

Chemical Science

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



www.rsc.org/chemicalscience

ARTICLE

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

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,

Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/Jack Davison,^a Jonathan Dorival,^a Hery Rabeharindranto,^{a,b} Hortense Mazon,^a Benjamin Chagot,^{*a} Arnaud Gruez^{*a} and Kira J Weissman^{*a}

The modular polyketide synthases (PKS) are gigantic multienzymes which synthesize diverse secondary metabolites of therapeutic value. Although structural information is increasingly available for the '*cis*-AT' class of modular PKS, almost nothing is known about the evolutionarily divergent '*trans*-AT' PKS, which characteristically incorporate an iteratively-acting acyl transferase. We report here the SAXS solution structure of a complete *apo* module from the virginiamycin *trans*-AT PKS, which is fundamentally different to structural models proposed for the *cis*-AT PKS based on the crystal structure of animal fatty acid synthase. The module, which serves as a platform for β -modification of the polyketide intermediate, consists of a ketosynthase (KS) and two acyl carrier protein (ACP) domains. In our solved structure, the homodimeric KS, which is flanked by well-folded linker regions, occupies the center of the module. While the first ACP is located close to the KS, the second is situated at the end of a flexible linker, and mobile. Taken together, these data provide a physical explanation for the functional non-equivalence previously observed for certain tandem ACPs of *trans*-AT PKS. Furthermore, the overall open shape of the module renders the second ACP highly accessible, which may be critical for its interaction with multiple *in trans* catalytic partners. Finally, our analysis redefines the function of a putative dimerization motif of tandem ACPs as a docking domain, suggesting that the module likely adopts a more closed form in order to affect transfer of the chain extension intermediate to the subsequent module.

Introduction

A multitude of bioactive polyketide natural products are biosynthesized using a division-of-labor strategy, in which the discrete steps in assembling the chains are carried out by individually-folding protein domains.¹ The domains are organized into functional units called 'modules' within large multienzyme polyketide synthases (PKS), where each module carries out one chain extension cycle. Thus, each module contains an acyl transferase (AT) domain to choose the appropriate building block and a ketosynthase (KS) domain to attach it to the growing chain using Claisen-type chemistry. A set of optional domains (ketoreductase (KR), dehydratase (DH) and enoyl reductase (ER)) then introduces the appropriate level of reduction at the initially-formed β -keto center. Finally, a dedicated thioesterase (TE) at the end of the last PKS multienzyme liberates the fully-processed chain either by macrocyclization or hydrolysis. The efficacy of this assembly-line strategy is augmented by the covalent tethering of the chain-extension intermediates to non-catalytic acyl carrier protein (ACP) domains, via a phosphopantetheine (Ppant) prosthetic group.²

Presumably in reflection of the efficiency of this mode of biosynthesis for complex secondary metabolites, Nature has convergently evolved two distinct classes of modular PKS. In the well-studied *cis*-AT systems, such as the prototypical PKS (DEBS) responsible for erythromycin assembly (Fig. 1a),³ the ATs are present as domains within the primary sequence of each module. For

such PKS, there is a strong correlation between the domain composition of the modules and the structure of the resulting chain extension intermediates. In contrast, in the more-recently discovered *trans*-AT PKS, the ATs are discrete enzymes which act iteratively to furnish a single extender unit to each module (Fig. 1b).⁴ As a result of further organizational features of the *trans*-AT pathways – the presence of duplicated and inactive domains, modules divided between two proteins, participation of large numbers of additional *trans*-acting enzymes, etc. – there is no strict co-linearity between the primary sequence of the PKS and the product structures.⁵ Another notable feature of these systems, which is shared by certain *cis*-AT PKS, is the presence of nonribosomal peptide synthetase (NRPS) modules allowing the incorporation and modification of amino acid building blocks (Fig. 1b). Despite the divergent evolutionary origins of the *trans*-AT PKS, their products exhibit remarkable structural similarity to those of the *cis*-AT PKS, and are likewise highly bioactive. Notable metabolites produced by *trans*-AT PKS include the anti-cancer bryostatins⁶ and the clinically-used antibiotics mupirocin⁷ and virginiamycin M.⁸

There is significant interest in the three-dimensional architectures of both *cis*- and *trans*-AT PKS, as such information is not only essential to understanding their complex mode of operation, but is vital to ongoing attempts to genetically re-engineer the systems to produce analogues of potential medicinal value.⁹ Resolution of the crystal or NMR structure of each of the domains of

ARTICLE

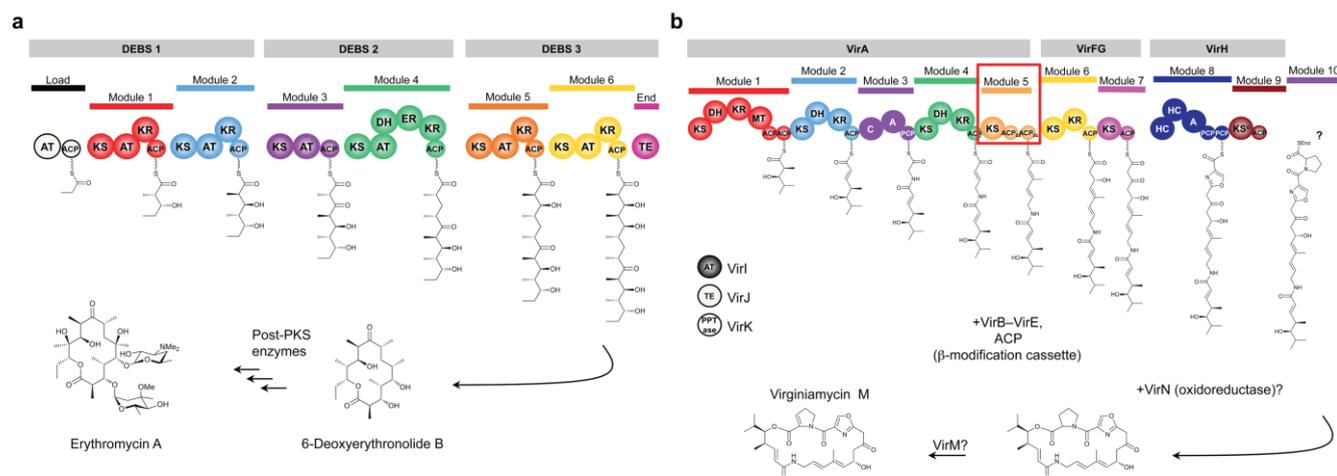


Fig. 1 Schema of the biosynthesis by representative *cis*-AT and *trans*-AT PKS systems. (a) The *cis*-AT PKS which assembles erythromycin A consists of three gigantic multienzymes (DEBS 1, DEBS 2 and DEBS 3) comprising loading, chain extension and termination modules. For this system, there is co-linearity between the order and domain composition of the modules and the structure of the resulting chain-extension intermediate. (b) The *trans*-AT PKS responsible for virginiamycin biosynthesis includes at least nine chain extension modules (seven polyketide synthase and two non-ribosomal peptide synthetase); the presumed tenth module for incorporation of proline was not identified in the sequenced region.⁸ In the course of this work, we determined that modules 6 and 7 are located on a single protein, renamed VirFG (for details see the Methods). This system exhibits several typical characteristics of the *trans*-AT PKS, including a discrete, iteratively acting AT (VirI), duplicated domains (ACPs of modules 1 and 5, PCPs of module 8), a condensation-inactive module (module 9), etc. In this study, we targeted module 5 (boxed) for structure elucidation by SAXS. Key: AT, acyltransferase; ACP, acyl carrier protein; KS, ketosynthase; KR, ketoreductase; DH, dehydratase; ER, enoyl reductase; TE, thioesterase; MT, methyltransferase; C, condensation; A, adenylation; PCP, peptidyl carrier protein; HC, condensation/heterocyclization; KS^o, inactive ketosynthase.

cis-AT PKS¹⁰ has revealed that the proteins have similar folds to the corresponding domains of animal fatty acid synthase (FAS), although there are notable differences in oligomerization state (for example, the ER of FAS is dimeric, while the PKS ER is monomeric¹¹). Taken together, these data have been used to propose several models for the overall structures of *cis*-AT PKS modules incorporating various complements of reductive activities,¹¹ by extrapolation from the 3.2 Å resolution crystal structure of porcine FAS.¹² No such model exists for *trans*-AT PKS, however, as only the structures of a *trans*-AT domain,¹³ an ACP-ACP didomain,¹⁴ and a KS-branching (B) didomain¹⁵ have been solved to date. We report here the first solution structure of a complete *apo* module from a model *trans*-AT system comprising KS and tandem ACP domains, obtained by small-angle X-ray scattering (SAXS)¹⁶ in combination with NMR and homology modeling.

Results

Strategy for resolution of the structure of a PKS module by SAXS

No three-dimensional crystallographic structure of an intact PKS module or multi-modular subunit has been published to date, presumably due to the high, inherent flexibility of the proteins.¹² We thus aimed to use SAXS to solve the structure of a representative module from a *trans*-AT PKS. SAXS has recently come to the fore as a powerful means to analyze large, flexible multidomain proteins and macromolecular complexes in solution due to advances in data

collection and analysis, which allow molecular shapes to be identified accurately.¹⁶

As a representative model system, we chose the ca. 100 kDa module 5 of the virginiamycin NRPS-PKS of *Streptomyces virginiae* (Fig. 1b), which comprises a KS, two consecutive ACP domains (ACP_{5a} and ACP_{5b}), and a putative C-terminal dimerization motif (DD) (see ESI, Fig. S1†),⁸ which are separated from each other by linker regions of variable length (Fig. 2). Module 5 interacts with a large number of catalytic partners both in *cis* and in *trans*: the ACP of module 4, the KS of module 6, the *trans*-AT VirI, the phosphopantetheinyl transferase VirK, the proof-reading thioesterase VirJ, and a cassette of enzymes (VirC–E) that introduces a β-methyl group into the chain-extension intermediate. Modules consisting of a KS and two tandem ACPs are present in fully a third of all *trans*-AT PKS characterized to date,^{4,17–22} and such duplicated ACPs are notably common in modules of *trans*-AT PKS in which the intermediate undergoes a β-methylation.²³ Our results are also likely to be relevant to simpler KS-ACP modules (see for example module 7 of VirFG (Fig. 1b)).⁴

In order to reconstitute the structure of the entire module from SAXS data,¹⁶ we used a ‘dissect and build’ strategy, using SAXS to study first KS₅ and a portion of the downstream linker (KS₅-linker), then the tandem ACPs (ACP_{5a}-ACP_{5b}), the tridomain KS₅-ACP_{5a}-ACP_{5b}, and finally the complete module (KS₅-ACP_{5a}-ACP_{5b}-DD) (Fig. 2). Each of the proteins was expressed in its *apo* form in *E. coli* and purified to homogeneity (Fig. S2†). Attempts to obtain KS₅-ACP_{5a} were unsuccessful, due to poor solubility (data not shown).

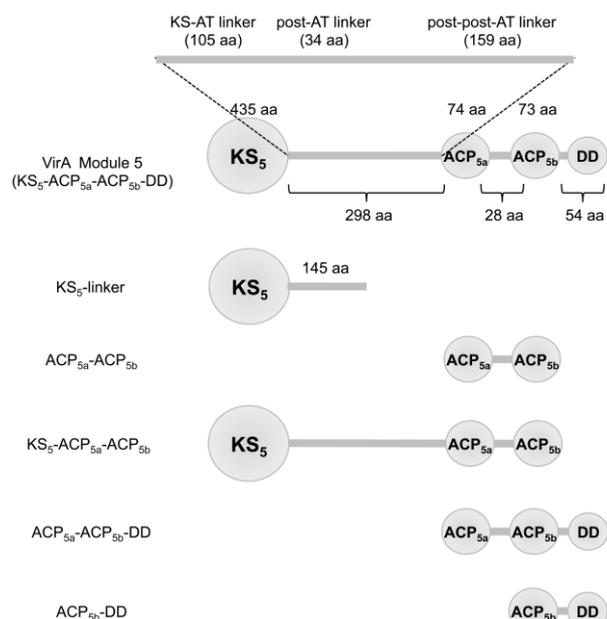


Fig. 2 Schematic of the constructs studied in this work. Six different constructs were evaluated by SAXS, including the entirety of module 5 (KS₅-ACP_{5a}-ACP_{5b}-DD), the KS and a portion of the downstream linker (KS₅-linker), the tandem of ACP domains and their intervening linker (ACP_{5a}-ACP_{5b}), and a construct encompassing all three domains but missing the putative dimerization region (KS₅-ACP_{5a}-ACP_{5b}). In order to further define the role of the dimerization domain, we also analyzed by gel filtration and SAXS, the tandem of ACPs and ACP_{5b} in the presence of the dimerization region (ACP_{5a}-ACP_{5b}-DD and ACP_{5b}-DD, respectively). All ACP-containing constructs were studied in their *apo* forms. The number of residues comprising the various domains and linkers is indicated.

Structural analysis of the KS₅-linker by SAXS

Modules 4 and 5 are joined within VirA by an approximately 55 residue linker (GenBank: BAF50727.1). The N-terminal boundary of KS₅ was chosen to lie within this linker, just upstream of the conserved domain as identified in the crystal structures of the KS-AT didomains from DEBS modules 3 and 5.^{24,25} The region downstream of the KS, which is conserved among *trans*-AT PKS,²⁶ has previously been described as a non-functional remnant of an AT domain, and suggested to play a role in docking the *trans*-AT with the PKS. Our reanalysis of the 298 amino acids separating the KS from ACP_{5a} revealed consecutive regions within the first 139 amino acids with convincing homology to the so-called KS-AT and post-AT linkers identified in the DEBS KS-AT didomain structures^{24,25} (Fig. S3†). In these didomains, the KS-AT linker which adopts an $\alpha\beta$ fold, and the post-AT linker which wraps back over both the AT domain and the KS-AT linker to interact with the KS, serve to fix the relative positions of the KS and AT domains. We therefore hypothesized that the KS and these 139 amino acids might form a structural unit. Consistent with this idea, the resulting construct which spanned these regions was obtained in satisfactory yield (6 mg L⁻¹), and shown to be homodimeric as expected²⁷ by native mass spectrometry (calculated dimer MW = 129,461.2 Da; found = 129,480 Da; monomer MW by denaturing mass spectrometry = 64730.0 Da (calc'd = 64730.6 Da)).

The SAXS data on the KS₅-linker were collected on the SWING beamline at the Synchrotron SOLEIL, which allows for online sample purification by gel filtration just prior to analysis. In this way, it is possible to select a population of molecules within the sample which have the same gyration radius (R_g), or average protein dimension. The buffer conditions used in our experiments are known to promote the catalytic activity of both *cis*- and *trans*-AT PKS.^{28–30}

Guinier's plot of the SAXS data acquired on the most abundant conformer of the KS₅-linker gave an R_g of 33.1 ± 0.5 Å, while calculation of the distance distribution function ($P(r)$) using GNOM³¹ yielded a maximum particle diameter (D_{max}) of 105 Å (Fig. 3a and 3b; all SAXS data are summarized in Table S1†). These data and a molecular weight of 137.2 kDa calculated using SAXSMOW³² are consistent with a globular, homodimeric protein (monomer MW = 64.8 kDa).

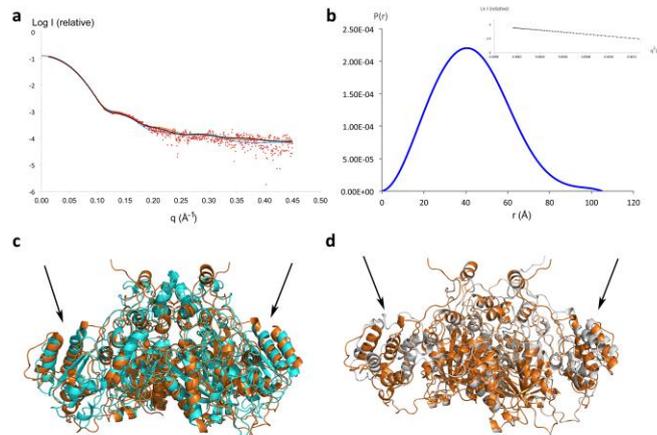


Fig. 3 SAXS analysis of the KS₅-linker. (a) Superimposition of the experimental SAXS data (red dots) with the theoretical curves for the KS₅-linker, calculated on the basis of the solved structures of the DEBS KS₅ (orange) and RhiE KS (cyan) domains as templates for the homology model. The curve in black corresponds to the theoretical curve for the KS₅-linker model derived from DEBS KS₅ as template, after varying the relative orientations of the KS and the KS-AT linker with SASREF. Improved agreement was obtained by rotating the two KS-AT linkers relative to their original orientations by approximately 18 and 21°, respectively, such that they occupy non-equivalent positions relative to the KS dimer. (b) The distance distribution function derived for the KS₅-linker calculated with GNOM. Inset is the Guinier plot. (c) Superimposition of the homology models of the KS₅-linker generated using DEBS KS₅ (orange) and RhiE KS (cyan) as template. (d) Superimposition of the DEBS-derived homology model of the KS₅-linker with that obtained from SASREF, illustrating the change in relative position of the KS-AT linkers (compare the regions indicated with the arrows in panels c and d).

We then derived *ab initio* three dimensional envelopes from the experimental data using the program DAMMIN³¹ with imposition of second order symmetry (Fig. 4a). In order to place the KS domain within the average envelope using SUPCOMB³³ in the absence of its high-resolution structure, we generated a homology model of the domain and adjacent linkers (Fig. 3c) with MODELLER³⁴ using KS₅ of DEBS as a template (40% sequence identity). The resulting model was then dimerized by alignment of the monomers with the structure of the DEBS KS₅-AT₅ didomain (PDB: 2HG4²⁴). (During preparation of this manuscript, the crystal structure of the didomain KS-branching (B) enzyme from the RhiE subunit of the rhizoxin *trans*-AT PKS was published.¹⁵ However, this KS was not a better model for the VirA KS₅, and so all analysis was based on the VirA KS₅-linker model derived from DEBS as template (for further discussion on this point, see the Methods)).

For values above 0.125 \AA^{-1} (Fig. 4a), there is a minor divergence between the calculated and experimental curves, which suggests that a portion of the KS₅-linker construct adopts multiple orientations in solution. Specifically, the different *ab initio* shape calculations reveal small differences in the molecular envelope corresponding to the linker region, suggesting that it is this part of the structure which is flexible. To validate this idea, we aimed to determine whether

ARTICLE

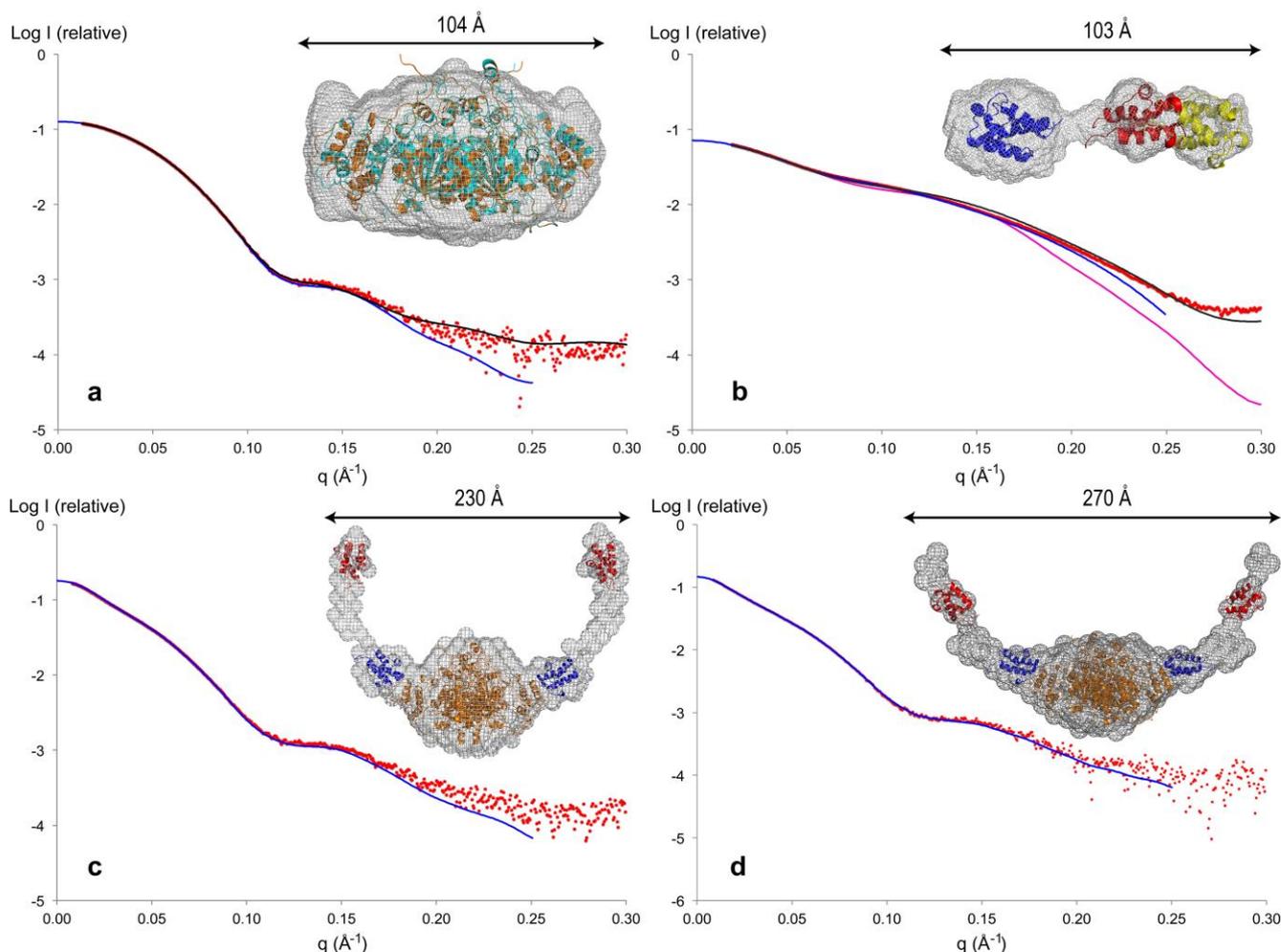


Fig. 4 SAXS profiles and domain organization within module 5. (a) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data obtained on the KS₅-linker (red dots). The fit to the data of the KS₅-linker model generated by rigid body modeling using SASREF is represented in black. Inset, the *ab initio* envelope of the KS₅-linker construct calculated from DAMMIN (grey mesh) is shown in superimposition by SUPCOMB with the KS₅-linker model obtained using the DEBS KS₅-linker as template (cyan), and the model calculated from SASREF (orange). (b) Fit between the *ab initio* model computed with DAMMIN (solid blue line) with the experimental SAXS data for *apo* ACP_{5a}-ACP_{5b} (red dots). Fit to the data of the theoretical and experimental curves computed with OLIGOMER of an 80:20 mixture of the compact and extended forms of the *apo* ACP_{5a}-ACP_{5b} di-domain (black). Inset, the *ab initio* envelope of *apo* ACP_{5a}-ACP_{5b} calculated from DAMMIN (grey mesh) with superimposition of one *apo* ACP (blue) and the second *apo* ACP in two locations (red and yellow) using SUPCOMB. (c) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data acquired on the *apo* KS₅-ACP_{5a}-ACP_{5b} (red dots). Inset, the *ab initio* envelope of the *apo* KS₅-ACP_{5a}-ACP_{5b} construct calculated from DAMMIN (grey mesh) is shown with superimposition by SUPCOMB of the KS₅-linker (orange), *apo* ACP_{5a} (blue) and *apo* ACP_{5b} (red). (d) Fit between the *ab initio* model computed with DAMMIN (solid blue line) and the experimental SAXS data obtained on *apo* KS₅-ACP_{5a}-ACP_{5b}-DD (red dots). Inset, the *ab initio* envelope of the *apo* KS₅-ACP_{5a}-ACP_{5b}-DD construct calculated using DAMMIN (grey mesh) with superimposition of the KS₅-linker (orange), *apo* ACP_{5a} (blue) and *apo* ACP_{5b} (red).

changing the position of the KS-AT linker relative to KS₅ would improve the agreement between the calculated and experimental data. To this end, we performed with SASREF³⁵ a rigid body simulated annealing using a dimeric KS and the two linkers as search models. Improved agreement was obtained for a model in which the two KS-AT linkers are rotated relative to their original orientations by approximately 18 and 21°, respectively, and occupy

apparently non-equivalent positions relative to the KS dimer (compare Fig. 3c and 3d). In contrast, comparison of the crystal structures of the KS-AT linkers from DEBS²⁴ and RhiE¹⁵ shows that their positions superimpose. Taken together, these data underline the power of SAXS relative to crystallography to reveal protein conformational flexibility in solution.

ARTICLE

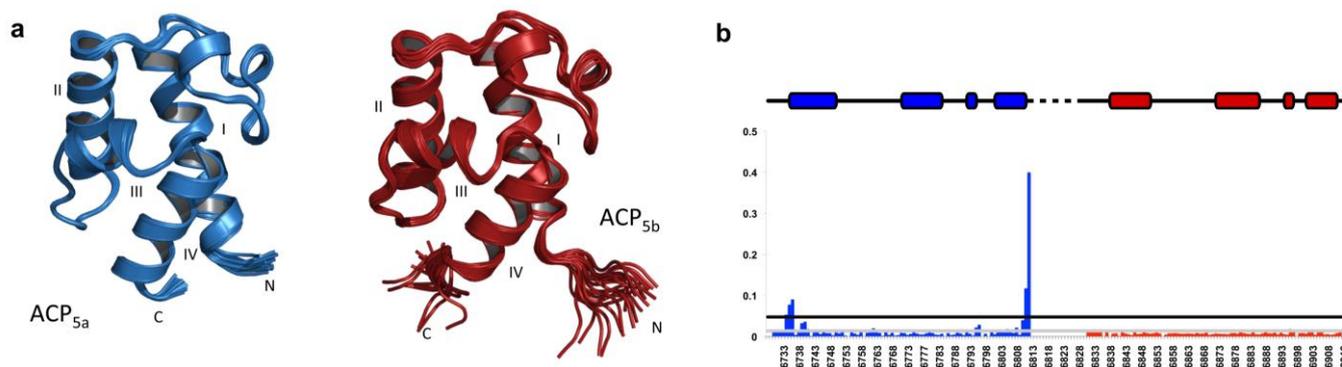


Fig. 5 NMR analysis of *apo* ACP_{5a}, *apo* ACP_{5b} and *apo* ACP_{5a}-ACP_{5b}. (a) The three-dimensional solution structures of the *apo* ACP_{5a} and ACP_{5b} domains were determined using multi-dimensional heteronuclear NMR spectroscopy. Structure calculation was carried out using a two-step procedure consisting of initial structure generation using CYANA followed by restrained molecular dynamics refinement within Amber. The quality of the structures was assessed using PROCHECK-NMR and the molprobity server. Shown are the final ensembles of 20 conformers representing the solution structures of *apo* ACP_{5a} (blue) and *apo* ACP_{5b} (red), depicted with all heavy atoms. As expected based on their sequence identity (52%), *apo* ACP_{5a} and ACP_{5b} have similar structures (RMSD over 72 residues is 1.2 Å). (b) Average ¹⁵N-¹H chemical shift changes plotted as a function of ACP_{5a}-ACP_{5b} residue number. The secondary structure elements present in *apo* ACP_{5a} and ACP_{5b} are shown on top of the plot (the linker is indicated by a dotted line as its secondary structure is unknown). ¹⁵N-¹H chemical shift differences are calculated between chemical shift of equivalent residues in *apo* ACP_{5a}-ACP_{5b} and ACP_{5b} (red). The grey and black lines indicate the mean average chemical shift and the average plus one standard deviation, respectively, and were used to identify significant changes. This analysis demonstrates that the chemical shifts of only a few residues vary between the isolated *apo* ACP domains and the didomain construct.

Structural analysis of ACP_{5a}-ACP_{5b} by SAXS and NMR

The tandem ACPs 5a and 5b, which exhibit 52% mutual sequence identity, are separated by a region of 28 residues rich in Ala and Pro (68%), characteristic of extended, but flexible linkers³⁶ (Fig. S4a†). SAXS analysis of a construct encompassing the two ACPs in their *apo* forms yielded an R_g of 29.0 ± 0.2 Å, and a D_{max} of 103 Å, which are large for a 187-amino acid protein. The distance distribution function (Fig. S5a†) revealed two peaks, consistent with a dumbbell-like architecture, in which two independent domains are joined by an unstructured region. Independent confirmation of this overall shape was provided by the recent publication of the NMR structure of an *apo* ACP-ACP didomain from subunit MmpA of the mupirocin *trans*-AT PKS, which also incorporates a 28-residue ACP-ACP linker (Fig. S4a†).¹⁴

We also solved the NMR solution structures of the two *apo* ACP domains (Fig. 5, Tables S2 and S3†). The structures of the ACP_{5a} and ACP_{5b} domains exhibit a typical ACP fold composed of three major antiparallel α helices (I, II and IV) and a short perpendicular α helix (III). The helices in ACP_{5a} and ACP_{5b} are as follows: helix I, R6734–V6749/A6837–T6850; helix II, S6770–A6783/S6871–Y6885; helix III, A6792–D6795/I6893–D6896; and helix IV P6799–L6810/L6900–R6910. The long loops between helix I and II of ACP_{5a} and ACP_{5b} include two short helical regions (P6754–R6756 and F6763–L6766 for ACP_{5a} and P6855–D6857 and F6864–D6857 for ACP_{5b}). The Ser which serves as the point of attachment of the phosphopantetheine prosthetic group (Ser6770 and S6871, respectively) is located at the N-terminus of helix II. These two structures allowed the sizes of the ACPs to be estimated at 35 Å, which when taken together with the SAXS results, reveal that the intervening linker is approximately 33 Å in length. Analysis of 19 additional sequences separating ACP domains in *trans*-AT PKS (Fig. S4a†) shows that these regions vary in length from 12 to 72

residues, and exhibit no strong homology. However, in common with the Vir ACP_{5a}-ACP_{5b} linker, many of the sequences, including that of MmpA, are biased towards Ala and Pro, suggesting that the global properties of the linkers may be shared. Indeed, the linker regions connecting the five tandem ACPs of the polyunsaturated fatty acid (PUFA) synthase of *Photobacterium profundum* also show this biased amino acid composition (Fig. S4b†), and the overall protein adopts an elongated, ‘beads-on-a-string’ architecture.³⁷

Using SUPCOMB, we next fitted the NMR structures of the two *apo* domains into the average of ten *ab initio* molecular envelopes ($\chi^2 = 1.785$) (Fig. 4b). This analysis showed that the position of one of the ACPs is well-defined, while that of the second is not. Indeed, the shape of the average molecular envelope reveals a spring-like movement along the inter-ACP axis, and thus that the intervening linker is flexible. Structural superimposition of one ACP on the didomain using PyMOL (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) confirms the locational variability of the second ACP, which occupies several positions relative to the other, giving rise to overall compact or extended shapes. The relative percentages of these two configurations were estimated at 80% and 20%, respectively, using OLIGOMER (<http://www.embl-hamburg.de/biosaxs/manuals/oligomer.html>). Notably, this finding is consistent with the form of the $P(r)$ function, in which the second peak presents two maxima (Fig. S5a†).

Our SAXS results additionally suggest that the ACPs do not interact at least in their *apo* forms. In order to confirm this result, we analyzed the *apo* ACP_{5a}-ACP_{5b} didomain by NMR. Comparison of the [¹H, ¹⁵N]-HSQC spectrum of *apo* ACP_{5a}-ACP_{5b} with those of the individual *apo* ACPs revealed no significant differences in chemical shift for key residues, with the exception of amino acids located at or near the covalent linkage between the two domains (Fig. 5b). These data are consistent with the absence of any direct contact between

the tethered ACPs, as found for the ACPs of MmpA¹⁴ and the PUFA synthase.³⁷

Structural analysis of KS₅-ACP_{5a}-ACP_{5b} by SAXS

We next analyzed by SAXS an *apo* construct incorporating all of the functional domains of module 5 in the absence of the putative C-terminal dimerization region, giving an R_g of 56.5 ± 0.7 Å and a D_{max} of 230 Å (Fig. S5b†). An average *ab initio* three dimensional envelope was then calculated from the experimental data with DAMMIN, with imposition of second order symmetry (Fig. 4c). This parameter is justified by the homodimeric nature of the KS and by the fact that lack of such symmetry imposition caused the calculation to diverge, likely due to the large number of amino acids in the protein. Use of MONSA,³⁵ then allowed us to fit simultaneously the obtained diffusion data on the KS₅-linker and the *apo* KS₅-ACP_{5a}-ACP_{5b} constructs into the molecular shape, clearly identifying the position of the dimeric KS domain at the center of the protein ($\chi^2 = 1.885$). The two *apo* ACP NMR structures were then placed into the envelope using as a constraint the D_{max} value of 105 Å obtained for the *apo* ACP_{5a}-ACP_{5b} didomain.

This analysis revealed that, together, the KS and the two ACP domains form an ‘open arms’ structure, wherein ACP_{5a} is located directly adjacent to the KS, and ACP_{5b} lies at the extremity of the arm. The respective locations of the two ACPs suggests that the position of ACP_{5a} is relatively fixed, whereas that of ACP_{5b} is much less well-defined, consistent with the inherent mobility observed for one of the two ACPs of ACP_{5a}-ACP_{5b}. Nonetheless, as ACP_{5a} sits at a distance of approximately 50 Å from the KS (as measured from the KS active site Cys to the Ser of the ACP which is modified by phosphopantetheine), which is out of reach of the 20 Å long prosthetic group, an interaction between the two domains would necessitate a structural rearrangement.

The other notable result from the form calculation is that the 159-residue region C-terminal of the KS-AT and post-AT linkers (designated as the ‘post-post-AT linker’ (Fig. 2)) must adopt a compact structure. It was not possible to model this region, however, due to the complete absence of homologous sequences in the protein database. Nonetheless, inspection of the linker reveals a high proportion of residues which are correlated with natively unfolded proteins (i.e. Q, S, P, E, K, G and A,³⁸ 102/159 residues), an analysis supported by disorder prediction using PONDR-FIT³⁹ (Fig. S6a and S6b†). Such regions often adopt a defined fold only in the presence of other proteins, and are implicated in molecular recognition processes with multiple partners.⁴⁰ To evaluate the generality of this finding, we analyzed the comparable region from 35 additional *trans*-AT PKS modules containing consecutive KS and ACP domains (Fig. S6c†). Although no clear homology to the VirA linker was detected in these sequences (with the exception of the corresponding linker from the pristinamycin II cluster of *Streptomyces pristinaespiralis* which generates the same metabolite²⁷ (Fig. S6d†)) nor any mutual sequence similarity, a large proportion of the regions also are predicted by PONDR-FIT to adopt an unfolded state (Fig. S6e†).

Analysis of KS₅-ACP_{5a}-ACP_{5b}-DD

In silico analysis of module 5 revealed a 50-residue C-terminal sequence with homology to a 51-amino acid region from the CurA subunit of the curacin *cis*-AT PKS (Fig. S1†). Notably, CurA contains a triplet of ACP domains and is also the site of β -modification of the polyketide intermediate. As this region was shown to increase the proportion of dimeric relative to monomeric forms of constructs incorporating the three tandem ACPs,⁴¹ it was designated as a dimerization motif. To evaluate the effect of this putative dimerization element on the overall structure of module 5,

we compared the SAXS data obtained on the entire *apo* module (KS₅-ACP_{5a}-ACP_{5b}-DD) to that of the *apo* KS₅-ACP_{5a}-ACP_{5b} construct. Guinier’s plot yielded an R_g of 69 ± 1 Å, while the distance distribution function produced a D_{max} of 279 Å (Fig. S5c†). As both parameters are higher than that of KS₅-ACP_{5a}-ACP_{5b}, the SAXS analysis revealed that, contrary to expectation, the major conformer of the *apo* module also adopts an open structure in the presence of the putative dimerization domain. An average *ab initio* three dimensional envelope was then calculated from the experimental data with DAMMIN ($\chi^2 = 2.204$), with imposition of second order symmetry (Fig. 4d). The structures of the KS₅-linker and the two *apo* ACPs were then placed within the obtained form by comparison to the data obtained on *apo* KS₅-ACP_{5a}-ACP_{5b}. The important result of this analysis is that the putative dimerization domains on each polypeptide do not interact with each other and that, indeed, the module still assumes an open shape. Furthermore, comparison of the structures of *apo* KS₅-ACP_{5a}-ACP_{5b} and *apo* KS₅-ACP_{5a}-ACP_{5b}-DD shows that the ‘arms’ can move, leading to relatively more open or closed conformations.

This observed lack of dimerization prompted us to determine the oligomerization state of further constructs incorporating the C-terminal dimerization motif and the tandem ACP domains (ACP_{5a}-ACP_{5b}-DD) or the second ACP alone (ACP_{5b}-DD). Both constructs were analyzed in their *apo* forms. Analysis by analytical gel filtration demonstrated that while *apo* ACP_{5a}-ACP_{5b} is a monomer, both *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD elute as oligomers (Fig. S7a–d†). Independent analysis, however, by native mass spectrometry ((Fig. S7e–h†) showed that both constructs incorporating the DD domain maintained the largely monomeric structure of *apo* ACP_{5a}-ACP_{5b}. To understand this discrepancy, we analyzed *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD by SAXS (Fig. 6). This experiment yielded an R_g of 37.0 ± 0.3 Å and a D_{max} of 140 Å for ACP_{5b}-DD (relative to $R_g = 11.1$ Å and $D_{max} = 36.4$ Å for *apo* ACP_{5b} alone, calculated with CRY SOL⁴² based on the NMR structure), and an R_g of 43.5 ± 0.2 Å and a D_{max} of 170 Å for ACP_{5a}-ACP_{5b}-DD (Fig. S5d–g†).

We calculated the *ab initio* molecular forms of the two proteins using DAMMIN with and without imposition of second order symmetry; the calculation performed in the absence of symmetry yielded a significantly better fit to the experimental data. We next attempted using MONSA to determine the contribution of each *apo* ACP to the three-dimensional forms of both constructs, but despite repeated rounds of calculation, this exercise was unsuccessful (for analysis of this calculation, see the Methods). Nonetheless, from the shapes, dimensions and molecular weights of the molecules (calculated using SAXSMOW³² (*apo* ACP_{5b}-DD: calc’d MW = 14.3 kDa, found = 31.4 kDa; *apo* ACP_{5a}-ACP_{5b}-DD: calc’d = 24.5 kDa, found = 48.9 kDa) it is clear that both proteins are dimeric, and adopt extended structures via dimerization of the DD regions (Fig. 6). In view of this result, we interpret the failure to detect dimers by mass spectrometry as evidence that the interfaces have significant hydrophobic character, as such contacts are often lost under the gas-phase conditions of the analysis.⁴³

Although the DD was shown to dimerize small, unconstrained proteins such as ACP_{5b}, it had no effect on the much larger KS₅-ACP_{5a}-ACP_{5b} tridomain, leading to the conclusion that the mechanism of dimerization is conformationally unavailable in the native system. This non-native mode of dimerization prompted us to consider whether the DD region might instead play a role in docking the VirA protein with its downstream partner, VirFG (Fig. 1b). Analysis of the N-terminus of VirFG identifies 70 residues upstream of the conserved KS domain (Fig. S8†). While the end of this region shows sequence characteristics of an intrinsically disordered protein (see the Methods for the sequence of VirFG), the first 28 residues

exhibit convincing sequence homology to the VirA DD region (Fig. S1a†). Thus, the non-native self-association observed for the VirA DD in the *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD constructs is likely a result of the inherent similarity between the DD and its natural VirFG N-terminal docking domain partner. Taken together, these observations are consistent with the idea that the C-terminal terminal DD participates in a docking interaction with the N-terminus of VirFG, possibly through the formation of a four α -helical bundle – a structure which is consistent with the compact, roughly spherical shape of the DD docking complex observed by SAXS (Fig. 6).

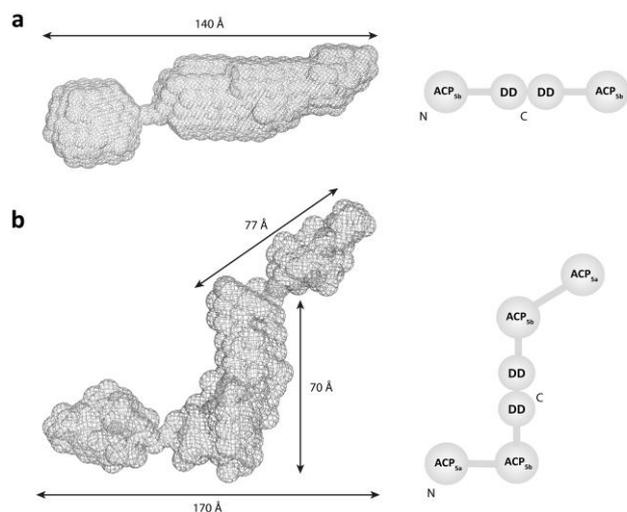


Fig. 6 Molecular envelopes of *apo* ACP_{5b}-DD and ACP_{5a}-ACP_{5b}-DD. (a) The *ab initio* envelope of the *apo* ACP_{5b}-DD calculated with DAMMIN in the absence of imposition of symmetry, shows that the construct adopts an extended, largely cylindrical shape, in which the ACPs are located at the extremities, and the two docking domains dimerize at the center. All of the calculations used to derive the average envelope show that the ACPs are not symmetrical, as the position of one of the two is less well-defined. (b) The *ab initio* envelope of the *apo* ACP_{5a}-ACP_{5b}-DD construct derived with DAMMIN in the absence of imposed symmetry. Measured dimensions of the shapes are indicated with arrows. Interestingly, the SAXS analysis shows that while the ACP_{5b}-DD portion of ACP_{5a}-ACP_{5b}-DD assumes a conformation similar to that in the structure of ACP_{5b}-DD alone, the ACP_{5b}-DD/DD-ACP_{5b} distance is reduced (70 vs. 140 Å). Similarly, the ACP_{5a}-ACP_{5b} distance (77 Å) shows that this region of the structure is also compacted relative to ACP_{5a}-ACP_{5b} alone (103 Å) (Fig. 5). Thus, our SAXS data reveal that the presence of nearby partner domains and linker regions can have a significant effect on the shape and flexibility of specific portions of the PKS.

In support of this hypothesis, the VirA DD region and the N-terminus of VirFG also show low, but convincing homology to the so-called ‘class 2’ C-terminal and N-terminal docking domain regions, respectively, identified in PKS subunits from a range of *cis*-AT systems in both cyanobacteria and myxobacteria^{44,45} (Fig. S1a†). Indeed, the CurA DD also shows a comparable level of sequence similarity to the C-terminal docking domains, and so should be included in this classification. The crystal structures of two such pairs of docking domains from the curacin system were recently reported,⁴⁵ revealing three important findings with relevance to our study: firstly, that the C-terminal docking domains are monomeric and so do not contribute to subunit dimerization; secondly, that docking is mediated by formation of an α -helix bundle; and finally, that the docking interface consists of a complementary hydrophobic surface, surrounded by electrostatic interactions that mediate specificity. Take together with our data, these results strongly suggest that *trans*-AT PKS systems also use such ‘class 2’-like

docking domains to ensure faithful transfer of intermediates between subunits.

Discussion

For modular PKS systems, an X-ray crystal structure of a complete module or multi-modular subunit is still awaited. In any event, as work on the vertebrate fatty acid synthase has shown,⁴⁶ static crystal structures do not capture key features of the highly dynamic native multienzyme. We demonstrate here that SAXS, although an inherently lower resolution technique, can provide invaluable information on whole-module architecture and dynamics in solution. We used a ‘dissect and build’ approach to solve the SAXS structure of *apo* module 5 from the virginiamycin *trans*-AT PKS; starting from the detailed study of smaller protein fragments by SAXS and allied computational methods, it proved possible to progress to analyze both the architecture and the dynamic behavior of the entire module.

Previous structural insights into modular PKS have drawn heavily on data gleaned from mono- and didomain proteins.¹⁰ In particular it has been shown that the KS domains of KS-AT didomains adopt their native homodimeric structures,^{24,25} so we anticipated that our model tridomain protein *apo* KS₅-ACP_{5a}-ACP_{5b}-DD would also reflect the native KS structure. Module 5 is located at the end of subunit VirA where it engages in chain transfer with module 6 located in the downstream subunit VirFG (Fig. 1b). This interaction likely necessitates the formation of a specific intersubunit protein-protein complex. As we have studied isolated module 5, it is relevant to consider whether the complex with module 6 is transient (in which case a free VirA C-terminus is physiologically relevant) or long-lived. Measured K_d values for interactions between several *cis*-AT PKS subunits are in the range of 2.6–23.7 μ M,^{45,47,48} placing them in the class of transient protein-protein interactions which characteristically exhibit a mixture of oligomeric states.^{49,50} Indeed, the three subunits of the erythromycin (DEBS) PKS do not form a stable complex, at least *in vitro*.⁵¹ The advantage of such temporary associations is that they would allow for the exchange of damaged subunits which would otherwise block the entire assembly line. In view of measured intracellular protein concentrations (most notably 0.05–0.1 μ M for the PksR subunit of the bacillaene *trans*-AT system,⁵² and more generally, 1 nM–10 μ M average concentration for *E. coli* proteins⁵³), this weak binding argues that a significant fraction of VirA module 5 will be uncomplexed at equilibrium. While the bacillaene NRPS-PKS was found to form a membrane-associated megacomplex in *Bacillus subtilis* cells,⁵² the molecular basis for this macro-association and thus its consequences for the affinity and lifetimes of individual intersubunit interfaces in this and other actinomycete NRPS-PKS are unknown. Therefore, taken together, the available data support the validity of isolated VirA module 5 as a model system for studying the structure of a PKS module.

Our analysis of *apo* module 5 has revealed the average locations within the module of the dimeric KS and the two ACP domains. The divergent positioning of the two ACPs raises questions as to their respective interactions with the KS, and indeed with other enzymatic partners of the module. The role of such tandem ACPs in the mupirocin *trans*-AT PKS has previously been investigated by gene inactivation experiments *in vivo*.⁵⁴ For the doublet ACP of this system, the data suggest that the two domains play essentially equivalent roles, while for the triplet cluster, the first ACP appears to be arranged in series relative to the following two. The non-equivalent locations of ACP_{5a} and ACP_{5b} relative to the KS in Vir module 5, as well as differences in their inherent mobility, suggest a physical explanation for this in-series operation. In this mode of

ARTICLE

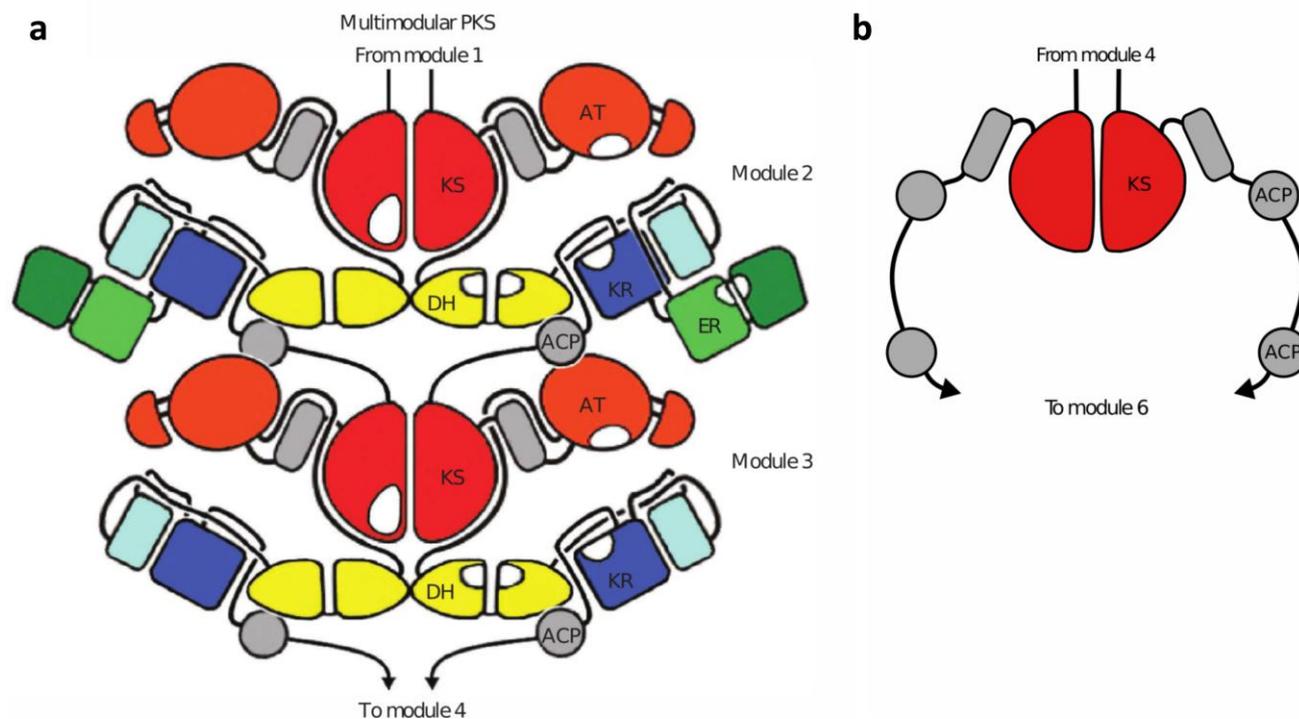


Fig. 7 Structures of PKS modules. (a) Structural model of *cis*-AT PKS modules incorporating various complements of reductive domains, based on the X-ray structure of animal FAS. The overall structure is homodimeric, with the ketosynthase (KS; red) and dehydratase (DH; yellow) domains comprising the dimer interface. The remaining monomeric catalytic domains (acyl transferase (AT; orange), ketoreductase (KR; blue) and enoyl reductase (ER; green)) loop out from the two-fold axis. Thus, each module comprises two independent reaction centers, each of which is accessible by a single acyl carrier protein (ACP) domain (the involved active sites are indicated in white). Each ACP additionally interacts with a downstream KS domain to which, in this example, it is covalently tethered. The grey box between the KS and AT domains represents the intervening, folded linker, a homologue of which is found in virginiamycin module 5 (see discussion in the text). Artwork reprinted by permission from Macmillan Publishers Ltd: *Nat. Chem. Biol.* (8, 615), copyright (2012), and from AAAS (*Science* (2008) 321, 1315). (b) Schematic of the structure of an *apo trans*-AT PKS module as revealed by SAXS coupled with NMR and homology modeling. The core of the structure is also formed by the KS domain, but together, the two ACP domains, the linker regions and the terminal dimerization/ docking domains form an open architecture.

biosynthesis, ACP_{5a} would participate in chain extension with the KS, while ACP_{5b} would serve as the site for modification reactions on the extended unit, with direct transfer of the intermediate between the two ACPs.^{54,55} Nonetheless, the validity of this mechanism for the virginiamycin system remains to be directly demonstrated by analogous mutagenesis experiments *in vivo* or *in vitro*.

Our SAXS experiments have also provided important insights into the structure and function of interdomain linkers within *trans*-AT PKS. As in *cis*-AT systems, the KS domain and the first portion of its downstream linker, the sequence of which is conserved among many *trans*-AT systems (Fig. S3†),²⁶ form a structural unit. As this linker additionally shares an interface with the AT domain in *cis*-AT PKS,^{24,25} it may serve as a docking point for the *trans*-acting AT, as previously proposed.²⁶ Thus, despite the distinct evolutionary origin of the *cis*- and *trans*-AT PKS,⁵ these key operational aspects appear to be common to the two systems. The short linker joining the two ACP domains is relatively extended and flexible, allowing a significant mobility of ACP_{5b}. In contrast, the remaining 159 residues of the KS-ACP linker, which exhibit convincing sequence

characteristics of a natively-disordered region,³⁸ adopt a compact structure, presumably due to contacts with other portions of the module. Although the precise lengths and sequences of these regions are not conserved among KS-ACP and KS-ACP-ACP modules in *trans*-AT PKS (Fig. S6†), their global properties appear to be shared. Taken together, these data have clear implications for attempts to reconfigure these systems by genetic engineering within the linker regions, as experiments which fail to respect the overall chemical character and function of these sequences are unlikely to succeed.

Module 5 ends in a 50-residue region which is predicted to include two terminal α -helices (Fig. S1†). The presumed helical regions within this structure show homology to a 51-residue motif which follows a triplet of ACP domains at the end of the curacin PKS subunit CurA.⁴¹ As judged by analytical ultracentrifugation, CurA *apo* ACP-ACP-ACP constructs appeared to be largely monomeric whether or not this region was present. However, analysis by size-exclusion chromatography and negative-stain electron microscopy revealed that its presence led to a substantial

shift in the monomer-dimer equilibrium towards dimer. Given these data, our finding that inclusion of the putative dimerization domain did not lead to closure of the module was surprising. An explanation is provided by our SAXS analysis (Fig. 6), which revealed that two N-terminally truncated versions of the module incorporating the DD region dimerize in a non-native fashion, a mode of association which is not possible in the context of the complete module. We propose that this unnatural dimerization mimics the natural interaction between the C-terminal DD and the N-terminus of module 6, as both of these regions show homology to class 2 docking domains of *cis*-AT PKS⁴⁵ (Fig. S1†).

Modules in which β -processing reactions occur are positioned at the termini of *trans*-AT PKS modules as in the case of the virginiamycin system, but they also occur at internal locations within the multienzyme subunits (such as within the bacillaene, myxovirescin, pederin, onnamide, etnangien, oocydin, kalimanticin, psymberin, coralopyronin and bongkreic acid PKS^{4,18–21,56}). In the latter cases, the ACPs are covalently attached to the downstream KS domain of the next module. However, in these systems, the linkers joining the ACPs of β -modification modules to the downstream KS domains are on average significantly longer than the corresponding linkers following non-modifying modules (ca. 63 vs. 46 residues; Fig. S9a†). It is also striking that all of the ACP-KS linkers from *trans*-AT PKS are substantially longer than the comparable regions from *cis*-AT PKS (average length 20 amino acids, Fig. S9b†). Given that the average size of an amino acid is 3.4 Å in a maximally-extended polypeptide,⁵⁷ these ACP-KS linkers would appear to be compatible with an open conformation of the preceding β -modification modules, even if a portion of the linker adopts secondary structure. These data further support the physiological relevance of the module 5 structure observed by SAXS.

Nevertheless, it remains likely that module 5 clamps down transiently during transfer of the fully-processed intermediate to the homodimeric KS of module 6, a structural change which may be induced by formation of a compact complex between the putative C- and N-terminal docking domains.⁴⁵ Thus, overall, our SAXS data lead us to propose that the module adopts several distinct conformations from open to relatively closed. Indeed, it is possible that phosphopantetheinylation of the two ACPs to generate the *holo* module might in itself induce an extent of modular closure. On the other hand, the observed open form, in which in particular ACP_{5b} is highly accessible, may be necessary to allow one or both of the carrier protein domains to interact with their many *in trans* catalytic partners, possibly through formation of a large, multi-protein complex. Our results underscore the inherent flexibility of the ACP-ACP region, which may facilitate this suite of interactions. Full enzymatic processing of the intermediate may then trigger structural or compositional changes in the module which permit formation of a docking interaction with the downstream subunit VirFG, and efficient transfer of the chain to the module 6 KS domain. Such a mechanism, when coupled with the intrinsic substrate specificity of the receiving KS,⁵ could aid in programming the biosynthesis, ensuring that the full set of modification reactions occurs prior to transacylation. Experiments to evaluate this hypothesis are under way. In the meantime, it is worth emphasizing that the structure of *apo* Vir module 5 differs considerably from models proposed for the *cis*-AT PKS systems (compare Fig. 7a and 7b) in which chain assembly occurs in two discrete reaction chambers,^{11,58} underlining the necessity to study the two types of systems independently.

Methods

Cloning, expression and purification

Eight constructs (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, ACP_{5a}-ACP_{5b}-DD, KS₅-linker, KS₅-ACP_{5a}-ACP_{5b}, and KS₅-ACP_{5a}-ACP_{5b}-

DD) were amplified from *S. virginiae* genomic DNA using forward (f) and reverse (r) primers incorporating *Bam*HI and *Hind*III restriction sites, respectively, and were ligated into the equivalent sites of vector pBG-106 (KS₅-linker, KS₅-ACP_{5a}-ACP_{5b}, and KS₅-ACP_{5a}-ACP_{5b}-DD) or pBG-102 (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, and ACP_{5a}-ACP_{5b}-DD). The primer sequences were as follows (the introduced restriction sites are underlined, while the bases annealing to gDNA are capitalized): KS f, (*Bam*HI) 5'-tttccggaatccGCGATCGCCGTGATCGGCGTTG-3' (used for KS₅-linker, KS₅-ACP_{5a}-ACP_{5b}, KS₅-ACP_{5a}-ACP_{5b}-DD); KS r (*Hind*III) 5'-tttccgaagctttcaCAGGTGGGCGATCTGCAGGTCGTACCAG-3'; ACP_{5a} f (*Bam*HI) 5'-tttcgcggaatccGCGGGCCGTCAGGAGGAG-3' (used for ACP_{5b}-ACP_{5b}, ACP_{5a}-ACP_{5b}-DD); ACP_{5a} r (*Hind*III) 5'-tttccgaagctttcaGGCTCCGAGCGACTCGGC-3' (used for ACP_{5a}); ACP_{5b} f (*Bam*HI) 5'-tttcgcggaatccGCCGTGGCCGTCGACCCCG-3' (used for ACP_{5b}, ACP_{5b}-DD); ACP_{5b} r (*Hind*III) 5'-tttccgaagctttcaGACCGGGGCGCGCGTCC-3' (used for ACP_{5b}, ACP_{5a}-ACP_{5b}); DD r (*Hind*III) 5'-tttcgcaagcttTCAGCGCCGCGGGGTGAGCAG-3' (used for ACP_{5b}-DD, ACP_{5a}-ACP_{5b}-DD, KS₅-ACP_{5a}-ACP_{5b}-DD). Vector pBG-102 codes for a SUMO-His₆ tag, while vector pBG106 codes for a His₁₀ tag and a flexible linker containing the precision 3C-protease cleavage site (Dr. L. Mizoue, Center for Structural Biology, Vanderbilt University). Following cleavage of the SUMO-His₆ tag, the proteins incorporate a non-native N-terminal Gly-Pro-Gly-Ser sequence. The vectors were used to transform *E. coli* BL21, and constructs were expressed as His₁₀ tagged (KS₅, KS₅-ACP_{5a}-ACP_{5b}, and KS₅-ACP_{5a}-ACP_{5b}-DD) or SUMO-His₆-tagged (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, and ACP_{5a}-ACP_{5b}-DD) fusions by growth in LB medium at 37 °C to an A₆₀₀ of 0.8, followed by induction with IPTG (0.2 mM) and incubation at 20 °C for a further 12–18 h. The *E. coli* cells were collected by centrifugation and resuspended in buffer 1 (50 mM sodium phosphate (pH 7.5), 250 mM NaCl). Cells were lysed by sonication, and cell debris were removed by centrifugation and filtration (0.45 μ m). The cell lysates were loaded onto a HisTrap 5 mL column (GE), equilibrated in buffer 1. The column was washed extensively with buffer 1 containing 75 mM imidazole, and His-tagged proteins were eluted at 350 mM imidazole. The SUMO-His₆ tagged constructs (ACP_{5a}, ACP_{5b}, ACP_{5b}-DD, ACP_{5a}-ACP_{5b}, and ACP_{5a}-ACP_{5b}-DD) were then incubated with His-tagged human rhinovirus 3C protease (1 μ M) for 12–16 h at 4 °C to cleave the affinity-solubility tags. The constructs were then separated from the remaining His-tagged proteins by loading onto a HisTrap 5 mL column (GE) followed by elution in buffer 1 containing 20 mM imidazole. Size exclusion chromatography on a Superdex 75 26/60 column (GE) in buffer 1 resulted in a homogeneous preparation of each protein (Fig. S2†). Following the HisTrap column, the His₁₀-tagged KS construct was purified to homogeneity by size exclusion chromatography on a Superdex 200 26/60 column (GE) in buffer 1 (Fig. S2†). The His₁₀-tagged constructs KS₅-ACP_{5a}-ACP_{5b} and KS₅-ACP_{5a}-ACP_{5b}-DD were purified by ion exchange chromatography on a HiLoad 16/10 Q Sepharose column (GE) equilibrated in buffer 2 (20 mM sodium phosphate (pH 8)) which was washed with buffer 2 containing 250 mM NaCl. The constructs were then eluted in 400 mM NaCl. Size exclusion chromatography on a Superdex 200 26/60 column (GE) in buffer 1 resulted in a homogeneous preparation of each protein (Fig. S2†). Production of ¹³C,¹⁵N-enriched ACPs (ACP_{5a}, ACP_{5b}, ACP_{5a}-ACP_{5b}) for NMR structure elucidation was carried out by growth in M9 minimal medium. The minimal medium was supplemented with ¹⁵NH₄Cl (0.5 g L⁻¹) and glucose or ¹³C₆-glucose (2.0 g L⁻¹) as the sole sources of nitrogen and carbon.

Correction of the sequence of *virFG*

In order to correct the end of *VirF*, we used a reverse primer complementary to the original end of the gene (5'-ttcggagctcTCAGCACCGCGCGCCGG-3'; the capitalized bases correspond to nucleotides 29985–30002 of GenBank AB283030.2) and a forward primer complementary to the original beginning of *virF* (5'-aaaaaacatgcatgCCGCCGAGCGCTGGC-3'; the capitalized bases correspond to nucleotides 32211–32226 of GenBank AB283030.2) and sequenced. This analysis revealed a missing C nucleotide at the end of *virF* (original position 30064 of *virF*). Addition of this nucleotide results in a frameshift in the gene which removes the stop codon. *virF* and *virG* are in fact a single gene, which we have designated as *virFG*. To correct the beginning of *virFG*, we designed a forward primer complementary to the upstream *virE* sequence (5'-aaaaaacatggcCCGCCGGCCTCCGACACCCTG-3'; the capitalized bases correspond to nucleotides 32615–32636 of GenBank AB283030.2), and a reverse primer complementary to the original *virF* gene (5'-ttcggagctcTCA TGTGACGGGGGCGGGGCGTTC-3'; the bases in bold correspond to the complement of nucleotides 30514–30534 of GenBank AB283030.2). The PCR products were inserted into the pCR-BLUNT vector (Invitrogen) and sequenced. This analysis revealed that a nucleotide (T) had been incorrectly added at position 32254 of the published *virF* sequence, and that in fact, the gene begins at nucleotide 32539. The corrected sequence of *virFG* corresponds much more closely to the comparable region of the homologous pristinamycin cluster (gene *snaE3*).⁵⁹ The final sequence of the translated VirFG is provided in Fig. S8†.

Native and denaturing electrospray ionization mass spectrometry

Prior to mass analysis, the samples were subjected to buffer exchange into 50 mM ammonium acetate (pH 6.9), using an Ultrafree-0.5 centrifugal filter device with a cut-off of 10 kDa (Millipore). Mass spectrometry measurements were performed in positive ion mode on a microTOF-Q (BRUKER Daltonics) equipped with an ESI source. The injection rate for the samples was 6 $\mu\text{L min}^{-1}$. Spectra were smoothed using the Gaussian method and were baseline subtracted. For native analysis, the desalted sample was diluted into 50 mM ammonium acetate to a final concentration of 20 μM . The ESI needle voltage was set to 4.5 kV, the nebulization gas (Ar) pressure was 1.7 bar, the drying gas flow was 6 L min^{-1} , the source temperature was 200 $^{\circ}\text{C}$ and the capillary exit voltage was set to 198 V. The acquisition range was m/z 2000–10,000. For denaturing analysis, the desalted sample was diluted in $\text{H}_2\text{O}/\text{ACN}/\text{FA}$ (50:50:1 volume) at a final concentrations of 10 μM . The ESI needle voltage was set to 4.5 kV, the nebulization gas (Ar) pressure was 0.8 bar, the drying gas flow was 4 L min^{-1} , the source temperature was 190 $^{\circ}\text{C}$ and the capillary exit voltage was set to 155 V. The acquisition range was m/z 500–2,500.

Small angle X-ray scattering (SAXS) data collection

The SAXS data were collected on the SWING beamline at the Synchrotron SOLEIL, using an AVIEX170170 CCD detector. Frames were recorded using an energy of 12 keV. The sample-to-detector distance was set to 1899 mm, leading to scattering vectors q ranging from 0.0005 \AA^{-1} to 0.5 \AA^{-1} . The scattering vector is defined as $q = 4\pi/\lambda \sin\theta$, where 2θ is the scattering angle. The protein samples were injected with an automatic sample changer into a pre-equilibrated HPLC-coupled size-exclusion chromatography column (Bio-SEC 300, Agilent). The online purification system that delivers the eluted fractions into the measurement cell was developed on the SWING beamline.⁶⁰ After equilibrating the column in the protein buffer (50 mM sodium phosphate (pH 7.5), 150 mM NaCl, 5%

glycerol), 100 successive frames of 0.75 s were recorded of the buffer background. This buffer was chosen for the SAXS analysis as phosphate, salt and glycerol are known to be important for the catalytic activity of *cis*-AT PKS,^{28–30} presumably through effects on the structural integrity of the proteins. Similarly, the buffers used for recent enzymatic analysis of KS domains from *trans*-AT PKS contained both salt and glycerol⁶¹ while that for activity experiments with a *trans*-AT contained phosphate.¹³ 80 μL of the protein sample (at 15 mg mL^{-1}) were then injected and complete data sets acquired. All experiments were performed at 10 $^{\circ}\text{C}$. In contrast to classical SAXS experiments which are conducted in batch using several protein concentrations within a standard range (e.g. 0.1–10 mg mL^{-1}), the fact that data collection is coupled to a gel filtration column means that analysis of the required multiple concentrations of the protein occurs within a single experiment (as measured downstream via the absorbance at 280 nm with an *in situ* spectrophotometer), as many different positions within the elution peak are sampled during the course of the measurement (typically 50–100 frames are acquired).⁶⁰ Each acquisition frame of the experiment yields a scattering spectrum which is then analyzed by the dedicated, in-house program FoxTrot, yielding a R_g (radius of gyration) as well as an $I(0)$ value (which depends on the protein concentration at that location in the elution peak) from the Guinier law (approximation $I(q) = I(0) \exp(-q^2 R_g^2/3)$ for $qR_g < 1.3$). It is thus the fact that the concentration naturally varies across a gel filtration peak that provides a full range of dilutions as used in classical SAXS experiments. Under our experimental conditions (buffer, dilution over the gel filtration step, etc...), the fact that we observed a constant R_g for a significant proportion of the concentrations present in our gel filtration peaks means that these measurements were independent of concentration, and thus, that we were effectively under conditions of infinite dilution. In the following step of data analysis, all the frames that exhibited the same R_g as a function of $I(0)$ were averaged, thus ensuring that the data reflected only the signal arising from the protein structure, and not from intermolecular interactions. Processing of the averaged data was performed using the program package PRIMUS.⁶² The distance distribution function $P(r)$ and the maximum particle diameter D_{max} were calculated by Fourier inversion of the scattering intensity $I(q)$ using GNOM.³¹

Ab initio shape determination

The overall *ab initio* protein shapes were calculated from the experimental SAXS data using the bead-modeling program DAMMIN.³¹ At least ten to fifteen independent fits were carried out without imposition of symmetry for the monomeric *apo* ACP_{5a}-ACP_{5b} construct, and with a 2-fold symmetry restraint for the KS₅-linker, *apo* KS₅-ACP_{5a}-ACP_{5b} and *apo* module 5 proteins; in the absence of such an imposed symmetry, the fit between the experimental and model SAXS curves was poorer (greater χ^2). The results of several DAMMIN models were then compared using the program SUPCOMB³³ in order to determine a consensus model. Low resolution models were then averaged using the DAMAVER and DAMFILT packages.⁶³ In addition, *ab initio* shape determination by a multiphase dummy atom mode using MONSA³⁵ which permitted the simultaneous fitting of scattering data sets obtained on KS₅-ACP_{5a}-ACP_{5b} and KS₅-ACP_{5a}-ACP_{5b}-DD constructs and the individual KS₅-linker domain, allowed for the localization of the domains within the various multienzymes.

In the case of the *apo* ACP_{5b}-DD and *apo* ACP_{5a}-ACP_{5b}-DD constructs, the calculation of the *ab initio* molecular forms was carried out without imposition of symmetry (P1) because this yielded a better fit to the data than when second order symmetry was imposed (P2) ($\chi^2 = 1.21$ vs. 1.6 for ACP_{5b}-DD; 1.3 vs. 1.55 for ACP_{5a}-ACP_{5b}-DD). We attempted to determine using MONSA the

individual contributions of the *apo* ACPs to the calculated molecular forms (that is, to try to place the ACPs into the forms), by fitting the diffusion curves of the two ACPs (calculated from the NMR structures using CRY SOL) and ACP_{5a}-ACP_{5b} in the case of ACP_{5a}-ACP_{5b}-DD, into the diffusion curves obtained on the overall proteins. In our hands, this calculation proved impossible due to the high mutual similarity of the ACPs, their inherent positional flexibility, and to the fact that the inclusion of ACP_{5a} and the ACP_{5a}-ACP_{5b} linker changed the relative orientations and distances between ACP_{5b} and the DD.

Molecular modeling

A three-dimensional structural model of the KS₅-linker was generated by pairwise comparison of profile hidden Markov models using the HHpred server (<http://toolkit.tuebingen.mpg.de/hhpred>).⁶⁴ Local sequence alignment was performed using PSI-BLAST and a multiple sequence alignment including information about predicted secondary structures was produced. The closest homologues to the VirA KS₅-linker within the Protein Data Bank are the DEBS KS₅-AT₅ didomain (40% sequence identity; PDB: 2HG4²⁴) and a didomain KS-branching (B) enzyme from the subunit RhiE (37% identity; PDB: 4KC5¹⁵). Three-dimensional models of the KS₅-linker were then constructed using either PDB 2HG4 or 4KC5 as template, using MODELLER.³⁴ The resulting models were then dimerized by superposition onto the KS domains of the dimeric crystal structures from the PDBs 2HG4 and 4KC5. Manual inspection of the resulting dimeric models revealed no steric clashes within the models. CRY SOL⁴² was then used to calculate the theoretical scattering curve of the two homology models. The quality of the 3D modeling was determined using the discrepancy χ^2 , defined according to Konarev and colleagues.⁶²

The theoretical scattering curves predicted by CRY SOL⁴² from the DEBS-derived and RhiE-derived models of the VirA KS₅-linker gave an equally good fit to the experimental data ($\chi^2 = 1.581$ vs. 1.582, respectively) (Fig. 3c). Part of the reason that the RhiE KS did not produce a superior model may be that its KS-AT linker region contains a 23-residue insertion relative to VirA, which is also not present in DEBS module 5 (Fig. S3†). Correspondingly, analysis of the RhiE KS-B structure shows that the KS-AT linker incorporates an additional β -strand relative to the DEBS KS-AT linker region.

Having selected the model based on the DEBS KS₅ template, the atomic model was superimposed onto the *ab initio* calculated shape of the protein using SUPCOMB.³³ Rigid body modeling against SAXS data was performed using SASREF.³⁵ In particular, SASREF was employed to position the KS and the linker with respect to each other by simulated annealing, while minimizing the discrepancy between the SAXS experimental and calculated curves. Figures displaying the protein structures were generated using PYMOL (Version 1.3, Schrödinger, LLC) and the low-resolution bead models from DAMMIN were converted into mesh.

Analysis of the relative positions of ACP_{5a} and ACP_{5b} was also attempted using SASREF. However, ten independent rounds of calculation produced divergent results that yielded poor fits to the experimental data for angles between 0.0005 and 0.2 Å⁻¹, with χ^2 between 25.0 and 25.3. Furthermore, the R_g values calculated for the models generated by SASREF are smaller than those derived from Guinier and the $P(r)$ function. The failure of SASREF to more accurately place the two domains relative to one other reveals that the SAXS data cannot be accounted for by a single conformation of the tandem of ACP domains.

NMR data acquisition

To minimize the amount of protein needed, 300 μ L of protein solution was loaded into a 4 mm NMR tube and the tube was

inserted into a 5 mm NMR tube containing 150 μ L D₂O.⁶⁵ The *apo* ACP_{5a} and ACP_{5b} domains were analyzed at a concentration of 1.0 mM and the *apo* ACP_{5a}-ACP_{5b} didomain at a concentration of 0.3 mM. All NMR data were recorded at 298 K on a Bruker DRX600 spectrometer equipped with a cryogenic probe (Plateforme de Biophysico-chimie Biologie Structurale – FR 3209 Université de Lorraine/CNRS). Backbone and sequential resonance assignments were obtained by the combined use of 2D ¹⁵N-¹H and ¹³C-¹H HSQC spectra and 3D HNCACB, CBCA(CO)NH and HNCO experiments (reviewed in ref. 66). Aliphatic side chain resonance assignments were obtained from 3D (H)CC(CO)NH, H(CCCO)NH, and HCCH-TOCSY experiments. ¹H chemical shift assignments of aromatic side chains were based on 2D heteronuclear (HB)CB(CGCD)HD and (HB)CB(CGCD)HE and 2D homonuclear COSY, TOCSY and NOESY experiments. Tautomeric states of histidines were determined using 2D heteronuclear long range ¹⁵N-¹H HSQC.⁶⁷ To collect NOE-based distance restraints for the ACP_{5a} and ACP_{5b} structure calculations, 2D ¹H homonuclear NOESY experiments were carried out on unlabelled samples of *apo* ACP_{5a} and ACP_{5b}, while 3D ¹⁵N NOESY-HSQC and ¹³C NOESY-HSQC were recorded on uniformly ¹³C,¹⁵N-enriched *apo* ACP_{5a} and ACP_{5b} samples. The mixing time used in all NOESY experiments was set to 120 ms. NMR data were processed using Topspin 3.1 (Bruker) and were analyzed using Sparky (Goddard TD and Kneller DG, SPARKY3, University of California, San Francisco, 2003 (<http://www.cgl.ucsf.edu/home/sparky/>)).

Chemical shift changes

To express the chemical shift variations of the corresponding amide groups in *apo* ACP_{5a} vs. *apo* ACP_{5a}-ACP_{5b} and *apo* ACP_{5b} vs. *apo* ACP_{5a}-ACP_{5b}, the chemical shift change was defined as $\Delta\delta_{av} = ((\Delta\delta H^2/2) + \{\Delta\delta N^2/10\})^{1/2}$.⁶⁸ The most significant changes were observed for Arg6734, Gln6735, Gly6811 and Ala6812 of ACP_{5a} (Fig. 5). Consideration of the *apo* ACP_{5a} structure strongly suggests that the changes can be attributed to the presence of the linker in the didomain. In the ACP_{5a} structure, the β -methyl group of Ala6812 is located in proximity to the Arg6734-Gln6735 amide (measured distance less than 4 Å). However, when the linker is present (residues 6811–6828), Ala6812 is displaced from its position near Arg6734 and Gln6735, affecting not only its chemical shift, but those of the two previously neighboring residues. The chemical shift of Gly6811 is altered by the change in covalent linkage. This interpretation, in combination with the fact that the chemical shifts of ACP_{5b} are essentially unchanged in the didomain, leads us to conclude that the two *apo* ACPs behave as independent domains.

NMR structure calculation

The first round of structure calculation was performed with CYANA 3.⁶⁹ Starting from a set of manually assigned NOEs, the standard CYANA protocol of seven iterative cycles of calculations was performed with NOE assignment by the embedded CANDID routine combined with torsion angle dynamics structure calculation.⁷⁰ In each cycle, 100 structures starting from random torsion angle values were calculated with 10,000 steps of torsion angle dynamics-driven simulated annealing. 1968 NOE-based distance and 114 backbone angle restraints were used for structure calculation of *apo* ACP_{5a}, while 1846 NOE-based distance and 114 backbone angle restraints were used for *apo* ACP_{5b}. The angle restraints were obtained from ¹³C _{α} , ¹³C _{β} , ¹³C', ¹⁵N, ¹HN and ¹H _{α} chemical shifts using TALOS+⁷¹ with an assigned minimum range of $\pm 20^\circ$. No hydrogen-bond restraints were used for structure calculation. The second stage consisted of the refinement of the 50 lowest CYANA target function conformers by rMD simulations in Amber 11⁷² using the generalized Born solvent model. Possible atoms clashes within CYANA

structures were regularized by a 1 ps energy minimization in the Amber force field, and then 20 ps of rMD (restrained molecular dynamic) was performed with the following protocol: the system was heated for 5 ps at 600 K with tight coupling for heating and equilibration (TAUTP = 0.4), 15 ps of cooling to 0 K with 13 ps of slow cooling (loose coupling, TAUTP = 4.0–1.0) followed by 1 ps of faster cooling (TAUTP = 1.0), and a final 1 ps of very fast cooling (TAUTP = 0.1–0.05). During rMD, the restraints were slowly ramped from 10–100% of their final values over the first 3 ps. Force constants for distance and angle restraints were set to 32 kcal mol⁻¹ Å⁻² and 50 kcal mol⁻¹ rad⁻², respectively. The final *apo* ACP_{5a} and ACP_{5b} representative ensembles correspond to the 20 conformers from each calculation with the lowest restraint and Amber energy terms. The quality of the structures was assessed using PROCHECK-NMR⁷³ and the MolProbity⁷⁴ server. The Ramachandran statistics for *apo* ACP_{5a} and ACP_{5b} are respectively: most favored regions, 99.1% and 91.9%; additionally allowed regions, 0.9% and 7.6; generously allowed regions, 0.0% and 0.5%; and disallowed regions, 0.0% and 0.0%. MolProbity scores (percentiles) for the lowest restraint energy structures of ACP_{5a} and ACP_{5b} are 0.99 (100th) and 0.88 (100th), respectively.

Sequence analysis

The sequences of the analyzed *cis*- and *trans*-AT PKS⁴ were accessed via PubMed, or in the case of the class 2 C- and N-terminal docking domains, from ref. 45. Sequence alignments were generated using Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>), and displayed using BoxShade (http://www.ch.embnet.org/software/BOX_form.html). The boundaries of the KS domain, and the KS-AT and post-AT linkers were assigned by comparison with the sequences of homologous regions from the erythromycin polyketide synthase,²⁴ while the boundaries of the structured ACP domains were determined by NMR (this study). The putative dimerization/docking regions were assigned as the sequences lying downstream from the last ACP domains of the respective subunits.

Accession numbers

Chemical shifts for *apo* ACPs 5a and 5b of VirA module 5 have been deposited in the Biological Magnetic Resonance Bank with accession codes 19536 (ACP_{5a}) and 19539 (ACP_{5b}). Protein Data Bank coordinates have been deposited under accession codes 2MF4 and 4CA3, respectively.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR JCJC 2011 PKS-PPIs to K.J.W.), the Centre National de la Recherche Scientifique (CNRS), the University of Lorraine (UL), and the Lorraine Region (BQR grants to K.J.W. and B.C.). Access to the 600 MHz NMR spectrometer of the 'SCBIM' (Service Commun de Biophysicochimie des Interactions Moléculaires) of the FR3209 CNRS/Université de Lorraine and to the Bruker MicrOTOF-Q (Service Commun de Spectrométrie de Masse et de Chromatographie, UL) was greatly appreciated. We thank Dr J. Perez and Dr P. Roblin for help during data collection on the beamline SWING at the synchrotron SOLEIL (GIF-sur-YVETTE), and Prof. P. F. Leadlay, Dr. B. Wilkinson and Dr. V. Receveur-Bréchot for helpful comments on the manuscript.

Notes and references

^a Molecular and Structural Enzymology Group, Université de Lorraine, IMoPA, UMR 7365, Vandœuvre-Les-Nancy, F-54500, France. CNRS, IMoPA, UMR 7365, Vandœuvre-Les-Nancy, F-54500, France. E-mail: benjamin.chagot@univ-lorraine.fr, arnaud.gruez@univ-lorraine.fr, kira.weissman@univ-lorraine.fr

^b 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.

† Electronic Supplementary Information (ESI) available: nine figures, three tables, and supplementary references. See DOI: 10.1039/b000000x/

- 1 S. Smith and S. C. Tsai, *Nat. Prod. Rep.*, 2007, **24**, 1041–1072.
- 2 J. Crosby and M. P. Crump, *Nat. Prod. Rep.*, 2012, **29**, 1111–1137.
- 3 L. Katz, *Methods Enzymol.*, 2009, **459**, 113–142.
- 4 J. Piel, *Nat. Prod. Rep.*, 2010, **27**, 996–1047.
- 5 T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, and J. Piel, *Nat. Biotechnol.*, 2008, **26**, 225–233.
- 6 S. Sudek, N. B. Lopanik, L. E. Waggoner, M. Hildebrand, C. Anderson, H. Liu, A. Patel, D. H. Sherman and M. G. Haygood, *J. Nat. Prod.*, 2007, **70**, 67–74.
- 7 C. M. Thomas, J. Hotherhall, C. L. Willis and T. J. Simpson, *Nat. Rev. Microbiol.*, 2010, **8**, 281–289.
- 8 N. Pulsawat, S. Kitani and T. Nihira, *Gene*, 2007, **393**, 31–42.
- 9 K. J. Weissman and P. F. Leadlay, *Nat. Rev. Microbiol.*, 2005, **3**, 925–936.
- 10 A. T. Keatinge-Clay, *Nat. Prod. Rep.*, 2012, **29**, 1050–1073.
- 11 J. Zheng, D. C. Gay, B. Demeler, M. A. White and A. T. Keatinge-Clay, *Nat. Chem. Biol.*, 2012, **8**, 615–621.
- 12 T. Maier, M. Leibundgut and N. Ban, *Science*, 2008, **321**, 1315–1322.
- 13 F. T. Wong, X. Jin, I. I. Mathews, D. E. Cane and C. Khosla, *Biochemistry*, 2011, **50**, 6539–6548.
- 14 A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hotherhall, E. Ploskoń, P. Wattana-amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi, J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas and M. P. Crump, *Nat. Chem. Biol.*, 2013, **9**, 685–692.
- 15 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.
- 16 M. Czjzek, H.-P. Fierobe and V. Receveur-Bréchot, *Methods Enzymol.*, 2012, **510**, 183–210.
- 17 R. Dehn, Y. Katsumaya, A. Weber, K. Gerth, R. Jansen, H. Steinmetz, G. Höfle, R. Müller and A. Kirschning, *Angew. Chem. Int. Ed Engl.*, 2011, **50**, 3882–3887.
- 18 N. Moebius, C. Ross, K. Scherlach, B. Rohm, M. Roth and C. Hertweck, *Chem. Biol.*, 2012, **19**, 1164–1174.
- 19 W. Mattheus, L.-J. Gao, P. Herdewijn, B. Landuyt, J. Verhaegen, J. Masschelein, G. Volckaert and R. Lavigne, *Chem. Biol.*, 2010, **17**, 149–159.
- 20 O. Erol, T. F. Schäberle, A. Schmitz, S. Rachid, C. Gurgui, M. El Omari, F. Lohr, S. Kehraus, J. Piel, R. Müller and G. M. König, *ChemBioChem*, 2010, **11**, 1253–1265.
- 21 M. A. Matilla, H. Stöckmann, F. J. Leeper and G. P. C. Salmond, *J. Biol. Chem.*, 2012, **287**, 39125–39138.
- 22 H. Irschik, M. Kopp, K. J. Weissman, K. Buntin, J. Piel and R. Müller, *ChemBioChem*, 2010, **11**, 1840–1849.
- 23 C. T. Calderone, *Nat. Prod. Rep.*, 2008, **25**, 845–853.
- 24 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.
- 25 Y. Tang, A. Y. Chen, C.-Y. Kim, D. E. Cane and C. Khosla, *Chem. Biol.*, 2007, **14**, 931–943.
- 26 Y.-Q. Cheng, J. M. Coughlin, S.-K. Lim and B. Shen, *Methods Enzymol.*, 2009, **459**, 165–186.
- 27 A. M. Haapalainen, G. Meriläinen and R. K. Wierenga, *Trends Biochem. Sci.*, 2006, **31**, 64–71.
- 28 M. Bycroft, K. J. Weissman, J. Staunton and P. F. Leadlay, *Eur. J. Biochem.*, 2000, **267**, 520–526.
- 29 R. Pieper, S. Ebert-Khosla, D. Cane and C. Khosla, *Biochemistry*, 1996, **35**, 2054–2060.
- 30 K. E. H. Wiesmann, Ph. D. Thesis, University of Cambridge, 1995.
- 31 D. Svergun, *J. Appl. Crystallogr.*, 1992, **25**, 495–503.
- 32 H. Fischer, M. de Oliveira Neto, H. B. Napolitano, I. Polikarpov and A. F. Craievich, *J. Appl. Crystallogr.*, 2009, **43**, 101–109.
- 33 M. Kozin and D. Svergun, *J. Appl. Crystallogr.*, 2001, **34**, 33–41.

- 34 N. Eswar, B. Webb, M. A. Marti-Renom, M. S. Madhusudhan, D. Eramian, M.-Y. Shen, U. Pieper and A. Sali, *Curr. Protoc. Protein Sci.*, 2007, **50**, 2.9.1–2.9.31.
- 35 M. V. Petoukhov and D. I. Svergun, *Biophys. J.*, 2005, **89**, 1237–1250.
- 36 R. N. Perham, *Annu. Rev. Biochem.*, 2000, **69**, 961–1004.
- 37 U. Trujillo, E. Vázquez-Rosa, D. Oyola-Robles, L. J. Stagg, D. A. Vassallo, I. E. Vega, S. T. Arold and A. Baerga-Ortiz, *PLoS One*, 2013, **8**, e57859.
- 38 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.
- 39 B. Xue, R. L. Dunbrack, R. W. Williams, A. K. Dunker, and V. N. Uversky, *Biochem. Biophys. Acta*, 2010, **1804**, 996–1010.
- 40 P. E. Wright and H. J. Dyson, *Curr. Opin. Struct. Biol.*, 2009, **19**, 31–38.
- 41 L. Gu, E. B. Eisman, S. Dutta, T. M. Franzmann, S. Walter, W. H. Gerwick, G. Skiniotis and D. H. Sherman, *Angew. Chem. Int. Ed Engl.*, 2011, **50**, 2795–2798.
- 42 D. Svergun, C. Barberato and M. H. J. Koch, *J. Appl. Crystallogr.*, 1995, **28**, 768–773.
- 43 C. Bich, S. Baer, M. Jecklin and R. Zenobi, *J. Am. Soc. Mass Spectrom.*, 2010, **21**, 286–289.
- 44 M. Thattai, Y. Burak and B. I. Shraiman, *PLoS Comput. Biol.*, 2007, **3**, 1827–1835.
- 45 J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman and J. L. Smith, *Chem. Biol.*, 2013, **20**, 1340–1351.
- 46 E. J. Brignole, S. Smith and F. J. Asturias, *Nat. Struct. Mol. Biol.*, 2009, **16**, 190–197.
- 47 T. J. Buchholz, T. W. Geders, F. E. Bartley, 3rd, K. A. Reynolds, J. L. Smith and D. H. Sherman, *ACS Chem. Biol.*, 2009, **4**, 41–52.
- 48 R. S. Gokhale, D. Hunziker, D. E. Cane and C. Khosla, *Chem. Biol.*, 1999, **6**, 117–125.
- 49 I. M. Nooren and J. M. Thornton, *J. Mol. Biol.*, 2003, **325**, 991–1018.
- 50 S. E. A. Ozbabacan, H. B. Engin, A. Gursoy and O. Keskin, *Protein Eng. Des. Sel.*, 2011, **4**, 635–648.
- 51 P. Caffrey, D. J. Bevitt, J. Staunton and P. F. Leadlay, *FEBS Lett.*, 1992, **304**, 225–228.
- 52 P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner and R. Kolter, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 305–310.
- 53 Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili and X. S. Xie, *Science*, 2010, **329**, 533–538.
- 54 A. S. Rahman, J. Hothersall, J. Crosby, T. J. Simpson and C. M. Thomas, *J. Biol. Chem.*, 2005, **280**, 6399–6408.
- 55 I. Thomas, C. J. Martin, C. J. Wilkinson, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2002, **9**, 781–787.
- 56 K. M. Fisch, C. Gurgui, N. Heycke, S. A. van der Sar, S. A. Anderson, V. L. Webb, S. Taudien, M. Platzner, B. K. Rubio, S. J. Robinson, P. Crews and J. Piel, *Nat. Chem. Biol.*, 2009, **5**, 494–501.
- 57 G. Yang, C. Ceconi, W. A. Baase, I. R. Vetter, W. A. Breyer, J. A. Haack, B. W. Matthews, F. W. Dahlquist and C. Bustamante, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 139–144.
- 58 J. Zheng, C. D. Fage, B. Demeler, D. W. Hoffman and A. T. Keatinge-Clay, *ACS Chem. Biol.*, 2013, **8**, 1263–1270.
- 59 Y. Mast, T. Weber, M. Gözl, R. Ort-Winklbauer, A. Gondran, W. Wohlleben and E. Schinko, *Microb. Biotechnol.*, 2011, **4**, 192–206.
- 60 G. David and J. Pérez, *J. Appl. Crystallogr.*, 2009, **42**, 892–900.
- 61 M. Jenner, S. Frank, A. Kampa, C. Kohlhaas, P. Pöplau, G. S. Briggs, J. Piel and N. J. Oldham, *Angew. Chem. Int. Ed Engl.*, 2013, **52**, 1143–1147.
- 62 P. V. Konarev, V. V., Volkov, A. V. Sokolova, M. H. J. Koch and D. I. Svergun, *J. Appl. Crystallogr.*, 2003, **36**, 1277–1282.
- 63 V. Volkov and D. Svergun, *J. Appl. Crystallogr.*, 2003, **36**, 860–864.
- 64 J. Söding, A. Biegert and A. N. Lupas, *Nucleic Acids Res.*, 2005, **33**, W244–248.
- 65 M. W. Voehler, G. Collier, J. K. Young, M. P. Stone and M. W. Germann, *J. Magn. Reson.*, 2006, **183**, 102–109.
- 66 J. Cavanagh, W. Fairbrother, A. Palmer and N. Skelton, *Protein NMR Spectroscopy: Principles and Practice*, 1996, Academic Press Inc., New York.
- 67 J. G. Pelton, D. A. Torchia, N. D. Meadow and S. Roseman, *Protein Sci.*, 1993, **2**, 543–558.
- 68 L. Tran, R. W. Broadhurst, M. Tosin, A. Cavalli and K. J. Weissman, *Chem. Biol.*, 2010, **17**, 705–716.
- 69 P. Güntert, *Methods Mol. Biol.*, 2004, **278**, 353–378.
- 70 T. Herrmann, P. Güntert and K. Wüthrich, *J. Mol. Biol.*, 2002, **319**, 209–227.
- 71 Y. Shen, F. Delaglio, G. Cornilescu and A. Bax, *J. Biomol. NMR*, 2009, **44**, 213–223.
- 72 D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668–1688.
- 73 R. A. Laskowski, J. A. Rullmann, M. W. MacArthur, R. Kaptein and J. M. Thornton, *J. Biomol. NMR*, 1996, **8**, 477–486.
- 74 V. B. Chen, W. B. Arendall, 3rd, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr. D Biol. Crystallogr.*, 2010, **66**, 12–21.