JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Characterization of Intersubunit Communication in the Virginiamycin *trans*-Acyl Transferase Polyketide Synthase

Jonathan Dorival,[†,⊥] Thibault Annaval,[†,⊥] Fanny Risser,[†] Sabrina Collin,[†] Pierre Roblin,[‡,§] Christophe Jacob,[†] Arnaud Gruez,[*,†] Benjamin Chagot,[*,†] and Kira J. Weissman[*,†]

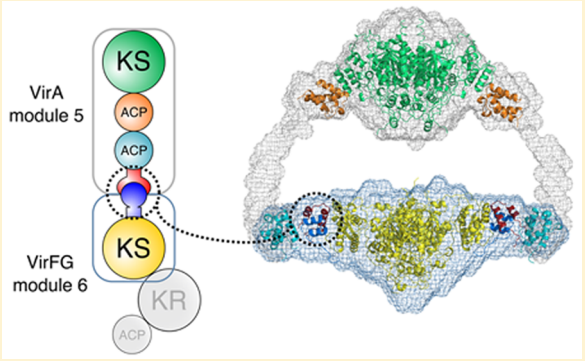
[†]UMR 7365, Ingénierie Moléculaire et Physiopathologie Articulaire (IMoPA), CNRS-Université de Lorraine, Biopôle de l'Université de Lorraine, Campus Biologie Santé, 9 Avenue de la Forêt de Haye, CS 50184, 54505 Vandœuvre-lès-Nancy CEDEX, France

[‡]Synchrotron SOLEIL, L'Orme des Merisiers, Saint-Aubin BP 48, 91192 Gif-sur-Yvette CEDEX, France

[§]UR1268 Biopolymères, Interactions Assemblages (BIA), INRA, Rue de la Géraudière BP 71627, 44316 Nantes CEDEX 3, France

Supporting Information

ABSTRACT: Modular polyketide synthases (PKSs) direct the biosynthesis of clinically valuable secondary metabolites in bacteria. The fidelity of chain growth depends on specific recognition between successive subunits in each assembly line: interactions mediated by C- and N-terminal "docking domains" (DDs). We have identified a new family of DDs in *trans*-acyl transferase PKSs, exemplified by a matched pair from the virginiamycin (Vir) system. In the absence of C-terminal partner (VirA ^{C}DD) or a downstream catalytic domain, the N-terminal DD (VirFG ^{N}DD) exhibits multiple characteristics of an intrinsically disordered protein. Fusion of the two docking domains results in a stable fold for VirFG ^{N}DD and an overall protein−protein complex of unique topology whose structure we support by site-directed mutagenesis. Furthermore, using small-angle X-ray scattering (SAXS), the positions of the flanking acyl carrier protein and ketosynthase domains have been identified, allowing modeling of the complete intersubunit interface.



INTRODUCTION

Synthesis of structurally complex, reduced polyketides in bacteria occurs on gigantic multienzymes called polyketide synthases (PKSs).[1] These systems are likened to molecular-scale assembly lines because each round of chain extension and chemical tailoring of newly incorporated building blocks is carried out by a specific module of enzymatic domains. Nature has evolved two distinct classes of modular PKSs: *cis*-acyl transferase (AT) and *trans*-AT. Within the majority of *cis*-AT PKSs, such as the prototypical 6-deoxyerythronolide B synthase (DEBS) responsible for assembling the core of erythromycin A, all of the component domains are present within the gigantic polypeptides.[2] *trans*-AT PKSs, however, are characterized by one or more free-standing enzyme activities (including notably the ATs) that collaborate in *trans* with the PKS multienzymes to accomplish both chain building and processing (Figure 1).[3,4] They also typically include modules of nonribosomal peptide synthetase (NRPS) that incorporate amino acids into the growing intermediates.

One conserved feature of *cis*- and *trans*-AT PKSs is the distribution of the functional modules among multiple polypeptide subunits.[5] Thus, generation of native polyketide products critically depends on the ability of the subunits to arrange themselves in the correct order. Conversely, rational manipulation of interpolypeptide interfaces to change the sequence of the subunits represents an attractive strategy for generating polyketide analogues because no modifications are made to the component modules. Indeed, modules are increasingly viewed as discrete operational units, given the high interdependence of the functional domains.[6,7]

In *cis*-AT PKSs, it has been shown that communication at intersubunit junctions is mediated by portable recognition sequences called docking domains (DDs), located at the extreme termini of the subunits, and these have already been exploited to generate polyketide analogues.[8−12] Three evolutionarily related docking domain classes have been identified to date in such PKS systems,[13] as well as a fourth class which operates at NRPS−NRPS/−PKS junctions in mixed PKS-NRPSs.[14] The NMR or X-ray crystal structures of covalently fused complexes of the PKS docking domains have been solved[15−17] (Figure 2), and the N-terminal DDs were shown to adopt the same structures when expressed together with the downstream ketosynthase (KS) domains.[17−19] The close structural homology between the two actinobacterial classes of DDs prompted their grouping into Class 1 (referred to here as 1a and 1b), whereas the third class, which is present

Received: December 22, 2015
Published: March 16, 2016

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

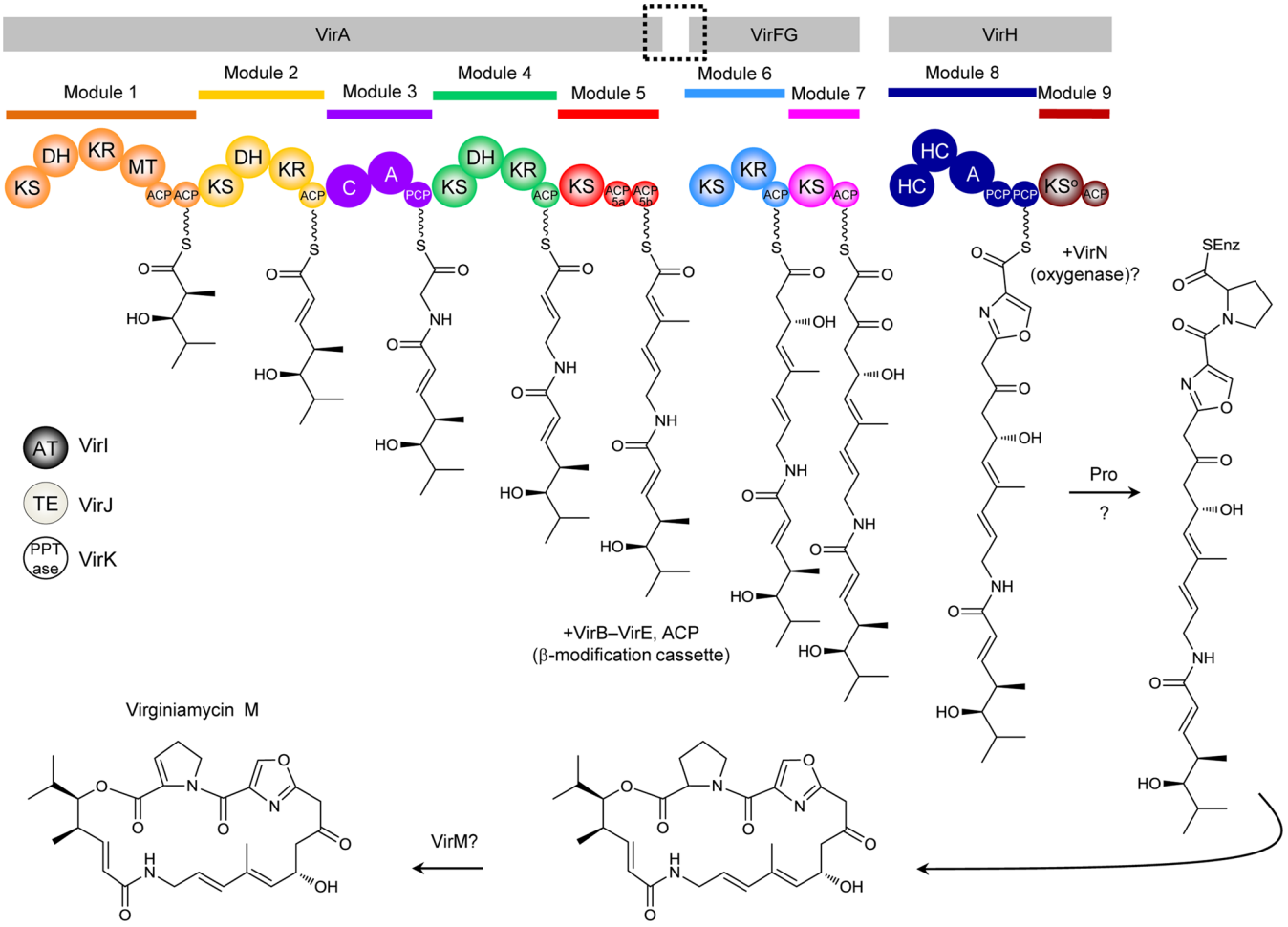


Figure 1. Virginiamycin (Vir) *trans*-AT polyketide synthase. The PKS comprises three subunits, VirA, VirFG (whose N-terminal sequence was recently revised by us),[21] and VirH as well as additional modules that were not identified in the initial cluster description[4] for starter unit selection and introduction of proline. The ? indicate additional steps where the enzyme responsible is not yet clear. The system incorporates many features characteristic of *trans*-AT PKS systems, including a *trans*-acting acyl transferase VirI, duplicated domains (acyl carrier proteins (ACPs) of modules 1 and 5; peptidyl carrier proteins (PCPs) of module 8), nonribosomal peptide synthetase modules (3 and 8), an inactive domain (KS° of module 9), and a cassette of enzymes that together introduce a *β*-methyl functionality. The experiments described here reveal the molecular details of the interface between subunits VirA and VirFG (boxed).

in myxo- and cyanobacteria, has been designated as Class 2.[17] (By extension, the mixed PKS-NRPS DDs constitute Class 3.)

Class 1 and 2 DDs share two central features: the N-terminal docking domains are homodimeric and form a coiled-coil motif that serves as a platform for docking of the likewise α-helical C-terminal docking domains. The docking interfaces are largely hydrophobic in character and complementary in shape, whereas specific charge−charge interactions at critical positions contribute to docking specificity. One notable difference between the two classes is that Class 1 C-terminal docking domains (Figure 2A,B) incorporate two α helices upstream of the docking α helix which together form a dimerization motif. Class 2 C-terminal docking domains, however, comprise only two α helices, both of which associate with the N-terminal docking domain. Furthermore, the coiled-coil motif of the Class 2 N-terminal DD is preceded by an additional α helix; thus, docking gives rise to an overall eight α-helical bundle (Figure 2C), in contrast to the four α-helical bundles characteristic of Class 1 DDs. Despite this architectural divergence and an approximately 2-fold difference in interface size, the affinity of association for both DD classes when the immediate upstream/downstream functional domains are present is remarkably consistent, with K_ds in the 2−20 μM range.[16,17,20]

We recently characterized by small-angle X-ray scattering (SAXS)[21] the structure of *apo* module 5 from the VirA subunit of the virginiamycin hybrid *trans*-AT PKS-NRPS[4] of *Streptomyces virginiae* (Figure 1). In the course of this work, we noted that the extreme C-terminal sequence (designated "VirA ^{C}DD") showed similarity to a region from the CurA subunit of the curacin *cis*-AT PKS, which had been described as a dimerization motif.[22] However, comparison by SAXS of *apo* module 5 in the presence and absence of this region did not reveal a change in the structure of the C-terminus because the module retained its open form. This prompted us to investigate whether VirA ^{C}DD was rather a docking domain, mediating a specific interaction with a partner DD located at the N-terminus of subunit VirFG ("VirFG ^{N}DD"). We show here that VirA ^{C}DD and VirFG ^{N}DD constitute the founding members of a new family of docking domains from *trans*-AT PKSs. Using a multidisciplinary approach combining NMR, site-directed mutagenesis, circular dichroism (CD), isothermal titration calorimetry (ITC), and SAXS, we have characterized their interaction in detail. We have furthermore employed SAXS to

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

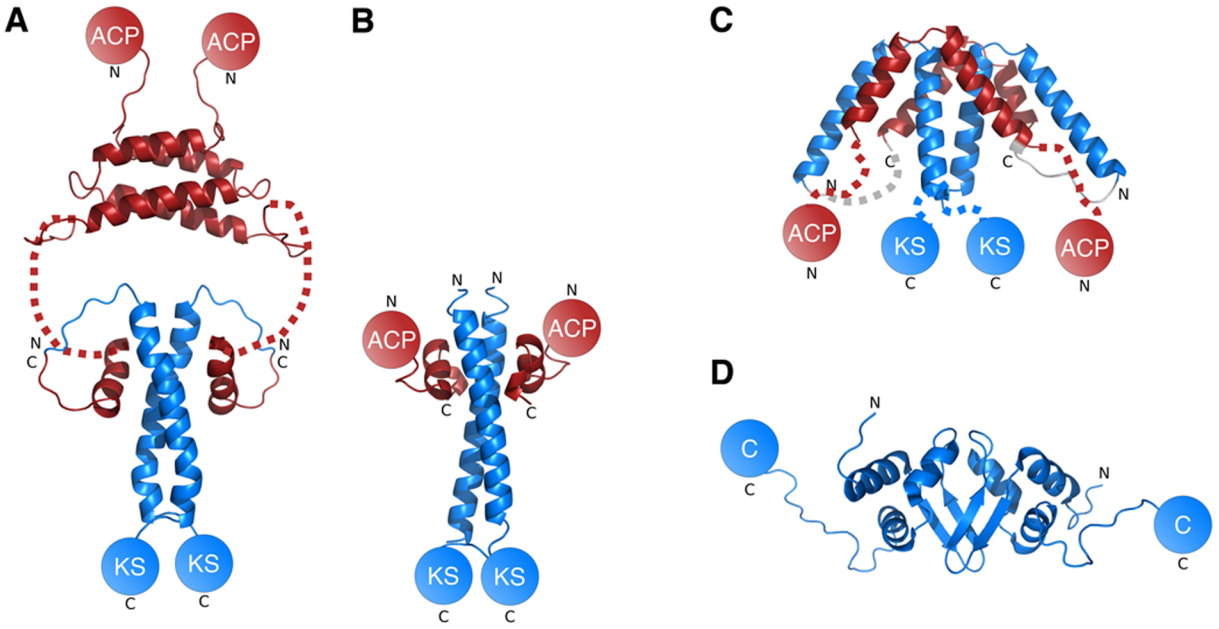


Figure 2. Three classes of docking domains from *cis*-AT PKSs. (A) NMR structure of a complex of covalently fused Class 1a docking domains (Protein Databank (PDB): 1PZQ, 1PZR).[15] The C-terminal DDs (red) comprise three α helices, of which the first two form a four α-helical bundle that acts as a dimerization element. The third helix wraps around the coiled-coil motif formed by the partner N-terminal docking domain (blue). The long, flexible linker joining the second and third α helices of the C-terminal DD is represented as a dashed line. C and N indicate C- and N-termini, respectively. (B) X-ray crystal structure of a covalent complex of Class 1b docking domains (PDB: 3F5H).[16] Although by sequence analysis this class also incorporates a C-terminal dimerization motif, it was not included in the investigated construct. The overall structure resembles that in A, but the C-terminal DD comprises two smaller α helices, and the precise hydrophobic and polar interactions at the interface differ. (C) X-ray crystal structure of a covalent complex of class 2 docking domains (PDB: 4MYY).[17] The C-terminal DD does not incorporate a dimerization region. Instead, its two α helices associate with two α helices provided by the N-terminal docking domain to form an overall 8 α-helical bundle. (D) NMR structure of a class 3 N-terminal docking domain (PDB: 2JUG).[14] Each protomer within the DD homodimer comprises three α helices and two β strands, which adopt an overall $\alpha\beta\beta\alpha\alpha$ topology. For all docking domain complexes, charge−charge interactions are critical determinants of specificity. Key to the domains up- and downstream of the docking domain complexes: ACP, acyl carrier protein; KS, ketosynthase; and C, condensation.

locate the immediate upstream acyl carrier protein (ACP) and downstream ketosynthase (KS) domains in positions that are fully compatible with our recent SAXS analysis of virginiamycin module 5 (ref 21). Furthermore, modeling of the overall VirA/VirFG interface strongly suggests that interaction between successive subunits establishes a closed reaction chamber for chain extension, a fundamentally different mode of operation compared to that of *cis*-AT PKSs.

METHODS

Sequence Analysis. The sequences of the analyzed *trans*-AT PKSs were accessed via PubMed, or in the case of the Class 2 C- and N-terminal docking domains, from ref 17. Sequence alignments were generated using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) with secondary structure prediction by PsiPred[23] and displayed using BoxShade (http://www.ch.embnet.org/software/BOX_form.html), or using PROMALS3D.[24] The C-terminal boundaries of the ACP domains were determined by multiple sequence alignment of the subunit ends and comparison with domain boundaries identified in the solved structures of ACPs 5a and 5b from the virginiamycin PKS.[21] Similarly, the start sites of the domains downstream of the N-terminal docking domains were assigned by alignment of the beginning of the subunits and comparison with the sequences of the solved structures of KS,[18,19] condensation (C),[25] and enoyl reductase (ER)[26] domains.

Materials and DNA Manipulation and Sequencing. *S. virginiae* genomic DNA was obtained from the National Institute of Agrobiological Sciences Genebank, Japan. Biochemicals and media were purchased from ThermoFisher Scientific (Tris, EDTA), Merck (NaPi), Carlo ERBA (NaCl), BD (peptone, yeast extract), VWR (glycerol), and Sigma-Aldrich (imidazole, IPTG). The enzymes for genetic manipulation were purchased from ThermoFisher Scientific. DNA isolation and manipulation in *Escherichia coli* was carried out according to standard methods.[27] PCR amplifications were carried out on a Mastercycler Pro (Eppendorf) using Phusion high-fidelity DNA polymerase (Thermo Scientific). Primers were obtained from Sigma-Aldrich (France), and DNA sequencing was carried out by GATC Biotech (Mulhouse, France).

Cloning, Expression, and Purification. Five constructs (VirA ^{C}DD, VirFG ^{N}DD, VirFG ^{N}DD-linker, VirA ACP$_{5b}$-^{C}DD, and VirFG ^{N}DD-KS$_6$) were amplified from *S. virginiae* genomic DNA using forward and reverse primers incorporating *Bam*HI and *Hind*III restriction sites, respectively (Table S1), and were ligated into the equivalent sites of vector pBG-102 (Center for Structural Biology, Vanderbilt University); in the case of VirFG ^{N}DD-KS$_6$, an internal *Bam*HI site was removed by site-directed mutagenesis using Phusion high-fidelity DNA polymerase, without introducing changes in the amino acid sequence. Vector pBG-102 codes for a His$_6$-SUMO tag. Following cleavage of the SUMO-His$_6$ tag, the proteins incorporated a non-native N-terminal Gly−Pro−Gly−Ser sequence. From these vectors and *S. virginiae* genomic DNA, covalent fusions (VirA ^{C}DD−VirFG ^{N}DD, VirA ACP$_{5b}$-^{C}DD−VirFG ^{N}DD, and VirA ACP$_{5b}$-^{C}DD−VirFG ^{N}DD-KS$_6$) were constructed using overlap PCR. Specifically, fragments encoding VirA ^{C}DD, VirFG ^{N}DD, VirA ACP$_{5b}$-^{C}DD, and VirFG ^{N}DD-KS$_6$ were amplified and then used as templates in a second PCR to amplify the fusion construct, and the resulting fragments were ligated into the vector pBG-102. Site-directed mutants of VirA ^{C}DD and VirFG ^{N}DD were obtained by site-directed mutagenesis using Phusion high-fidelity DNA polymerase from vector pBG-102 containing VirA ^{C}DD and VirFG ^{N}DD wild-type sequences.

The vectors were used to transform *E. coli* BL21 (DE3), and constructs were expressed as His$_6$-SUMO-tagged fusions by growth in LB medium at 37 °C to an A_{600} of 0.8, followed by induction with

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

IPTG (0.1 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) in the case of VirA ^{C}DD, VirA ACP$_{5b}$-^{C}DD, VirA ACP$_{5b}$-^{C}DD−VirFG ^{N}DD (covalent fusion), and all VirA ^{C}DD mutants and in buffer 2 (50 mM Tris-HCl (pH 7.5), 150 mM NaCl) in the case of VirFG ^{N}DD, VirFG ^{N}DD-linker, VirA ^{C}DD−VirFG ^{N}DD (covalent fusion), and all VirFG ^{N}DD mutants, and in buffer 3 (50 mM sodium phosphate (pH 7.5), 400 mM NaCl, 10% glycerol) for VirA ACP$_{5b}$-^{C}DD−VirFG ^{N}DD-KS$_6$ (covalent fusion). Cells were lysed by sonication, and cell debris was removed by centrifugation and filtration (0.45 μm). The cell lysates were loaded onto a HisTrap 5 mL column (GE), equilibrated in buffer 1, 2, or 3 as appropriate. The column was washed extensively with buffer 1, 2, or 3 containing 60 mM imidazole, and His-tagged proteins were eluted using 300 mM imidazole. The VirA ACP$_{5b}$-^{C}DD−VirFG ^{N}DD-KS$_6$ covalent fusion was further loaded onto a HiTrap Q HP column (GE) equilibrated in buffer (20 mM sodium phosphate (pH 7.5), 2 mM EDTA, 5% glycerol), which was then washed extensively with this buffer. The fusion was eluted using a NaCl gradient, up to a final concentration of 1 M. All His$_6$-SUMO-tagged constructs were then incubated with His-tagged human rhinovirus 3C protease (1 mM) 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, 2, or 3 containing 20 mM imidazole.

Following this step, size-exclusion chromatography on a Superdex 75 26/60 column (GE) in 100 mM sodium phosphate (pH 7.5) or on a Superdex 200 26/60 column (GE) in the case of VirA ACP$_{5b}$-^{C}DD−VirFG ^{N}DD-KS$_6$ (covalent fusion) in 20 mM sodium phosphate (pH 7.5), 250 mM NaCl, 2 mM EDTA, and 5% glycerol resulted in a homogeneous preparation of each protein. Production of ^{15}N- or ^{13}C,^{15}N-enriched VirA ^{C}DD, VirFG ^{N}DD, and VirA ^{C}DD−VirFG ^{N}DD (covalent fusion) for NMR structure elucidation was carried out by growth in M9 minimal medium. The minimal medium was supplemented with ^{15}NH$_4$Cl (0.5 g/L) and ^{13}C$_6$-glucose (2.0 g/L) as the sole sources of nitrogen and carbon. Isotopically-labeled chemicals were purchased from Eurisotop.

Analytical Gel Filtration. VirA ^{C}DD and VirFG ^{N}DD (initial concentrations of 670 μM) were loaded separately onto a Superdex 75 10/300 GL column (GE) equilibrated with 100 mM sodium phosphate (pH 6.5) buffer and elution of the two proteins followed via their absorbance at 280 nm. The same volume of a mixture of the two proteins (both at 670 μM) was then reinjected. Calibration was carried out using Gel Filtration Standard (Bio-Rad).

Size-Exclusion Chromatography Multiangle Laser Light Scattering. The oligomeric state of multiple constructs was determined by size-exclusion chromatography multiangle laser light scattering (SEC-MALLS) by the Macromolecular Interaction Platform of I2BC (UMR 9198, http://www.i2bc.paris-saclay.fr). For this, SEC was first carried out on a Superdex 200 10/300 GL Increase column (GE Healthcare Life Sciences) at 17 °C in phosphate buffer (50 mM NaPi (pH 7.5), 250 mM NaCl, 5% glycerol) using a Shimadzu Prominence HPLC. Multiangle laser light scattering (MALLS) was measured using a MiniDAWN TREOS (Wyatt Technology) equipped with a QELS module (Wyatt Technology). Refractometry was monitored using an Optilab T-rEX (Wyatt Technology). Data processing was carried out with ASTRA 6 software (Wyatt Technology) using a refractive-index increment of 0.183 mL/g (dn/dC) to calculate the protein concentration from the refractive-index measurements.

Isothermal Titration Calorimetry. ITC was performed using a MicroCal ITC 200 (GE) at 20 °C. A 300 μL aliquot of VirA ^{C}DD (wild type or mutant) at 60 μM was placed in the calorimeter cell, and VirFG ^{N}DD (wild type or mutant at 600 μM) or VirFG ^{N}DD-linker (at 1.4 mM) were added as follows: 0.4 μL over 0.8 s for the first injection, followed by 19 injections of 2 μL over 4 at 120 s intervals. For the experiment with VirA ACP$_{5b}$-^{C}DD, the construct (900 μM) was added to VirFG ^{N}DD (60 μM), as described above. The heat of reaction per injection (microcalories/second) was determined by integration of the peak areas using either the Origin (MicroCal) software or Nitpic.[28] These software provide the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N), and the dissociation constant (K_d). The heats of dilution were determined by injecting VirA ^{C}DD and VirFG ^{N}DD alone into the buffer (initial concentrations of 600 μM) and were subtracted from the corresponding experiments before curve fitting. This experiment additionally showed that VirA ^{C}DD was undergoing homodimerization as the concentration in the calorimeter cell increased because a change in the heat signal was detected up to a concentration of 30 μM. No such change was observed over the whole concentration range for VirFG ^{N}DD (up to 78 μM).

NMR Data Acquisition. To minimize the amount of protein needed, 300 μL of protein solution were loaded into a 4 mm NMR tube. ^{15}N−^{1}H HSQC spectra of VirA ^{C}DD were recorded at a concentration of 100 μM in the absence of partner and in the presence of 200−400 μM ^{14}N VirFG ^{N}DD. ^{15}N−^{1}H HSQC spectra of VirFG ^{N}DD were recorded at a concentration of 100 μM either alone or in the presence of 400 μM ^{14}N VirA ^{C}DD or 200 μM ^{14}N VirA ACP$_{5b}$-^{C}DD. Heteronuclear spectra were acquired on ^{13}C,^{15}N-VirA ^{C}DD at a concentration of 1.2 mM in the presence of a 2-fold excess of ^{12}C,^{14}N-VirFG and on ^{13}C,^{15}N-VirA ^{C}DD−VirFG ^{N}DD at a concentration of 0.9 mM. All NMR data were recorded at 17 °C on a Bruker DRX600 spectrometer equipped with a cryogenic probe (Plateforme de Biophysicochimie Biologie Structurale−FR 3209 Université de Lorraine/CNRS).

Backbone and sequential resonance assignments were obtained by the combined use of 2D ^{15}N−^{1}H and ^{13}C−^{1}H HSQC spectra and 3D HNCA, HNCACB, CBCA(CO)NH, HNHA, HN(CA)CO, and HNCO experiments. Assignments of aliphatic side chain resonance were based on 2D aromatic ^{13}C−^{1}H HSQC, 3D (H)CC(CO)NH, H(CC) (CO)NH, HCCH−COSY, and HCCH-TOCSY experiments (reviewed in ref 29). To collect NOE-based distance restraints for the VirA ^{C}DD−VirFG ^{N}DD structure calculations, 3D ^{15}N NOESY-HSQC and ^{13}C NOESY-HSQC were recorded on uniformly ^{13}C,^{15}N-enriched VirA ^{C}DD−VirFG ^{N}DD 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).

NMR Structure Calculation. Initial structures were generated using CYANA 3.0 software.[30] 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.[31] In each cycle, 100 structures starting from random torsion angle values were calculated with 10 000 steps of torsion angle dynamics-driven simulated annealing. A total of 478 NOE-based distance and 54 backbone angle restraints were used for structure calculation of the VirA ^{C}DD domain, whereas 1372 NOE-based distance and 98 backbone angle restraints were used for VirA ^{C}DD−VirFG ^{N}DD. The angle restraints were obtained from ^{13}Cα, ^{13}Cβ, ^{13}C′, ^{15}N, ^{1}HN, and ^{1}Hα chemical shifts using TALOS+[32] with an assigned minimum range of ±20°. 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 restrained molecular dynamic (rMD) simulations in Amber 14 (refs 33 and 34) using the generalized Born solvent model. Possible atoms clashes within CYANA initial structures were regularized by a 1 ps energy minimization in the Amber force field, and then 20 ps of rMD 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 VirA ^{C}DD−VirFG ^{N}DD representative

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

ensembles correspond to the 20 conformers from each calculation with the lowest restraint energy terms.

Analysis by Circular Dichroism. Circular dichroism (CD) was carried out on a Chirascan CD from Applied Photophysics. Data were collected at 0.5 nm intervals in the wavelength range of 180−260 nm at 20 °C, using a temperature-controlled chamber. A 0.01 cm cuvette containing 30 μL of protein sample at 100 μM was used for all the measurements except that of 10 μM VirA ^{C}DD (100 μL of sample were used in a 0.1 cm cuvette). Measurements were made at least in triplicate, and sample spectra were corrected for buffer background by subtracting the average spectrum of buffer alone. To evaluate the extent of induced folding when the two docking domains were together, we acquired a combined spectrum of the two docking domains by placing VirA ^{C}DD into one cuvette and VirFG ^{N}DD into another cuvette (30 μL each at 100 μM) and analyzing the two cuvettes simultaneously. The same result was obtained by averaging the spectra of VirA ^{C}DD and VirFG ^{N}DD acquired separately. We then combined the two docking domains, and placed the mixture into both cuvettes (30 μL at 100 μM), again analyzing the two simultaneously. Spectrum deconvolution was carried out using the CDNN software.[35]

Small-Angle X-ray Scattering Data Collection. SAXS data were collected on the SWING beamline at the Synchrotron SOLEIL (France). Frames were recorded using an AVIEX170170 CCD detector at the energy of 12 keV. The distance between the sample and the detector was set to 1803 mm, leading to scattering vectors q ranging from 0.0005 to 0.5 Å^{-1}. The scattering vector is defined as $4\pi/\lambda \sin\theta$, where 2θ is the scattering angle. The protein samples were injected using the online automatic sample changer into a pre-equilibrated HPLC-coupled size-exclusion chromatography column (Bio-SEC 100, Bio-SEC 150 or Bio-SEC 300, Agilent) depending on the molecular weight of the protein sample, at a temperature of 15 °C. The online purification system that delivers the eluted fractions into the measurement cell was developed on the SWING beamline.[36] After equilibrating the column in the protein buffer (50 mM sodium phosphate (pH 7.5), 150 mM NaCl, 5% glycerol (as radioprotectant)), 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 maintain the catalytic activity of *cis*-AT PKSs,[37,38] presumably by stabilizing the structure of the proteins. A 50 μL aliquot of the protein sample (at 10 or 15 mg/mL) was then injected, and complete data sets were acquired. The protein concentration downstream of the elution column was followed via the absorbance at 280 nm with an *in situ* spectrophotometer. In contrast to classical SAXS experiments that 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 because many different positions within the elution peak are sampled during the course of the measurement (typically 50−100 frames are acquired).[36] Data reduction to absolute units, frame averaging, and subtraction were performed using FOXTROT, a dedicated in-house application. Each acquisition frame of the experiment yielded a scattering spectrum, which was then analyzed using FOXTROT, giving 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 $I(q) = I(0) \exp(-q^2R_g^2/3)$ 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. (See Supporting Information Methods for a model data set.) In the following step of data analysis, all the frames that exhibited the same R_g as a function of $I(0)$ were corrected for buffer signal and averaged, thus ensuring that the data reflected only the signal arising from the protein structure and not from intermolecular interactions. 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.[39]

Determination of Construct Molecular Weights. Classically, molecular weights can be derived from the SAXS data using the $I(0)$ and the known protein concentration. Although in our experimental setup the protein concentration is known when the peak elutes from the gel-filtration column, the fact that there is a delay between exiting the column and the acquisition of the SAXS data means that the concentration could have changed prior to the measurement. Therefore, to determine the molecular weights of our constructs, we relied on two alternative methods: SAXS MoW,[40] which relies on integration of the distance distribution function (because this is known to be proportional to molecular weight), and size-exclusion chromatography multiangle laser light scattering (SEC-MALLS), for which the intensity of light scattered off protein molecules in solution is proportional to their concentration and molecular weight.[41]

***Ab Initio* Shape Determination.** *Ab initio* protein shapes were calculated from the experimental SAXS data using the bead-modeling program DAMMIN.[42] At least 20 independent fits were carried out without imposition of symmetry for the monomeric VirFG ^{N}DD, the VirA ^{C}DD/VirFG ^{N}DD and VirA ACP$_{Sb}$-^{C}DD/VirFG ^{N}DD non-covalent complexes, and the VirA ^{C}DD−VirFG ^{N}DD and VirA ACP$_{Sb}$-^{C}DD−VirFG ^{N}DD covalent fusions and with a 2-fold symmetry restraint for the VirA ACP$_{Sb}$-^{C}DD−VirFG ^{N}DD-KS$_6$ fusion protein. (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 SUPCOMB[43] in order to determine a consensus model. Low-resolution models were then averaged using the DAMAVER and DAMFILT packages.[44,45] The quality of the 3D modeling was determined using the discrepancy χ^2, defined according to Konarev et al.[44] Values lying in the range of 0.9−1.1 are accepted to indicate a good fit between the models and the data. However, the calculation of χ^2 is inversely proportional to the measurement error, that is, the lower the error in the measurement, the higher the χ^2. Because the detector at SOLEIL yields inherently low-error, high-quality data, the effect is to raise the determined χ^2 values above this standard range.[46] Nonetheless, visual inspection of the agreement between the theoretical curves calculated from the average molecular forms and the acquired SAXS data confirms the goodness of fit in all cases. Therefore, we decided not to rescale artificially the error on the measurement to give χ^2 of approximately 1, as has been suggested.[46]

High-resolution structures or models were then superimposed into the *ab initio* protein shapes using SUPCOMB.[43] Figures displaying the protein structures were generated using PYMOL (Schrödinger, LLC), and the low-resolution bead models from DAMMIN were converted into mesh.

Molecular Modeling. A 3D structural model of the KS$_6$-linker region was generated by pairwise comparison of profile hidden Markov models using the HHpred server (http://toolkit.tuebingen.mpg.de/hhpred).[47] Via this program, local sequence alignment using PSI-BLAST was performed, and a multiple sequence alignment including information about predicted secondary structures was produced. The closest homologue to the VirA KS$_6$-linker within the Protein Data Bank is KS$_2$ of PksJ (PDB: 2NA1) (35% sequence identity; 63% similarity).[48] This structure was then used to construct a model of the KS$_6$-linker using the associated MODELER software.[49] The KS$_6$ homodimer was reconstituted by superimposing two models onto the dimeric crystal structures of the PksJ KS.

Demonstration by SAXS That VirFG ^{N}DD Undergoes a Conformational Change. *Ab initio* molecular forms for VirFG ^{N}DD were calculated using DAMMIN.[42] Twenty of the resulting low-resolution models (χ^2 of 2.321−2.489 vs the SAXS data) were compared using SUPCOMB,[43] revealing an extended shape that is not consistent with the conformation of VirFG ^{N}DD adopted in the VirA ^{C}DD−VirFG ^{N}DD fusion. Indeed, calculation of theoretical scattering curves for VirFG ^{N}DD alone using CRYSOL[50] based on the NMR structure of the VirA ^{C}DD−VirFG ^{N}DD fusion (all 20 structures within the ensemble were used) uniformly give poor fits to the

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

experimental data for scattering angles between 0.0005 and 0.3 Å^{-1} (χ^2 = 6.61−10.39). Furthermore, the Kratky plot[51] computed from the experimental SAXS curve is consistent with an elongated shape/flexible particle. Taken together, these calculations clearly show that free VirFG ^{N}DD and VirFG ^{N}DD in complex with partner do not adopt the same structure, consistent with an induced conformational change.

Accession Numbers. Chemical shifts for VirA ^{C}DD in complex with VirFG ^{N}DD and for VirA ^{C}DD−VirFG ^{N}DD have been deposited in the Biological Magnetic Resonance Bank with accession codes 25721 and 25706, respectively. PDB coordinates for VirA ^{C}DD−VirFG ^{N}DD have been deposited under accession code 2N5D.

RESULTS

In Silico Analysis of Docking in *trans*-AT PKSs. In earlier work, we identified VirA ^{C}DD as comprising the 50 residues C-terminal of the conserved ACP$_{5b}$ domain of module 5 (Figure 1).[21] Similarly, inspection of the N-terminus of VirFG revealed 70 amino acids N-terminal to the conserved KS domain as a candidate docking domain partner (Figure S1). Although BLAST[52] analysis of this region returned no homologues within the PDB, the first 28 residues of VirFG ^{N}DD show low sequence homology to VirA ^{C}DD (28% identity[53]) and to multiple Class 2 docking domains (e.g., 4−21% to the six reported docking domains of the curacin PKS;[17] Figure S1). The C-terminal 42 amino acids, however, are not conserved, suggesting no involvement in docking. These observations prompted us to consider whether homologous sequences were present at intersubunit junctions in additional *trans*-AT PKSs, allowing putative ^{C}DD and ^{N}DD families to be identified in these systems for the first time.

In *cis*-AT PKSs, interpolypeptide junctions are formed between partner docking domains with additional contacts contributed by the flanking C-terminal ACP and N-terminal KS domains.[6,15−17,54] Our analysis of 44 published *trans*-AT PKSs reveals that such interfaces are formed by a much greater diversity of domain pairs (Table S2). It is also striking that in the majority of cases the domains comprising the junction are in fact members of a single chain-extension module (i.e., the functional module is split between two different multienzyme polypeptides). Nonetheless, in a number of cases, the intersubunit interfaces are formed by the ends of modules. For a subset of such interfaces, analysis of the sequences directly downstream from the conserved ends of the ACP domains and those directly upstream from the conserved start sites of the N-terminal domains (KS, ketoreductase (KR), ER, or C/heterocyclization (HC)) revealed regions with low but suggestive homology to VirA ^{C}DD (10−53% identity; avg. 22%)[53] and VirFG ^{N}DD (8−39% identity; avg. 23%),[53] respectively. All of the regions were predicted by secondary structure analysis (PsiPred[23] and PROMALS3D)[24] to contain two α helices or, more rarely, one long α helix (Figure S2).

The 27 identified junctions (Table S3) include the ACP/KS interfaces present in the anthracimycin, chivosazol, elansolid, enacyloxin, kirromycin, oxazolomycin, rhizoxin, and thailandamide systems (four of these junctions are of the specific type tandem ACP-ACP/KS as in the Vir system), the ACP/C or ACP/HC (PKS/NRPS) interfaces within the bacillaene, chivosazol, disorazol, kalimanticin/batumin, kirromycin, oxazolomycin, rhizopodin, and thailandamide assembly lines, an ACP/KR interface within the macrolactin system, and ACP/ER junctions in the bryostatin, etnangien, and myxovirescin machineries. *Streptomyces pristinaespiralis* also harbors a hybrid *trans*-AT PKS-NPRS that generates virginiamycin (though the product is referred to as pristinamycin IIA), and as expected, sequences with homology to VirA ^{C}DD and VirFG ^{N}DD were present on the respective subunits SnaE2 and SnaE3.[55] Nonetheless, the ^{N}DD equivalent of SnaE3 was N-terminally truncated, suggesting incorrect identification of the *snaE3* site start. Indeed, inspection of the genomic region of *S. pristinaespiralis* (GenBank: FR682001) revealed a putative alternative start codon 147 base pairs 5′ of the published start site (new start site at base pair 65 054); translation of this longer region gave a sequence incorporating a complete N-terminal docking domain (Figure S2). Conversely, the published N-terminus of TaiM from the thailandamide PKS[56] begins some 300 residues upstream of a Met and the first conserved α helix of the putative docking domain, again suggesting an incorrectly assigned start site. In contrast, the TaiE subunit from the same PKS also contains a long extension N-terminal to the docking domain, but its presence cannot apparently be explained by a misassigned start codon.

VirA ^{C}DD and VirFG ^{N}DD Serve as Docking Domains. Having identified VirA ^{C}DD and VirFG ^{N}DD as representative of their respective families of C- and N-terminal docking domains, we aimed to investigate their interaction. For this, we expressed and purified (Figure S3) the following constructs in recombinant form from *E. coli* BL21(DE3) as His$_6$-SUMO-tagged proteins using plasmid pBG-102 (Vanderbilt University Center for Structural Biology): VirA ^{C}DD, VirFG ^{N}DD, VirA ACP$_{5b}$-^{C}DD, VirFG ^{N}DD-linker (all 70 residues up to the conserved KS$_6$), VirFG ^{N}DD-KS$_6$, and a protein (VirA ^{C}DD−VirFG ^{N}DD) in which the two docking domains were covalently fused via the same eight-amino-acid linker (G_3SG_3S) used previously to study the Class 2 docking domains of *cis*-AT PKSs[17] (Supporting Information Methods). For the sequences of all constructs, see Table S4. To allow both efficient monitoring of VirA ^{C}DD and VirFG ^{N}DD by UV−vis during purification and dosing for binding studies by ITC, we added a tyrosine residue at the N- and C-terminal ends, respectively; subsequent structural analysis (*vide infra*) showed that these additions had no effect on the native docking interaction involving the opposite termini. Following cleavage of the His$_6$-SUMO-tag using human rhinovirus 3C (H3C) protease, the N-termini of all of the proteins also harbored a non-native GPGS sequence; this sequence is highly similar to and occupies the same position relative to VirFG ^{N}DD as the G_3SG_3S sequence used to fuse the domains covalently, which we also show by NMR (*vide infra*) to have no effect on the docking interaction while in the case of VirA ^{C}DD, it lies upstream of the added Y. In the case of VirFG ^{N}DD-KS$_6$, we designed the expression construct to include 179 residues downstream of the KS$_6$ domain (encompassing consecutive regions showing homology to the so-called KS-AT and post-AT linkers observed in structures of KS-AT didomains from *cis*-AT PKSs)[18] because this choice of boundary previously allowed us to obtain the KS of module 5 as a stable protein.[21]

As an initial demonstration that the docking domains interact, we carried out analytical gel filtration in the presence of VirA ^{C}DD, VirFG ^{N}DD, and a 1:1 mixture of the two domains (all at an initial concentration of 670 μM). On their own, VirA ^{C}DD and VirFG ^{N}DD gave elution volumes of ca. 13.7 and 14.5 mL, respectively, whereas the 1:1 mixture gave a single peak with a new elution volume of 14 mL (Figure S4), providing the first evidence for a specific interaction between the two docking domains. Conversion into molecular weights yielded the following data: VirA ^{C}DD, 11.8 kDa (monomer

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

weight 4.3 kDa); VirFG ^{N}DD 7.4 kDa (monomer weight 4.1 kDa); and the VirA ^{C}DD/VirFG ^{N}DD complex, 9.9 kDa (heterodimeric complex weight, 8.4 kDa). These data suggested that both domains are oligomeric when isolated but recombine to form a 1:1 complex; in fact, VirFG ^{N}DD is monomeric but elutes anomalously because of its extended conformation (*vide infra*). Measurement of the affinity of the interaction by ITC yielded a value of 5.8 ± 0.2 μM (from three independent analyses, Figure S5), which is within the range measured for Class 1 and 2 docking domains (4−20 μM);[16,17] a similar K_d (4 μM) was measured for the VirA ACP$_{5b}$-^{C}DD/VirFG ^{N}DD pair (Figure S5). Inclusion of the 42-residue linker downstream of VirFG ^{N}DD did not significantly alter the affinity (K_d = 5.5 μM for the interaction between VirA ^{C}DD and VirFG ^{N}DD-linker, Figure S5), supporting our supposition that this region is not involved in docking.

Characterization of Docking between VirA ^{C}DD and VirFG ^{N}DD by NMR, SAXS, and Allied Methods. To investigate in detail the interaction between the two docking domain partners, we analyzed ^{15}N-labeled VirA ^{C}DD in isolation and in the presence of unlabeled (^{14}N) VirFG ^{N}DD so that in either case only signals arising from VirA ^{C}DD would be visible in the obtained spectra. In the absence of its partner docking domain, the ^{15}N−^{1}H HSQC spectrum of VirA ^{C}DD (100 μM) exhibited only 15 of 37 expected peaks (Figure S6). In contrast, when a 2- to 4-fold excess of ^{14}N VirFG ^{N}DD was also present, 36 of 37 peaks were observed (Figure S6). Two explanations account for the fewer than expected number of peaks in the spectrum of unpartnered VirA ^{C}DD: (i) an equilibrium between folded and unfolded states, on the intermediate NMR time-scale, or (ii) a mixed population of nonspecific homodimers (a state observed previously),[21] which also undergo exchange on the intermediate time-scale. (The set of residues at the interfaces are those which disappear because they experience multiple chemical environments.) We strongly favor the second explanation because data obtained by ITC as the concentration of VirA ^{C}DD was increased incrementally are consistent with a net monomer to dimer transition, with the protein dimerizing completely above 30 μM (Figure S5). Furthermore, molecular weight estimation by SAXS MoW following SAXS analysis of VirA ^{C}DD (Figure S7 and Table S5) yielded a molecular weight of 8.4 kDa (this procedure typically produces measurement errors of <10%;[40] homodimeric complex theoretical weight 8.6 kDa), and the result was confirmed by SEC-MALLS, which gave a molecular weight of 8.2 ± 0.6 kDa. Using this information, we calculated average molecular forms for VirA ^{C}DD in solution, with and without imposition of second-order symmetry. In all cases, the fits between the theoretical curves calculated from the forms and the SAXS data (for the q range 0.018 and 0.3 Å^{-1}) were good, but the χ^2 values were lower in the absence of imposed symmetry (1.732−2.543 vs 1.889−3.28, Figure S7). The overall heterogeneity of these shapes supports the idea that VirA ^{C}DD in solution is best described as a population of interchanging, nonspecific dimers. To investigate whether VirA ^{C}DD retains its structure even in the monomeric form, we analyzed it by CD at concentrations both above (100 μM) and below (10 μM) the estimated dimerization point. This experiment showed that although autoassociation did increase the overall secondary structure, VirA ^{C}DD retained 73% of its α-helical content in monomeric form (Figure S8).

Although isolated VirA ^{C}DD clearly homodimerizes, in the presence of excess VirFG ^{N}DD it undergoes subunit exchange to form a stable, heterodimeric docked complex, explaining the observation of almost all anticipated NMR peaks. This fact allowed us to fully assign ^{15}N,^{13}C-VirA ^{C}DD within the context of this complex using a combination of 2D and 3D NMR experiments in order to determine its secondary structure (Figure S6). Consistent with the secondary structure predictions, bound VirA ^{C}DD comprises two short α helices (residues Q6936−R6947 and I6953−L6960; numbering based on GenBank: BAF50727).

We next carried out these experiments in reverse, with ^{15}N-labeled VirFG ^{N}DD and unlabeled (^{14}N) VirA ^{C}DD. The ^{15}N−^{1}H HSQC spectrum of isolated VirFG ^{N}DD (100 μM) showed the expected number of peaks, but with poor chemical shift dispersion (Figure S6). Dynamical studies also showed it to be highly flexible (Figure S6). Taken together, these results indicated that VirFG ^{N}DD alone (i.e., in the absence of its docking partner or a downstream catalytic domain) lacks a well-defined 3D structure and so could be an intrinsically disordered protein (IDP).[57] (For a discussion of the applicability of these findings to other *trans*-AT PKS, see the Appendix). In agreement with this, CD analysis yielded data consistent with only partial structuration (22% α helix, Figure S8), whereas analysis by ITC under the same conditions as for VirA ^{C}DD provided no evidence for homodimerization (Figure S5). As further support for the IDP character of VirFG ^{N}DD, addition of 2,2,2-trifluoroethanol (TFE) to the buffer significantly increased the α-helical content as measured by CD (Figure S8). This reagent mimics the hydrophobic environment experienced by proteins in protein−protein interactions and so is widely exploited to reveal IDPs with a propensity to undergo induced folding.[58] We also analyzed VirFG ^{N}DD by SAXS under the NMR conditions, yielding an R_g of 16 ± 1 Å, a D_{max} of 66 Å, and M_ws of 5 kDa (SAXS MoW) and 4.2 ± 0.4 Da (SEC-MALLS), consistent with a monomer (Figure S7 and Table S5). The D_{max} value accords with an extended, albeit partially folded protein (predicted D_{max} for a fully extended protein of 40 residues =160 Å, Figure S7); this result also explains the anomalous behavior by analytical GF, in which VirFG ^{N}DD appeared to elute as a dimer because of its extended shape. The form of a Kratky plot[51] of the obtained data also confirmed that the protein is not globular (i.e., that it does not fold stably) but showed that it is nonetheless likely to contain residual α helix (Figure S7). However, despite extensive attempts, we were unable to obtain a recombinant protein in which VirFG ^{N}DD was attached to the downstream KS domain (VirFG ^{N}DD-KS$_6$) because the construct aggregated in solution. Thus, we cannot strictly rule out the possibility that VirFG ^{N}DD is structured in its native context even in the absence of VirA ^{C}DD, although clearly it shows multiple characteristics of an IDP.

In any case, addition of a 4-fold excess of ^{14}N-VirA ^{C}DD (ca. 400 μM) provoked a number of spectral modifications, including significant changes in chemical shift and an overall increase in chemical shift dispersion (Figure S6). These types of spectral changes are consistent with a coupled process of binding/induced structuration of VirFG ^{N}DD by VirA ^{C}DD. To bolster the case that VirFG ^{N}DD was gaining structure upon contact with VirA ^{C}DD, we carried out differential CD analysis, acquiring spectra of VirFG ^{N}DD and VirA ^{C}DD alone and in complex (Figure S8). This analysis showed that the spectrum of the complex does not correspond to a simple average of the spectra of the two individual partners (α-Helical content increased from 40 to 52%).[35] VirA ^{C}DD is already at least

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

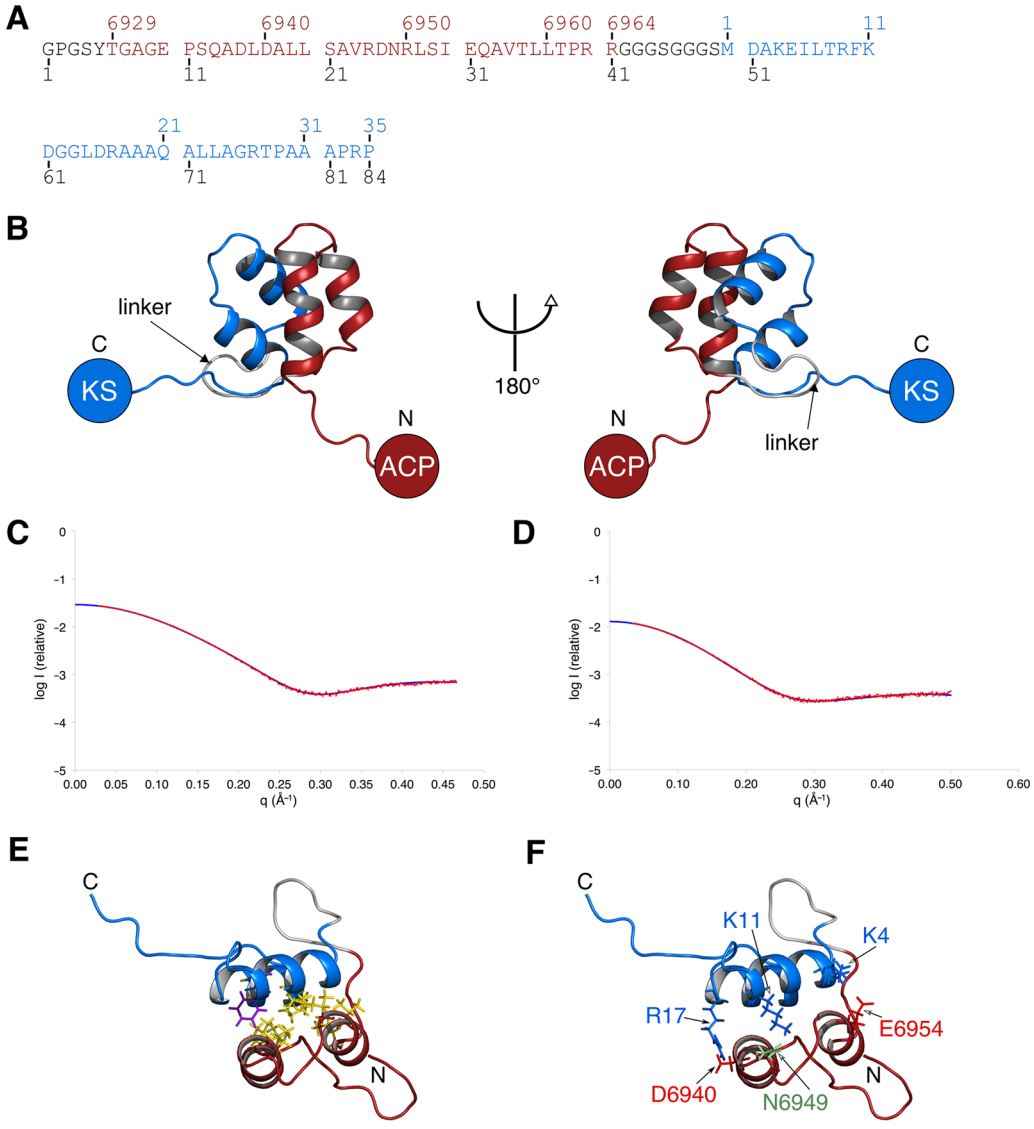


Figure 3. Analysis of docking between VirA ^{C}DD and VirFG ^{N}DD. (A) Sequence of the VirA ^{C}DD–VirFG ^{N}DD fusion construct, with VirA ^{C}DD indicated in red, and VirFG ^{N}DD indicated in blue. The N-terminal vector-derived residues and the added 8-residue linker are shown in black. The numbering above the sequence corresponds to that in GenBank, whereas that below is based on the construct itself. (B) NMR structure of the covalent docking domain complex VirA ^{C}DD–VirFG ^{N}DD color-coded as in A. (C and D) The strong correspondence between the SAXS curve for VirA ^{C}DD–VirFG ^{N}DD predicted from its NMR structure (CRYSOL[50]) (blue) and those obtained (red) for (C) the noncovalent VirA ^{C}DD/VirFG ^{N}DD complex and (D) VirA ^{C}DD–VirFG ^{N}DD supports not only the validity of the NMR structure but also the fact that covalently fusing the two DDs does not alter their interaction. (E) Hydrophobic residues at the docking interface within VirA ^{C}DD–VirFG ^{N}DD. (F) Electrostatic interactions at the docking interface within VirA ^{C}DD–VirFG ^{N}DD that were targeted for site-directed mutagenesis.

partially structured at the concentration used in this experiment because of self-association; thus, we attribute the structuration seen during complex formation, at least in part, to induced folding of VirFG ^{N}DD, although we cannot exclude that the secondary structure content of VirA ^{C}DD also increases. An essentially identical ^{15}N–^{1}H HSQC spectrum of ^{15}N-labeled VirFG ^{N}DD was obtained in the presence of VirA ACP$_{5b}$-^{C}DD (Figure S6), consistent with the result from ITC showing that the presence of the ACP does not impact the docking domain interaction.

We followed up these observations by elucidating the complete structure of the docking domain complex, using the construct VirA ^{C}DD–VirFG ^{N}DD in which the two docking elements were covalently fused.[17] (For the sequence of this protein, see Figure 3A.) Structural studies of ^{15}N,^{13}C-labeled VirA ^{C}DD–VirFG ^{N}DD (Figures 3 and S6) were carried out by multidimensional heteronuclear NMR spectroscopy. Structure calculation was performed using a two-step procedure consisting of initial structure generation using CYANA[30] followed by restrained molecular dynamics refinement within Amber.[33,34] The quality of the structures was assessed using PROCHECK-NMR[59] and the MolProbity[60] server (Table S6). The first important result is that the introduced linker between the two docking domains is unstructured in the complex (Figure 3B). Consistent with this, amide chemical shifts arising from this region are clustered together in the ^{15}N–^{1}H HSQC spectrum, and dynamical studies show it to be highly mobile (Figure S6). Taken together, these data confirm that the linker acts simply as a flexible tether and does not contribute to the docking interface, supporting the idea that our fusion strategy has faithfully captured the native interaction.

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155–4167

Within the context of this fusion construct, VirA ^{C}DD comprises two consecutive, short α helices as found for discrete VirA ^{C}DD in complex with VirFG ^{N}DD. Likewise, fused VirFG ^{N}DD incorporates two short α helices (residues A3−K11 and R17−L24, numbered according to our revised sequence of VirFG,[21] GenBank: KT308172), and its NMR spectrum superimposes on that of the domain in noncovalent complex with VirA ^{C}DD (Figure S6). The introduced Y residues at the extremities of VirA ^{C}DD and VirFG ^{N}DD both lie within unstructured linkers at a significant distance from these key secondary structure elements (7 and 11 residues, respectively, for VirA ^{C}DD and VirFG ^{N}DD) and do not contact the docked complex, confirming that they have not affected the interaction. Together, the four α helices form a four α-helical bundle in which the two sets of α helices are set at an angle of approximately 127° with respect to each other. Thus, unlike the Class 1 and 2 docking domains in which the upstream and downstream subunits each contribute two copies of the C- and N-terminal docking domains to a single docking complex,[15−17] the interaction between VirA and VirFG likely involves formation of two distinct docking complexes, each of which comprises one C- and one N-terminal docking domain.

Analysis by SAXS MoW and SEC-MALLS confirmed the heterodimeric nature of the covalent complex (calcd M_w 8.4 kDa; heterodimeric complex weight, 9 kDa via SAXS MoW; and 8.9 ± 0.9 kDa, via SEC-MALLS; Table S5). The docking interaction in *trans*-AT PKSs is therefore fundamentally different from that in *cis*-AT PKSs. It is also notable that a search of the PDB for homologues using both VAST[61] and DALI[62] did not reveal a comparable topology for any known interprotein complexes, consistent with previous findings that IDPs characteristically form unusual interactions.[63] Independent evidence for the structure was obtained by SAXS because the curve predicted using CRYSOL[50] from the NMR structure of the VirA ^{C}DD−VirFG ^{N}DD fusion protein superimposed nearly perfectly with the SAXS data obtained on this construct (χ^2 = 2.90; Figures 3C,D and S7 and Table S5). Furthermore, the SAXS curves predicted by CRYSOL for VirFG ^{N}DD alone if it adopted the same structure as in the complex (i.e., those arising from all 20 structures within the NMR ensemble) clearly do not correspond to the experimental data obtained on the isolated protein (χ^2 = 6.61−10.39), proving that VirFG ^{N}DD undergoes a structural transition in the presence of its partner VirA ^{C}DD. (See Figure S7 for a representative fit.) Within the complex, the interface is ∼730 Å^2, an extensive surface area given the limited number of participating residues (17 amino acids within a 4 Å radius). This too is consistent with VirFG ^{N}DD being an IDP, as such regions are typically able to achieve a greater interface area/residue than comparable ordered proteins.[64] If two such complexes form simultaneously *in vivo* at the VirA/VirFG junction, then the total interaction surface area would be ∼1460 Å^2, which is in between Class 1 (∼1100 Å^2) and Class 2 (∼1800 Å^2) docking domains and in accord with the affinity measured by ITC. The docking interface is hydrophobic, and comprises residues L6939, L6943, V6946, I6953, A6956, V6957, and L6960 of VirA ^{C}DD and residues A3, L7, F10, A20, and L24 of VirFG ^{N}DD (Figure 3E,F). The contact surface between the two VirA^{C}DD α helices includes L6942, A6945, L6951, A6956, and L6959, whereas that between the two VirFG ^{N}DD α helices comprises I6, L15, A19, and L23. Inspection of the sequence alignments of the *trans*-AT PKS C- and N-terminal docking domains (Figure S2 and Table S7) shows that hydrophobic residues are relatively well-conserved at the corresponding positions. Three electrostatic interactions are also present at the interface (residues (VirA ^{C}DD/VirFG ^{N}DD): D6940/R17, N6949/K11, and E6954/K4).

Analysis of Protein−Protein Interaction Determinants by Site-Directed Mutagenesis and ITC. Because our docking model derives from the structure of the covalent complex of the two partners, we aimed to test its applicability to the discrete domains. For this, we mutated both polar and hydrophobic residues lying at the interface of the complex in order to evaluate their effects on the interaction affinity by ITC. Each mutant was initially analyzed by circular dichroism (CD) to monitor for structural changes induced by the mutations because these would also have an effect on measured K_d values. In fact, the CD experiments revealed that VirFG ^{N}DD underwent architectural alterations when mutated, even gaining increasing secondary structure (Figure S8). This observation is again consistent with VirFG ^{N}DD being an IDP and with its disordered character needing to be actively maintained against mutations that can increase structural content.[65] However, this behavior complicated interpretation of ITC data obtained using mutant VirFG ^{N}DD because we could not clearly distinguish between mutations which disrupted an important interaction with VirA ^{C}DD and those which simply impacted the ability of VirFG ^{N}DD to undergo induced structuration. We therefore focused our attention on the VirA ^{C}DD variants. Specifically, we analyzed the effect of exchanging D6940, N6949, and E6954 for alanine, as well as mutations which reversed the charge on residues D6940 and E6954 (changed to R and K respectively) and introduced charge at the N6949 position. We were unable to purify N6949A as well as an additional L6942A mutant because of poor solubility, but all other mutant proteins were obtained in acceptable yield. Far-UV CD analysis confirmed that their structures were not perturbed significantly relative to wild-type VirA ^{C}DD (Figure S8).

Because of solubility limitations, we could not reach binding saturation with many of the mutants, so the binding affinities determined by ITC are only rough estimates. (The stated errors represent those from fitting of a single experiment, Figure S5.) Nevertheless, the effect of the mutation in each case was clear. The D6940A and E6954A charge-removal mutations decreased binding affinity relative to that of wild type (K_d = 5.8 μM) by factors of ca. 15 (85 ± 23 μM) and 4-fold (20 ± 2 μM), respectively. The effect of changing the charge state of the residues was equally significant, with the affinity of VirFG ^{N}DD binding to the VirA ^{C}DD N6949K and E6954K mutants reduced to 37 ± 4 and 49 ± 16 μM, respectively. Even more dramatically, interaction between the D6940R mutant and VirFG ^{N}DD was completely abolished. On the basis of these results, we decided to test the effect of a combined mutation, introducing D6940A and E6954A simultaneously into VirA ^{C}DD. Analysis of the double mutant by CD showed that the structure was essentially unchanged, and the measured affinity to wild-type VirFG ^{N}DD was >160 μM, demonstrating that the effect of the mutations was cumulative.

Taken together, our data clearly show that electrostatic interactions contribute to the assembly of the VirA ^{C}DD/VirFG ^{N}DD complex. Inspection of the residue positions corresponding to D6940, N6949, and E6954 in the other docking domains of the putative family shows that nonpolar residues are also present in a number of systems and that even when the residues are charged they are often but not systematically matched by an opposite charge on the partner docking domain (Figure S2 and

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

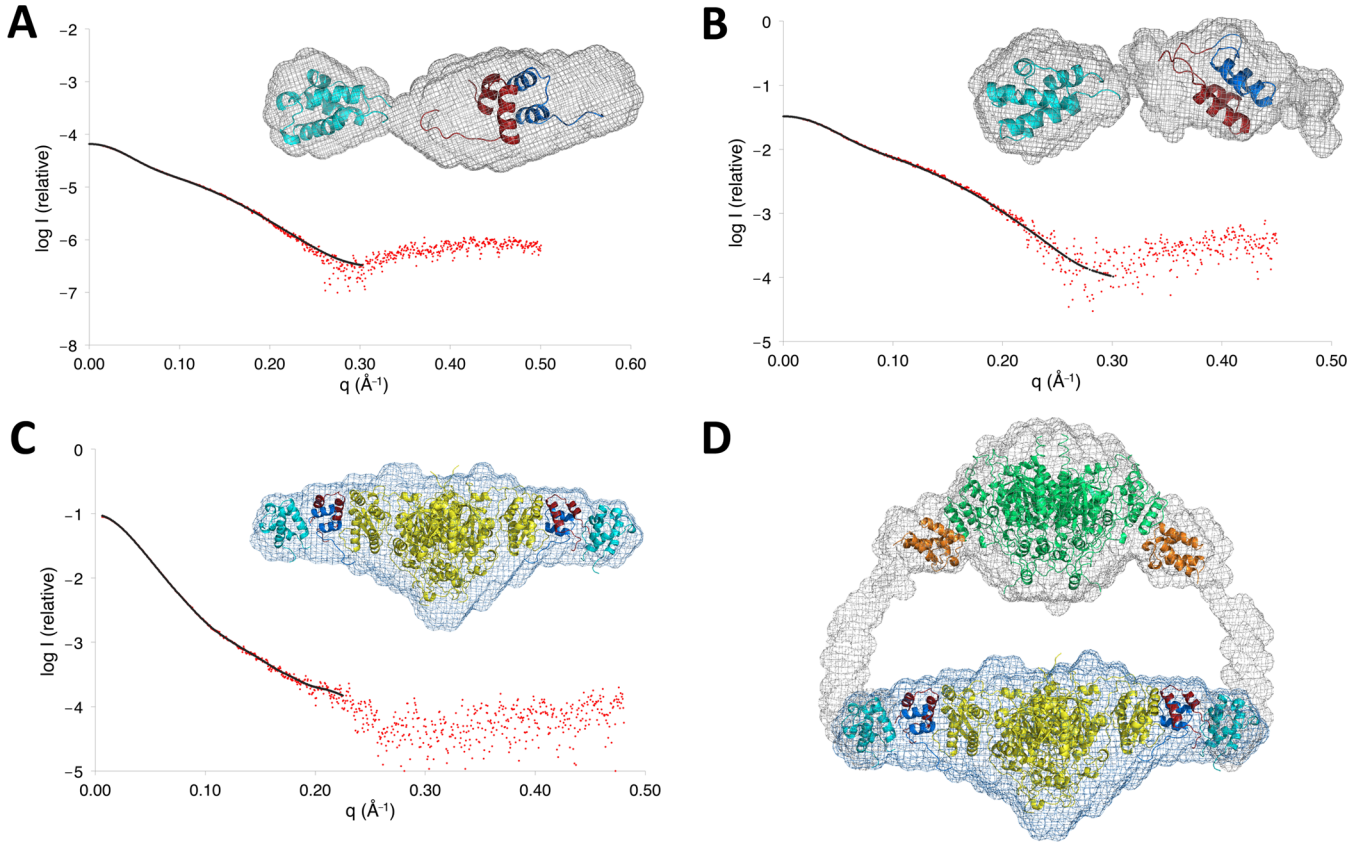


Figure 4. Analysis of the interface between subunits VirA and VirFG by SAXS. (A) Analysis by SAXS of the VirA ACP_{5b}-^{C}DD/VirFG ^{N}DD complex. Fit between the *ab initio* model computed with DAMMIN[42] (solid black line) and the experimental SAXS data (red dots). This analysis yielded a molecular envelope different from that of VirA ACP_{5b}-^{C}DD alone[21] because density was only observed for one ACP domain (light blue) and the docking domain complex (VirA ^{C}DD in red and VirFG ^{N}DD in dark blue). This result confirmed that docking occurs between a single copy each of the C- and N-terminal docking domains and that ACP_{5b} does not directly contact the complex of DDs. The presence of significant empty density likely reflects the fact that several species are present when the docking domains form a noncovalent complex. (B) Analysis of VirA ACP_{5b}-^{C}DD–VirFG ^{N}DD by SAXS, showing the fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots). Comparison with the average molecular form shown in A demonstrates that the covalent linkage of the two docking domains leads to a more homogeneous sample. (C) Analysis by SAXS of a covalent complex between VirA ACP_{5b}-^{C}DD and VirFG ^{N}DD-KS_6 (VirA ACP_{5b}-^{C}DD–VirFG ^{N}DD-KS_6). Fit between the *ab initio* model computed with DAMMIN (solid black line) and the experimental SAXS data (red dots). A homology model of KS_6 (yellow) was placed into the average molecular envelope using SUPCOMB,[43] whereas localization of the NMR structure of ACP_{5b}[21] was based on the assumption that the domains would be situated on the extremities of the structure. The remaining electronic density could then be assigned to the docking domain complex. Comparison with C indicates that further structuration occurs in the presence of both flanking functional domains. (D) Model for the complete interface between subunits VirA and VirFG generated by superimposing the SAXS-derived model of the KS_5-ACP_{5a}-ACP_{5b} region of VirA module 5 (ref 21) (KS_5 in green, ACP_{5a} in orange) on the average molecular envelope obtained for VirA ACP_{5b}-^{C}DD–VirFG ^{N}DD-KS_6. This analysis suggests that the interaction between the subunits creates a single, closed reaction chamber within module 5. However, because the estimated distance between the ACP_{5b} and KS_6 active sites would not allow for direct chain transfer, conformational adjustment is likely to occur in the presence of polyketide intermediate on ACP_{5b}.[7]

Table S7); thus, it is not possible at present to discern a clear "code" that drives interaction specificity. However, in no case do the docking domain pairs from the same system have identical sets of residues at these positions, consistent with their potential contribution to partner choice through effects on k_{on}.[66,67] Similarly, the precise nature of the hydrophobic interface residues varies between multiple docking domain pairs from the same system (e.g., chivosazol (Chi), oxazolomycin (Ozm), thailandamide (Tai), and enacyloxin (5922)); thus, they may also contribute to docking specificity through effects on k_{on}, k_{off}, or both (Table S7).[66,67]

Positioning of the Flanking ACP and KS Domains as Determined by SAXS. The challenge of achieving specific interactions is substantial when multiple, candidate interaction partners have similar sequences and 3D structures, as is the case within the C- and N-terminal docking domains from the Chi, Ozm, enacyloxin, and Tai systems (Table S3).[68] This raises the question of whether contacts between the flanking domains and the docking complex or direct contact between the domains themselves also contribute to specificity as observed for *cis*-AT PKSs.[54] To address this issue, we aimed to identify by SAXS the locations of the flanking VirA ACP_{5b} and VirFG KS_6 domains relative to the docking domains. For this, we analyzed the following constructs: VirA ACP_{5b}-^{C}DD (data from ref 21), VirA ACP_{5b}-^{C}DD/VirFG ^{N}DD complex, VirA ACP_{5b}-^{C}DD–VirFG ^{N}DD (a covalent fusion between ACP_{5b}-^{C}DD and VirFG ^{N}DD), and a fusion protein spanning from ACP_{5b} to KS_6 (VirA ACP_{5b}-^{C}DD–VirFG ^{N}DD-KS_6) (Figure S3); as mentioned previously, it was not possible to investigate noncovalent domain combinations involving VirFG ^{N}DD-KS_6 (i.e., VirA ^{C}DD/VirFG ^{N}DD-KS_6 complex and VirA ACP_{5b}-^{C}DD/VirFG ^{N}DD-KS_6 complex) because of aggregation. For the non-

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155–4167

covalent complex VirA ACP_{5b}-^{C}DD/VirFG ^{N}DD, we used equal concentrations of both partners well above the measured K_d values in order to favor complex formation and disfavor self-association of VirA ^{C}DD.

Previous analysis by SAXS of VirA ACP^{5b}-^{C}DD[21] yielded an average molecular envelope consistent with the homodimerization observed for VirA ^{C}DD by NMR, ITC, and SAXS. In the presence of equimolar concentrations of its partner VirFG ^{N}DD, however, a new average molecular envelope was obtained (Figures 4A and S7 and Table S5), in which density was observed only for a single ACP domain and the docking domain complex. Essentially identical data were acquired on VirA ACP_{5b}-^{C}DD−VirFG ^{N}DD (Figures 4B and S7), strengthening the conclusion that joining the docking domains covalently does not alter how they interact. The molecular weight obtained in each case from SAXS MoW shows that only one copy of each docking domain is present in the complex (Table S5). The data also confirm that the docking complex is formed from only a single copy each of the C- and N-terminal docking domains and that ACP_{5b} does not contact it directly. Analysis of VirA ACP_{5b}-^{C}DD−VirFG ^{N}DD-KS_6 (Figure S7) gave an average molecule envelope (χ^2 = 1.77) into which the homodimeric KS_6 domain could be confidently placed using SUPCOMB[43] (Figure 4C). It was not possible, however, to locate computationally the two ACP_{5b} domains and the two docking domain complexes because the scattering data are dominated by the signal from the KS_6. However, from the obtained molecular envelope it is clear that the 42-residue sequence separating VirFG ^{N}DD from KS_6 must be largely structured because there is no electronic density which could correspond to an unstructured region of this size. Furthermore, making the reasonable assumption that the ACP_{5b} domains are located on the two extremities of the structure (Figure 4C) allowed us to superimpose the molecular envelope on that obtained previously for the portion of module 5 encompassing the KS_5, ACP_{5a}, and ACP_{5b} domains (KS_5−ACP_{5a}−ACP_{5b})[21] (Figure 4D). This analysis shows that the two sets of ACP_{5b} domains superimpose nearly perfectly and therefore that the global form observed for VirA ACP_{5b}-^{C}DD−VirFG ^{N}DD-KS_6 is compatible with the previously characterized structure of module 5 (ref 21). In the absence of docking with the downstream subunit, module 5 adopts a wider open shape (distance ACP_{5b} to ACP_{5b} in module 5 alone = 279 Å[21] vs 200 Å in KS_5−ACP_{5a}−ACP_{5b}), so as proposed earlier,[21] interaction with VirFG induces closure of the module.

DISCUSSION

The majority of modular PKSs incorporate multiple polypeptide subunits, suggesting an evolutionary advantage in distributing the modules among several smaller proteins. Communication between subunits in all *cis*-AT PKSs characterized to date is orchestrated, at least in part, by matched terminal docking domains (DDs).[14−17] We have shown here that for an important subset of interpolypeptide junctions, analogous DDs operate in *trans*-AT PKSs. Of the 27 additional pairs of DDs identified by sequence/secondary structure analysis (Table S3), 10 are clearly predicted to exhibit properties (e.g., interaction propensity, extent of disorder, and lack of coiled-coil formation) in line with those of VirA ^{C}DD and VirFG ^{N}DD (Appendix), whereas the N-terminal docking domains at three other interfaces also appear similar to VirFG ^{N}DD. For the remaining nine interfaces, however, the N-terminal DD is predicted[69] to adopt a coiled-coil conformation. Comparative sequence analysis further suggests that this group of sequences may in fact be Class 2 type docking domains (Appendix), implying that multiple structural classes of docking elements also act in *trans*-AT PKS.

Overall, the recognition of four distinct structural classes of docking domains now permits a systematic comparison of docking strategies that give rise to both high specificity but relatively low affinity interactions (ca. 2−50 μM). All of the docking complexes incorporate contacts between α helices. Although architecturally distinct, the docking interfaces are largely hydrophobic and similar in size (1100−1800 Å^2). Selectivity among potential partners seems to arise in all cases from electrostatic interactions between a few, strategically placed charged residues, although many PKS assembly lines also incorporate several, structurally orthogonal classes of DDs (i.e., classes 1a and 1b[13]).

Our work also suggests that weak docking affinity arises in the case of multiple *trans*-AT PKSs through use of an intrinsically disordered protein as a docking partner. IDPs are rare in prokaryotes[70] (though relatively more common in high-GC, large-genome-size bacteria such as the Actinobacteria),[71] and VirFG ^{N}DD represents a strong candidate for the first IDP to be shown to participate in bacterial secondary metabolism. VirA ^{C}DD may also be an IDP, but a disorder-to-order transition could not be observed in the context of a discrete protein because of self-association. IDPs are expected to show much higher rates of sequence variation than those of structured regions of the same proteins.[65] This is indeed the case for the Vir[4] and Sna[55] systems, which both govern virginiamycin biosynthesis, because the corresponding subunits show much higher amino acid sequence identity than do the docking domain regions (e.g., 66% identity for VirA/SnaE2 but only 53% for the ^{C}DDs; 66% overall identity between VirFG/SnaE3 vs 39% for the ^{N}DDs).[53] This observation suggests that it is the amino acid character and not the specific sequences of these regions that is critical. However, given the unpredictable effects of mutations on induced folding and the structure of IDPs,[65] it seems likely that using this class of DD to engineer new interfaces will require transplantation of matched sets of docking domains. Our study identifies clear boundaries to be used in such experiments.

This work has also permitted for the first time visualization of a full interface formed between two consecutive subunits within a *trans*-AT PKS. Our previous analysis by SAXS showed VirA module 5 to have an open architecture but that flexibility on the level of the "arms" might allow it to clamp down to engage with the KS_6 domain of the next module. Indeed, the distance between the ACP_{5b} domains seems ideally adapted to accommodate the homodimeric KS_6 and two flanking docking domain complexes. In this way, docking between the two subunits in the context of a larger multienzyme complex (for example, incorporating the *trans*-AT and β-processing cassette) could give rise to a closed space containing a single reaction center in which reactive intermediates are protected. However, the estimated distance between ACP_{5b} and KS_6 active sites in this *apo* structure (ca. 80 Å) is not compatible with direct chain transfer between the two domains as required by the biosynthetic mechanism. This observation is consistent with recent studies by cryo-electron microscopy (cryo-EM) on the interface between subunits from the pikromycin (PIK) *cis*-AT PKS, in which a stable interaction between the partner ACP and KS domains was only observed in the presence of chain

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

extension intermediate tethered to the ACP.[6] We propose that the intrinsic flexibility of the linker regions separating the various functional and docking domains will allow adoption of a configuration in which ACP_{5b} is correctly positioned to interact with KS_6. However, visualizing this productive chain transfer complex by SAXS or cryo-EM will likely require studies of appropriately modified VirA module 5, work which is in progress. In the meantime, the absence of a specific interface between the ACP and KS domains is encouraging from an engineering perspective because it suggests that transplantation of docking domains may be sufficient to create novel functional intersubunit interfaces in *trans*-AT PKSs as long as the receiving KS has appropriate substrate specificity.[56]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b13372.

Appendix. (PDF)

Additional results, methods, and references. (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*arnaud.gruez@univ-lorraine.fr

*benjamin.chagot@univ-lorraine.fr

*kira.weissman@univ-lorraine.fr

Author Contributions

⊥J.D. and T.A. contributed equally to this work

Funding

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 Université de Lorraine (UL), and the Lorraine Region (Bonus Qualité Recherche (BQR) grants to K.J.W. and B.C).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Access to the 600 MHz NMR spectrometer of the "SCBIM" (Service Commun de Biophysiocochimie des Interactions Moléculaires) of the FR3209 CNRS/Université de Lorraine was greatly appreciated. We thank Dr. Javier Perez and Dr. Aurélien Thureau for help during data collection on the beamline SWING at the synchrotron SOLEIL (Gif-sur-Yvette), and Dr. Christophe Velours for help with the SEC-MALLS experiments. We would also like to acknowledge L. Ornella Bimai for help in preparing the docking domain constructs and Alexandre Kriznik for assistance in analyzing the ITC and CD data.

■ REFERENCES

(1) Hertweck, C. *Angew. Chem., Int. Ed.* **2009**, *48*, 4688.

(2) Katz, L. *Methods Enzymol.* **2009**, *459*, 113.

(3) Piel, J. *Nat. Prod. Rep.* **2010**, *27*, 996.

(4) Pulsawat, N.; Kitani, S.; Nihira, T. *Gene* **2007**, *393*, 31.

(5) Weissman, K. J.; Müller, R. *ChemBioChem* **2008**, *9*, 826.

(6) Dutta, S.; Whicher, J. R.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Congdon, G. R.; Narayan, A. R.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 512.

(7) Whicher, J. R.; Dutta, S.; Hansen, D. A.; Hale, W. A.; Chemler, J. A.; Dosey, A. M.; Narayan, A. R.; Håkansson, K.; Sherman, D. H.; Smith, J. L.; Skiniotis, G. *Nature* **2014**, *510*, 560.

(8) Gokhale, R. S.; Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Science* **1999**, *284*, 482.

(9) Tsuji, S. Y.; Cane, D. E.; Khosla, C. *Biochemistry* **2001**, *40*, 2326.

(10) Menzella, H. G.; Reid, R.; Carney, J. R.; Chandran, S. S.; Reisinger, S. J.; Patel, K. G.; Hopwood, D. A.; Santi, D. V. *Nat. Biotechnol.* **2005**, *23*, 1171.

(11) Menzella, H. G.; Carney, J. R.; Santi, D. V. *Chem. Biol.* **2007**, *14*, 143.

(12) Reeves, C. D.; Ward, S. L.; Revill, W. P.; Suzuki, H.; Marcus, M.; Petrakovsky, O. V.; Marquez, S.; Fu, H.; Dong, S. D.; Katz, L. *Chem. Biol.* **2004**, *11*, 1465.

(13) Thattai, M.; Burak, Y.; Shraiman, B. I. *PLoS Comput. Biol.* **2007**, *3*, e186.

(14) Richter, C. D.; Nietlispach, D.; Broadhurst, R. W.; Weissman, K. J. *Nat. Chem. Biol.* **2008**, *4*, 75.

(15) Broadhurst, R. W.; Nietlispach, D.; Wheatcroft, M. P.; Leadlay, P. F.; Weissman, K. J. *Chem. Biol.* **2003**, *10*, 723.

(16) Buchholz, T. J.; Geders, T. W.; Bartley, F. E., 3rd; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. *ACS Chem. Biol.* **2009**, *4*, 41.

(17) Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. *Chem. Biol.* **2013**, *20*, 1340.

(18) Tang, Y.; Kim, C.-Y.; Mathews, I. I.; Cane, D. E.; Khosla, C. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 11124.

(19) Tang, Y.; Chen, A. Y.; Kim, C.-Y.; Cane, D. E.; Khosla, C. *Chem. Biol.* **2007**, *14*, 931.

(20) Gokhale, R. S.; Hunziker, D.; Cane, D. E.; Khosla, C. *Chem. Biol.* **1999**, *6*, 117.

(21) Davison, J.; Dorival, J.; Rabeharindranto, H.; Mazon, H.; Chagot, B.; Gruez, A.; Weissman, K. J. *Chem. Sci.* **2014**, *5*, 3081.

(22) Gu, L.; Eisman, E. B.; Dutta, S.; Franzmann, T. M.; Walter, S.; Gerwick, W. H.; Skiniotis, G.; Sherman, D. H. *Angew. Chem., Int. Ed.* **2011**, *50*, 2795.

(23) McGuffin, L. J.; Bryson, K.; Jones, D. T. *Bioinformatics* **2000**, *16*, 404.

(24) Pei, J.; Grishin, N. V. *Methods Mol. Biol.* **2014**, *1079*, 263.

(25) Keating, T. A.; Marshall, C. G.; Walsh, C. T.; Keating, A. E. *Nat. Struct. Biol.* **2002**, *9*, 522.

(26) Zheng, J.; Gay, D. C.; Demeler, B.; White, M. A.; Keatinge-Clay, A. T. *Nat. Chem. Biol.* **2012**, *8*, 615.

(27) *Molecular Cloning: A Laboratory Manual*; Sambrook, J., Fritsch, E. F., Maniatis, T., Eds.; Cold Spring Harbor Laboratory Press: New York, 1989.

(28) Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. *Anal. Chem.* **2012**, *84*, 5066.

(29) Cavanagh, J.; Fairbrother, W.; Palmer, A.; Skelton, N. *Protein NMR Spectroscopy: Principles and Practice*; Academic Press Inc.: New York, 1996.

(30) Güntert, P. Automated NMR Structure Calculation with CYANA. In *Protein NMR Techniques*; Downing, A. K., Ed.; Methods in Molecular Biology series; Humana Press: Totowa, NJ, 2004; Vol. *278*, pp 353−378.

(31) Herrmann, T.; Güntert, P.; Wüthrich, K. *J. Mol. Biol.* **2002**, *319*, 209.

(32) Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. *J. Biomol. NMR* **2009**, *44*, 213.

(33) Case, D. A.; Berryman, J. T.; Betz, R. M.; Cerutti, D. S.; Cheatham, T. E., III; Darden, T. A.; Duke, R. E.; Giese, T. J.; Gohlke, H.; Goetz, A. W.; Homeyer, N.; Izadi, S.; Janowski, P.; Kaus, J.; Kovalenko, A.; Lee, T. S.; Legrand, S.; Li, P.; Luchko, T.; Luo, R.; Madej, B.; Merz, K. M.; Monard, G.; Needham, P.; Nguyen, H.; Nguyen, H. T.; Omelyan, I.; Onufriev, A.; Roe, D. R.; Roitberg, A.; Salomon-Ferrer, R.; Simmerlin, C. L.; Smith, W.; Swails, J.; Walker, R. C.; Wang, J.; Wolf, R. M.; Wu, X.; York, D. M.; Kollman, P. A. *AMBER 14*; University of California: San Francisco, CA, 2014.

(34) Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167

(35) Böhm, G.; Muhr, R.; Jaenicke, R. *Protein Eng., Des. Sel.* **1992**, *5*, 191.
(36) David, G.; Pérez, J. *J. Appl. Crystallogr.* **2009**, *42*, 892.
(37) Pieper, R.; Ebert-Khosla, S.; Cane, D.; Khosla, C. *Biochemistry* **1996**, *35*, 2054.
(38) Bycroft, M.; Weissman, K. J.; Staunton, J.; Leadlay, P. F. *Eur. J. Biochem.* **2000**, *267*, 520.
(39) Svergun, D. *J. Appl. Crystallogr.* **1992**, *25*, 495.
(40) Fischer, H.; de Oliveira Neto, M.; Napolitano, H. B.; Polikarpov, I.; Craievich, A. F. *J. Appl. Crystallogr.* **2010**, *43*, 101.
(41) Knobloch, J. E.; Shaklee, P. N. *Anal. Biochem.* **1997**, *245*, 231.
(42) Svergun, D. I. *Biophys. J.* **1999**, *76*, 2879.
(43) Kozin, M.; Svergun, D. *J. Appl. Crystallogr.* **2001**, *34*, 33.
(44) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *J. Appl. Crystallogr.* **2003**, *36*, 1277.
(45) Volkov, V.; Svergun, D. *J. Appl. Crystallogr.* **2003**, *36*, 860.
(46) Svergun, D. I.; Koch, M. H. J.; Timmins, P. A.; May, R. P. *Small Angle X-Ray and Neutron Scattering from Solutions of Biological Macromolecules*. OUP: Oxford, U.K., 2013.
(47) Söding, J.; Biegert, A.; Lupas, A. N. *Nucleic Acids Res.* **2005**, *33*, W244.
(48) Gay, D. C.; Gay, G.; Axelrod, A. J.; Jenner, M.; Kohlhaas, C.; Kampa, A.; Oldham, N. J.; Piel, J.; Keatinge-Clay, A. T. *Structure* **2014**, *22*, 444.
(49) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusdhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. In *Current Protocols in Protein Science*; Wiley: Hoboken, NJ, 2007; Chapter 2, Unit 2.9.
(50) Svergun, D.; Barberato, C.; Koch, M. *J. Appl. Crystallogr.* **1995**, *28*, 768.
(51) Receveur-Bréchot, V.; Durand, D. *Curr. Protein Pept. Sci.* **2012**, *13*, 55.
(52) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403.
(53) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. *Mol. Syst. Biol.* **2011**, *7*, 539.
(54) Wu, N.; Cane, D. E.; Khosla, C. *Biochemistry* **2002**, *41*, 5056.
(55) Mast, Y.; Weber, T.; Gölz, M.; Ort-Winklbauer, R.; Gondran, A.; Wohlleben, W.; Schinko, E. *Microb. Biotechnol.* **2011**, *4*, 192.
(56) Nguyen, T.; Ishida, K.; Jenke-Kodama, H.; Dittmann, E.; Gurgui, C.; Hochmuth, T.; Taudien, S.; Platzer, M.; Hertweck, C.; Piel, J. *Nat. Biotechnol.* **2008**, *26*, 225.
(57) Sibille, N.; Bernadó, P. *Biochem. Soc. Trans.* **2012**, *40*, 955.
(58) Bourhis, J.-M.; Receveur-Bréchot, V.; Oglesbee, M.; Zhang, X.; Buccellato, M.; Darbon, H.; Canard, B.; Finet, S.; Longhi, S. *Protein Sci.* **2005**, *14*, 1975.
(59) Laskowski, R. A.; Rullmann, J. A. C.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. *J. Biomol. NMR* **1996**, *8*, 477.
(60) Chen, V. B.; Arendall, W. B., 3rd; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12.
(61) Bailey, T. L.; Gribskov, M. *Bioinformatics* **1998**, *14*, 48.
(62) Holm, L.; Rosenström, P. *Nucleic Acids Res.* **2010**, *38*, W545.
(63) Uversky, V. N. *Protein Sci.* **2013**, *22*, 693.
(64) Liu, Z.; Huang, Y. *Protein Sci.* **2014**, *23*, 539.
(65) Oldfield, C. J.; Dunker, A. K. *Annu. Rev. Biochem.* **2014**, *83*, 553.
(66) Selzer, T.; Schreiber, G. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 190.
(67) Schreiber, G.; Haran, G.; Zhou, H.-X. *Chem. Rev.* **2009**, *109*, 839.
(68) Schreiber, G.; Keating, A. E. *Curr. Opin. Struct. Biol.* **2011**, *21*, 50.
(69) Lupas, A.; Van Dyke, M.; Stock, J. *Science* **1991**, *252*, 1162.
(70) Williamson, M. P.; Potts, J. R. *Biochem. Soc. Trans.* **2012**, *40*, 945.
(71) Pavlović-Lažetić, G. M.; Mitić, N. S.; Kovačević, J. J.; Obradović, Z.; Malkov, S. N.; Beljanski, M. V. *BMC Bioinf.* **2011**, *12*, 66.

DOI: 10.1021/jacs.5b13372
J. Am. Chem. Soc. 2016, 138, 4155−4167