGMAPLPAAVGLAALERLLAA
DGPEVVSLFHGDAATWRAHLAALHLERPAAAGQVTAAAPAAPAPSDTPVVASADPAAADDGRRAELADRVSRTV
ADTIGRPAGSIGGHTSLESMGLDSVMIRALASRLSAEVAPVGPEMLFGLRDLDELVDHLVAARPVPVAEPATPAVPP
APAATAATVFAPVAPADPAVVPAAPAALPVPAAAPASASRTLPAPSADDRFAIIGISGRYPQAPDLDAFWQNLLNGK
DTASDLPTDRWPDAAGVNARGHFLEGVDAFDPTFFGLSAHDGTLLDPQERLFLEVAWEALEDAGYTGSRLHDLVA
PDGERRSVGVFAGITSSDYKLLGAERWAAGHREMPSGHYWSLPNRLSYLLDLRGPSQPVDTACSSSLVALHLALDA
LRRGECAAALVGGVNLYVHPSRFRMLRRSGFLAEDGLCRSFGAGGAGFGPGEGGGAVVVKPLAVALADGDTVHA
VVRGSAVAHGGRTNGFTAPSPWAQARVLRAALRNSGTDPETVNVIEAHGTGTELGDPVELAGLQDAYGSGRVPCS
LGSVKSAVGHGESAAGIAALTKVVLQLRHGELVPTLHADPVNPGLRLEDTRFVLQHTPGRWERLTGADGSPLPRRA
GISSFGAGGVNAHVIVEEYLPEPHGRAAAAEPGRPELVLLSAPTREHLAATADRLARRLEGPEAPADLRAVAYSLRT
GRSAMDCRLAVVATDTAGLAAALGAFARSADAGEGESAVRYADLRDGRRAHRDLDTVPETTAFLADLWRNRRLH
QLGELWLSGLDIERAAPREPGTRLVPLPTSAFLRRRLWITDPVTADTTPTPDPLPTPDPLPVPLSAPAPAPAPAPQPAP
TTTPAPAAAPAAAPARSGEAVTEQLSRLVAGFVPDAAGSVDPDRTLLEHGIDSINLMNLRFEITERFGRTLPLQLLSE
STVPVLAAHLSADRAHDRA*

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

a


Ta-1_ACP-KS
BonA_ACP-KS_2
Bat2_ACP-KS
VirA_ACP5b
OocJ_ACP-KS
Ta-1_ACP-KS_M
EtnE_ACP-KS_M
BaeL_ACP-KS_M
OocL_ACP-KS
BonA_ACP-KS_M
BonA_ACP-KS
PedI_ACP-KS_M
OnnB_ACP-KS_M
PsyA_ACP-KS_M
EtnD_ACP-KS_M
Bat2_ACP-KS_M
Bat3_ACP-KS_M
OocN_ACP-KS
PedF_ACP-KS_2
PsyD_ACP-KS_2
OnnI_ACP-KS_2
BonD_ACP-KS_M
EtnE_ACP-KS
EtnF_ACP-KS
EtnD_ACP-KS
PedI_ACP-KS
PsyD_ACP-KS
OnnB_ACP-KS
PedF_ACP-KS
OnnI_ACP-KS
VirA_KS
BaeM_ACP-KS
Bat3_ACP-KS
OocL_ACP-KS_M
PsyD_ACP-KS_3
Bat3_ACP-KS_2
Ta-1_ACP-KS_2
TaO_ACP-KS
BaeJ_ACP-KS
BaeN_ACP-KS


Ta-1_ACP-KS
BonA_ACP-KS_2
Bat2_ACP-KS
VirA_ACP5b
OocJ_ACP-KS
Ta-1_ACP-KS_M
EtnE_ACP-KS_M
BaeL_ACP-KS_M
OocL_ACP-KS
BonA_ACP-KS_M
BonA_ACP-KS
PedI_ACP-KS_M
OnnB_ACP-KS_M
PsyA_ACP-KS_M
EtnD_ACP-KS_M
Bat2_ACP-KS_M
Bat3_ACP-KS_M
OocN_ACP-KS
PedF_ACP-KS_2
PsyD_ACP-KS_2
OnnI_ACP-KS_2
BonD_ACP-KS_M
EtnE_ACP-KS
EtnF_ACP-KS
EtnD_ACP-KS
PedI_ACP-KS
PsyD_ACP-KS
OnnB_ACP-KS
PedF_ACP-KS
OnnI_ACP-KS
VirA_KS
BaeM_ACP-KS
Bat3_ACP-KS
OocL_ACP-KS_M
PsyD_ACP-KS_3
Bat3_ACP-KS_2
Ta-1_ACP-KS_2
TaO_ACP-KS
BaeJ_ACP-KS
BaeN_ACP-KS



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

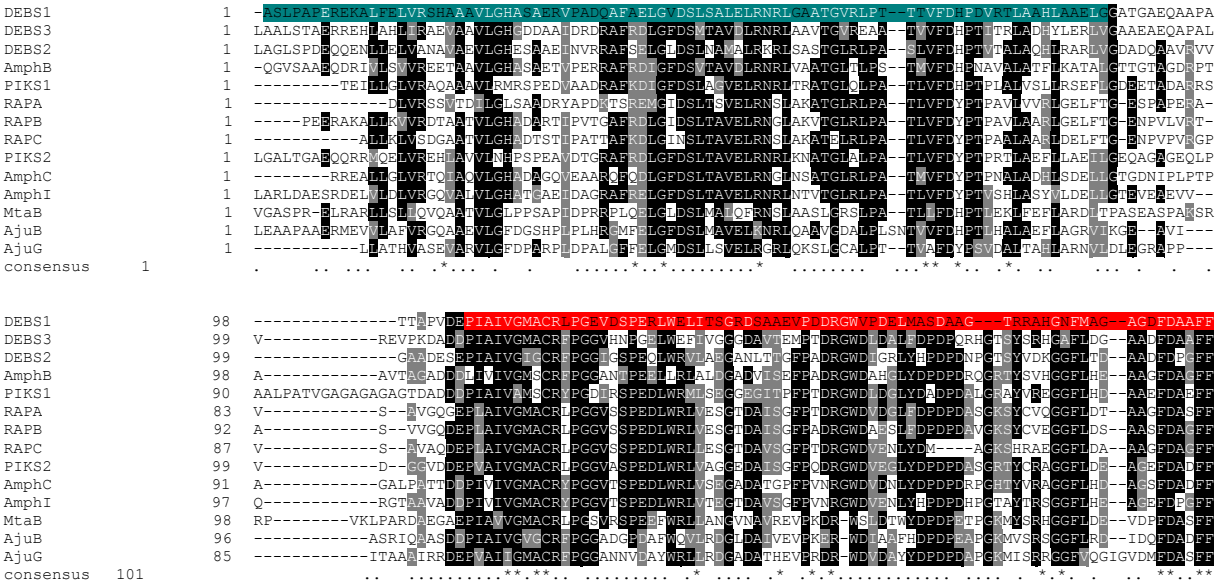


Fig. S9 Analysis of ACP-KS linkers. (a) Portion of a multiple alignment of the ACP−KS regions (junction between two modules) from a selection of *trans*-AT PKS subunits, to determine the lengths (shown in the table) of the linkers intervening between the conserved ACP and KS domains. The C-terminal and 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, kalimanticin; 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 ACP_2.[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 approximation[8]
[b] D_{max} is the maximal particle diameter derived from the distance distribution function ($P(r)$) using the program GNOM[9]
[c] χ^2 is the discrepancy between the experimental SAXS profile and the theoretical calculated curve using the programs DAMMIN,[9] SASREF[10] or MONSA[10]
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 (º)	0 ± 0
Max. dihedral angle violation (º)	0
Max. distance constraint violation (Å)	0.13
Deviations from idealized geometry	
Bond lengths (Å)	0.010 ± 0.001
Bond angles (º)	2.062 ± 0.018
Impropers (º)	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 (º)	0 ± 0
Max. dihedral angle violation (º)	0
Max. distance constraint violation (Å)	0.17
Deviations from idealized geometry	
Bond lengths (Å)	0.010 ± 0.001
Bond angles (º)	2.054 ± 0.018
Impropers (º)	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.