

## Non-Ribosomal Peptide Synthetase

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 14171–14178

International Edition: doi.org/10.1002/anie.202103498

German Edition: doi.org/10.1002/ange.202103498

## Cooperation between a T Domain and a Minimal C-Terminal Docking Domain to Enable Specific Assembly in a Multiprotein NRPS

Jonas Watzel, Elke Duchardt-Ferner, Sepas Sarawi, Helge B. Bode,\* and Jens Wöhnert\*

**Abstract:** Non-ribosomal peptide synthetases (NRPS) produce natural products from amino acid building blocks. They often consist of multiple polypeptide chains which assemble in a specific linear order via specialized N- and C-terminal docking domains (<sup>N/C</sup>DDs). Typically, docking domains function independently from other domains in NRPS assembly. Thus, docking domain replacements enable the assembly of “designer” NRPS from proteins that normally do not interact. The multiprotein “peptide-antimicrobial-*Xenorhabdus*” (PAX) peptide-producing PaxS NRPS is assembled from the three proteins PaxA, PaxB and PaxC. Herein, we show that the small <sup>C</sup>DD of PaxA cooperates with its preceding thiolation (T<sub>1</sub>) domain to bind the <sup>N</sup>DD of PaxB with very high affinity, establishing a structural and thermodynamical basis for this unprecedented docking interaction, and we test its functional importance *in vivo* in a truncated PaxS assembly line. Similar docking interactions are apparently present in other NRPS systems.

## Introduction

Non-ribosomal peptide synthetases (NRPSs) manufacture a very diverse range of natural products starting from amino acids. Many of these products such as daptomycin,

vancomycin or bleomycin are highly relevant for clinical applications.<sup>[1]</sup> The biosynthesis of these peptide products is accomplished by the orchestrated interplay of a number of functionally distinct catalytic domains, which are grouped into modules. Each module is responsible for incorporating one specific amino acid building block into the nascent product. A typical elongation module in an NRPS contains an adenylation (A) domain, a thiolation (T) or peptidyl-carrier protein (PCP) domain and a condensation (C) domain.<sup>[2]</sup> The A domain selects a specific amino acid and uses ATP to chemically activate this amino acid as an aminoacyl adenylate. The adjacent T domain in its catalytically active *holo* form contains a phosphopantetheinyl (Ppant) arm derived from coenzyme A covalently attached to a conserved serine residue. The reactive thiol group of the Ppant arm reacts with the aminoacyl intermediate activated by the preceding A domain and forms a thioester with the amino acid while releasing AMP. The subsequent condensation domain (C) then catalyzes the formation of a peptide bond between the amino acid bound to its preceding T domain and an amino acid bound to the T domain of the downstream module. Additional domains with specific tailoring functions such as methyltransferases or epimerization domains can be included in a functional module. The final product is cleaved off from the Ppant arm of the T domain in the last elongation module by a thioesterase (TE) domain localized at the C-terminus of a termination module, which releases either linear, cyclic or branched cyclic peptides. Multiple modules can be located on a single protein. An extreme example in this regard is the NRPS synthesizing kolossin A which consists of 15 modules on a single polypeptide chain.<sup>[3]</sup> In contrast, in multiprotein NRPS systems the modules are distributed between multiple protein chains.

In single-protein NRPSs the linear order of modules from the N- to the C-terminus directly determines the linear order of amino acid building blocks in the final product. This relationship is known as the collinearity rule.<sup>[4]</sup> In multiprotein NRPS a specific linear arrangement of the individual proteins and thereby of the different modules is established by non-covalent interactions between specialized structural elements or domains of the individual proteins. These domains are often localized at the N- and C-termini of the respective proteins of a multiprotein NRPS. They were initially referred to as “communication-mediating domains”<sup>[5]</sup> in NRPS or as “inter-polypeptide linkers”<sup>[6]</sup> or “docking domains (DD)”<sup>[7]</sup> in the architecturally related polyketide synthases (PKSs) and are now generally referred to as docking domains (DDs) in megasynthase systems. Since they are able to direct the correct linear assembly of multiprotein megasynthases<sup>[5,8,9]</sup> and are portable between different mega-

[\*] Prof. Dr. H. B. Bode

Department of Natural Products in Organismic Interactions  
Max-Planck-Institute for Terrestrial Microbiology  
35043 Marburg (Germany)

and

Senckenberg Gesellschaft für Naturforschung  
60325 Frankfurt am Main (Germany)  
E-mail: helge.bode@mpi-marburg.mpg.de

Dr. E. Duchardt-Ferner, S. Sarawi, Prof. Dr. J. Wöhnert  
Institute of Molecular Biosciences and Center for Biomolecular  
Magnetic Resonance (BMRZ)  
Goethe University Frankfurt  
60438 Frankfurt am Main (Germany)  
E-mail: woehnert@bio.uni-frankfurt.de

J. Watzel, S. Sarawi, Prof. Dr. H. B. Bode  
Molecular Biotechnology  
Institute of Molecular Biosciences  
Goethe University Frankfurt  
60438 Frankfurt am Main (Germany)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:  
<https://doi.org/10.1002/anie.202103498>.

© 2021 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

synthase systems often without diminishing the activity of the neighboring catalytic domains,<sup>[5,10,11]</sup> they are ideal tools to enable combinatorial biosynthesis by connecting protein chains of non-related megasynthases in a predictable, functional and specific way in order to synthesize structurally novel and diverse “designer” products not found in nature.<sup>[10]</sup>

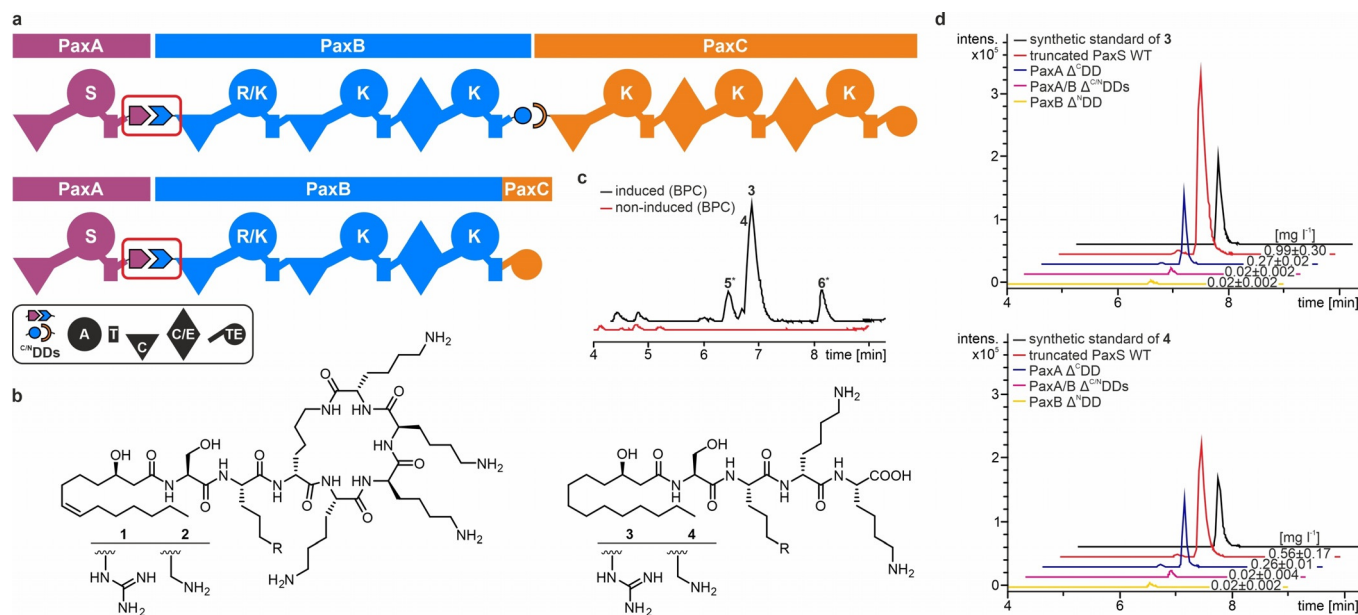
The structural analysis of docking domains and their complexes revealed a large structural diversity.<sup>[11–15]</sup> However, a majority of docking domain structures is dominated by  $\alpha$ -helical secondary structure elements. Typical docking domains are rather small (ca. 30–65 amino acids). They bind their cognate docking domains with medium affinities with dissociation constants between 5 and 25  $\mu\text{M}$ .<sup>[8,11,12,14]</sup> The successful transplantation of docking domains between different functionally unrelated megasynthase systems<sup>[9,10,16]</sup> already suggests that they act independently from the other functional domains in directing specific protein–protein interactions between megasynthase components. Affinity measurements for cognate docking domain pairs of megasynthases that included flanking functional domains such as the acyl carrier proteins (ACP)<sup>[14,17]</sup> in PKS, T domains or C domains<sup>[8]</sup> in NRPS showed that the affinity between the docking domains did not increase significantly in the presence of the flanking domains in comparison to the isolated docking domains. Thus, docking domains are thought to be functionally independent in mediating the specificity of the assembly in multiprotein megasynthases and not supported by the other domains in the protein chain.

Here, we focus on the multiprotein NRPS PaxS consisting of the three proteins PaxA, PaxB and PaxC (Figure 1 a). PaxS occurs in bacterial species belonging to the genus *Xenorhab-*

*du* and produces “peptide-antimicrobial-*Xenorhabdus* (PAX)” peptides.<sup>[18]</sup> PAX peptides such as **1** and **2** (Figure 1 b) are cyclic peptides with predominantly basic residues and N-terminally attached acyl chains that protect the bacterial producer cell against insect-derived basic antimicrobial peptides.<sup>[19]</sup> PaxA contains a specialized starter C domain transferring an acyl chain to the amino group of a serine which is activated by the A domain and stored on the downstream T domain (T<sub>1</sub>). PaxB consists of three canonical elongation modules with a preference for arginine/lysine or lysine. PaxC contains three additional canonical elongation modules and a C-terminal thioesterase (TE) domain which also catalyzes the cyclization between a lysine side chain amino group and the TE-bound C-terminus of the peptide. Based on the structure of the PAX peptides and the collinearity rule PaxA, PaxB and PaxC form a unidirectional assembly line where the C-terminus of PaxA supposedly interacts with the N-terminus of PaxB and the C-terminus of PaxB with the N-terminus of PaxC.

We recently established the structural basis for the specific interaction between the PaxB C-terminal docking domain (<sup>C</sup>DD) and the PaxC N-terminal docking domain (<sup>N</sup>DD) in the PaxS NRPS of *Xenorhabdus bovienii* SS-2004.<sup>[20]</sup> The structure of the PaxB <sup>C</sup>DD/PaxC <sup>N</sup>DD complex defined a novel type of docking domain interaction.<sup>[12]</sup> The PaxB <sup>C</sup>DD/PaxC <sup>N</sup>DD pair can be grafted onto unrelated multiprotein NRPS systems to induce functional and specific interchain interactions in these systems.<sup>[10]</sup>

Here, we investigate the interaction between PaxA and PaxB. This interaction is mediated by an unprecedented docking interface. It not only involves the <sup>C</sup>DD of PaxA and



**Figure 1.** The “Peptide-antimicrobial-*Xenorhabdus*” (PAX) peptide-producing NRPS PaxS. a) Schematic representation of the full-length and truncated PaxS producing the peptides **1–4**. For domain assignment the following symbols are used: adenylation (A) domain, large circles; thiolation (T) domain, rectangle; condensation (C) domain, triangle; dual condensation/epimerization (C/E) domain, diamond; thioesterase (TE) domain, small circle. For each adenylation domain its amino acid preference is given by the single letter code. The PaxA/B docking domain interface framed in red was investigated in this work. b) Main products **1/2** of the full-length and **3/4** of the truncated PaxS. c/d) HPLC/MS analysis of the peptide production **3–6** of the (modified) truncated PaxS, which are shown as base peak chromatograms (BPC) in (c) and extracted ion chromatograms (EICs) of the main products **3/4** in (d).

the <sup>N</sup>DD of PaxB as expected in analogy to previously described docking interactions but requires the cooperation of the T domain and the <sup>C</sup>DD of PaxA to bind the PaxB <sup>N</sup>DD with a surprisingly high affinity. This docking interaction is functionally important for efficient product formation in vivo and also occurs in other NRPS systems in bacteria not related to *Xenorhabdus* species.

## Results and Discussion

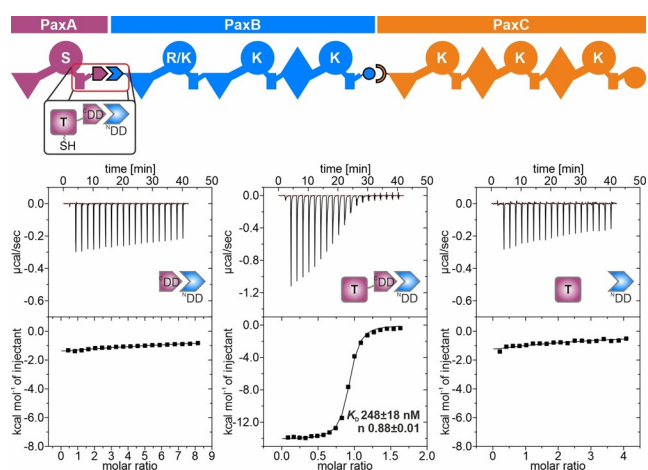
An initial bioinformatics analysis showed that the C-terminal T<sub>1</sub> domain of PaxA is followed by a short stretch of only 19 amino acids predicted to be unstructured while the N-terminal C domain of PaxB is preceded by approximately 36 amino acids predicted to contain a region with elevated  $\alpha$ -helical propensity. While these two stretches of amino acids could represent putative C- and N-terminal docking domains, they show no clear sequence homologies to known docking domain pairs. Furthermore, the putative PaxA <sup>C</sup>DD is significantly shorter than other known types of docking domains.

To test if the putative PaxA <sup>C</sup>DD and PaxB <sup>N</sup>DD and their interactions are important for efficient product formation in the *X. bovienii* SS-2004 PaxS NRPS in vivo, we created a truncated version of the PaxS NRPS consisting of only two proteins. This artificial NRPS contained full-length native PaxA and a modified PaxB where the thioesterase domain of the PaxC termination module (PaxB-TE<sub>PaxC</sub>, Figure 1a) replaces the native <sup>C</sup>DD of PaxB. It was expressed successfully in the heterologous host *E. coli* DH10B::mtaA<sup>[21]</sup> and shown to efficiently produce the two shortened linear PAX tetra-peptides **3** (SRkK; D-amino acids in italics and lower case throughout this work (Figure 1b)) and **4** (SKkK, Figure 1b) with an N-terminally attached (3*R*)-3-hydroxytetradecanoyl fatty acid moiety as expected. Two additional side products (\*; (3*R*,7*Z*)-3-hydroxytetradec-7-enoyl-SRkK (**5**)) and ((3*R*)-3-hydroxytetradecanoyl-SK (**6**)) were detected (Figure 1c). The amino acid and acyl chain composition of **3–6** was confirmed by feeding experiments followed by HR-HPLC/MS analysis (Figure S1/S2) and the comparison of the MS<sup>2</sup> fragmentation patterns with those of synthetic standards (**3**, **4**) (Figure S3). In this truncated NRPS product formation should only rely on a productive non-covalent interaction between PaxA and PaxB. Deletion of either the putative PaxB <sup>N</sup>DD or of both the PaxB <sup>N</sup>DD and the PaxA <sup>C</sup>DD almost completely abolished product formation in vivo while a deletion of the PaxA <sup>C</sup>DD alone significantly reduced product formation to 25 % for **3** and to 50 % for **4** (Figure 1d). This shows that the putative docking domains play an important role for mediating productive non-covalent interactions between PaxA and PaxB in vivo. However, since the deletion of the short PaxA <sup>C</sup>DD alone did not completely abolish product formation, other structural elements in PaxA might also contribute to the docking interaction in vivo.

To test the interaction of the putative PaxA <sup>C</sup>DD and PaxB <sup>N</sup>DD from the *X. bovienii* SS-2004 PaxS in vitro, we overexpressed and purified a PaxA<sub>1071–1089</sub> (PaxA <sup>C</sup>DD) and a PaxB<sub>1–36</sub> (PaxB <sup>N</sup>DD) peptide (Figure S4/S5a, Table S1).

Surprisingly, the interaction between the PaxA <sup>C</sup>DD and the PaxB <sup>N</sup>DD was too weak to be reliably quantified by isothermal titration calorimetry (ITC; Figure S5b). However, when we titrated PaxA<sub>986–1089</sub>, a di-domain construct containing both the entire T<sub>1</sub> domain in its *apo* state and the putative <sup>C</sup>DD of PaxA, with the PaxB <sup>N</sup>DD peptide in ITC experiments we measured a  $K_D$  of  $201 \pm 20$  nM (Figure S5c) for this interaction with a 1:1 stoichiometry ( $n$ :  $0.74 \pm 0.08$ ). In contrast, in a titration of the isolated T<sub>1</sub> domain of PaxA (PaxA<sub>986–1076</sub>) with the PaxB <sup>N</sup>DD peptide the interaction was again too weak to be quantified (Figure S5d). Thus, in the PaxS NRPS from *X. bovienii* SS-2004 the non-covalent interaction between the PaxA and PaxB proteins is dependent on the presence of both the T<sub>1</sub> and the <sup>C</sup>DD of PaxA which apparently have to cooperate to bind the PaxB <sup>N</sup>DD with an affinity that is remarkably high in comparison to the affinities for other characterized docking domain pairs.<sup>[8,12–15]</sup>

Unfortunately, despite its very high affinity the complex between the PaxA T<sub>1</sub>-<sup>C</sup>DD di-domain construct and the PaxB <sup>N</sup>DD peptide from the PaxS NRPS from *X. bovienii* SS-2004 was not suitable for a full structure determination. The closely related species *Xenorhabdus cabanillasii* JM26 also contains the PaxS NRPS with PaxA and PaxB (Figure S4). More importantly, the *X. cabanillasii* PaxA T<sub>1</sub>-<sup>C</sup>DD di-domain (PaxA<sub>981–1084</sub>) construct bound a slightly length-optimized *X. cabanillasii* PaxB <sup>N</sup>DD construct (PaxB<sub>1–30</sub>) with a similarly high affinity ( $248 \pm 18$  nM) and a 1:1 stoichiometry ( $n$ :  $0.88 \pm 0.01$ ) according to ITC (Figure 2/ Figure S6) as observed for *X. bovienii* SS-2004. The affinity of this interaction was not modulated by the presence of the Ppant moiety in the PaxA T<sub>1</sub> domain (Figure S7). No quantifiable interactions were observed by ITC between the isolated PaxA <sup>C</sup>DD (PaxA<sub>1066–1084</sub>) and the PaxB <sup>N</sup>DD as well as between the isolated PaxA T<sub>1</sub> domain (PaxA<sub>981–1071</sub>) and the PaxB <sup>N</sup>DD, respectively (Figure 2/ Figure S6). Thus, the interaction between PaxA and PaxB from the PaxS NRPS of *X. cabanillasii* has biophysical properties very similar to the one from the *X. bovienii* system



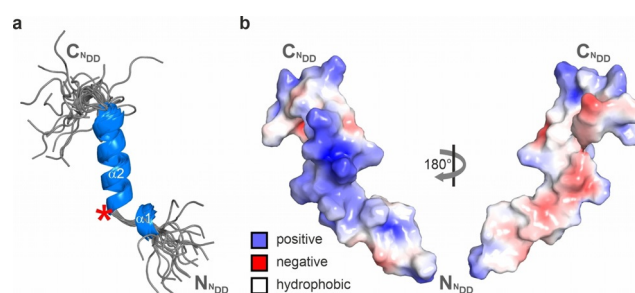
**Figure 2.** ITC analysis of the PaxA/B docking domain interface. Exemplary ITC thermograms and the derived binding curves for titrations between *X. cabanillasii* JM26 PaxA <sup>C</sup>DD/PaxB <sup>N</sup>DD, PaxA T<sub>1</sub>-<sup>C</sup>DD/PaxB <sup>N</sup>DD and PaxA T<sub>1</sub>/PaxB <sup>N</sup>DD, respectively.

and was amenable to a full structural characterization by NMR spectroscopy.

The secondary structure of the PaxA T<sub>1</sub>-<sup>C</sup>DD di-domain in its *apo* state and in the absence of the PaxB <sup>N</sup>DD was previously determined<sup>[22]</sup> based on the TALOS-N<sup>[23]</sup>-derived chemical shift index (CSI) and {<sup>1</sup>H},<sup>15</sup>N-hetNOE experiments (Figure S8). Apparently, the T domain features the characteristic four  $\alpha$ -helices ( $\alpha$ 1– $\alpha$ 4) typical of the canonical carrier protein fold.<sup>[24]</sup> The long loop between helix  $\alpha$ 1 and helix  $\alpha$ 2 is interrupted by a short, single-turn  $\alpha$ -helix as previously observed in the structure of other carrier proteins.<sup>[25,26]</sup> Importantly, residues Q1070–E1084 corresponding to the PaxA <sup>C</sup>DD are unstructured and flexible both in the framework of the di-domain construct based on the NMR data as well as in the absence of the T<sub>1</sub> domain according to CD spectroscopy (Figure S9).

The free PaxB <sup>N</sup>DD showed a surprisingly high  $\alpha$ -helical content (Figure S9/S10) in contrast to the intrinsically disordered protein (IDP) character typically observed for many other docking domains in the absence of their binding partner.<sup>[14,15,17]</sup> In our NMR structure of free PaxB (Figure 3a, Table S2, PDB ID: 7B2F) residues L11–K22 form a continuous three-turn  $\alpha$ -helix ( $\alpha$ 2).

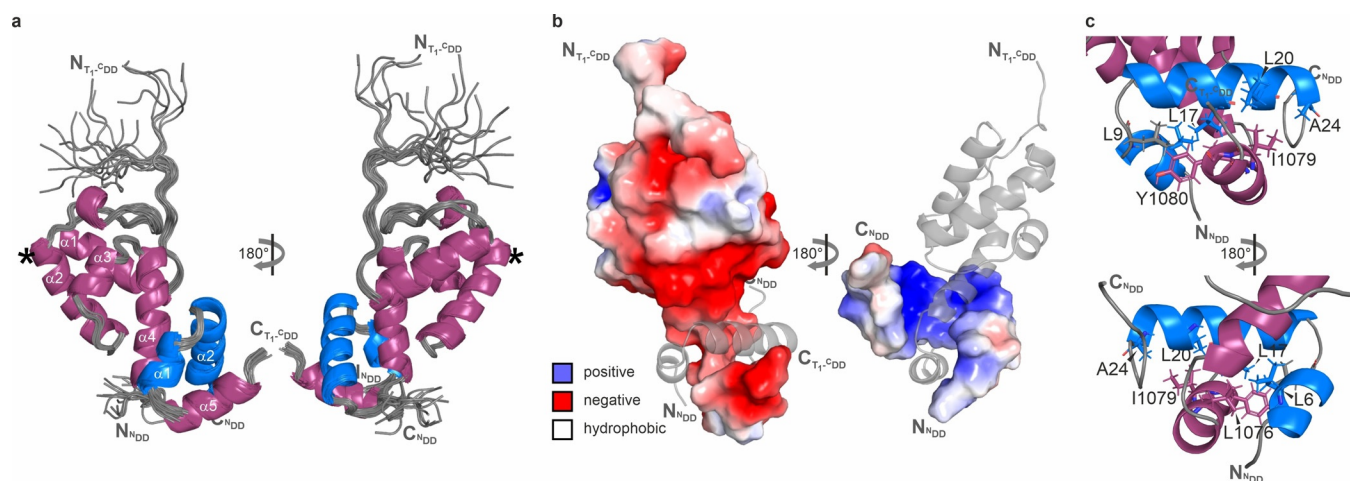
Residues L6–S8 form an additional  $\alpha$ -helical turn ( $\alpha$ 1) which is separated from  $\alpha$ 2 by a sharp kink (ca. 83°). The kink is induced by P10 which adopts a *trans* peptide bond conformation based on its <sup>13</sup>C $\gamma$  and <sup>13</sup>C $\beta$  chemical shifts.<sup>[27]</sup> According to the {<sup>1</sup>H},<sup>15</sup>N-hetNOE data for the free PaxB <sup>N</sup>DD the helix  $\alpha$ 2 from L11 to K22 is stable while the helical turn from L6 to S8 forms only transiently (Figure S10). The stability of  $\alpha$ 2 can be attributed to the presence of two pairs of residues with oppositely charged side chains—K15/E19 and E13/R16—spaced in a manner to enable favorable intrahelical electrostatic interactions. The structure of the free PaxB <sup>N</sup>DD reveals a very uneven charge distribution for  $\alpha$ 2. One side of the helix is highly positively charged while the



**Figure 3.** NMR solution structure of the free PaxB <sup>N</sup>DD. a) Solution structure bundle of the 20 energy-minimized conformers with the lowest CYANA target functions of the unbound PaxB <sup>N</sup>DD. The location of the proline residue (P10) is marked with a red asterisk. b) Electrostatic surface potentials mapped on the molecule surface of the free PaxB <sup>N</sup>DD. Negatively charged surface areas are colored in red, positively charged areas are colored in blue and white areas correspond to hydrophobic surfaces.

other surface is mostly hydrophobic with interspersed negative charges (Figure 3b).

The solution structure of the 1:1 complex between the PaxA T<sub>1</sub>-<sup>C</sup>DD di-domain construct and the PaxB <sup>N</sup>DD was determined based on a large number of intra- and intermolecular NOEs (Table S2) and its dynamics was characterized (Figure S11). The structural bundle representing the NMR solution structure of this complex (PDB ID: 7B2B) is shown in Figure 4. The PaxB <sup>N</sup>DD interacts extensively with the C-terminal half of helix  $\alpha$ 4 from the T<sub>1</sub> domain of PaxA as well as with the <sup>C</sup>DD of PaxA. Importantly, in the complex the <sup>C</sup>DD now forms a stable three-turn  $\alpha$ -helix ( $\alpha$ 5) involving residues A1073–S1082. The PaxB <sup>N</sup>DD now features an extended stable N-terminal  $\alpha$ -helix spanning residues N3–S8 ( $\alpha$ 1) while the sharp kink centered around P10 and helix  $\alpha$ 2 (L11–A24) are preserved. The pairs of interacting helices from PaxA ( $\alpha$ 4 and  $\alpha$ 5) and PaxB ( $\alpha$ 1 and  $\alpha$ 2) each form



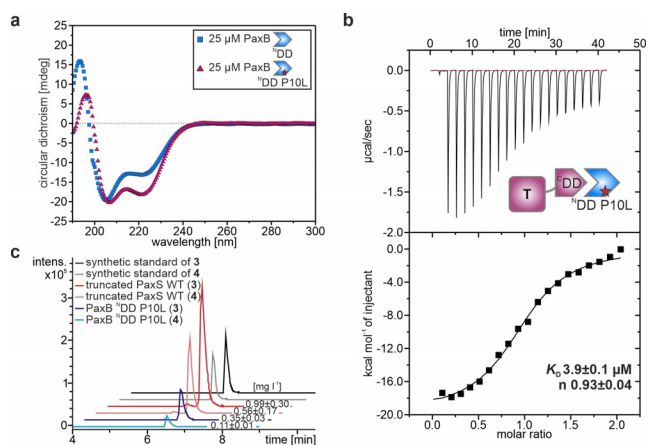
**Figure 4.** NMR solution structure of the PaxA T<sub>1</sub>-<sup>C</sup>DD/PaxB <sup>N</sup>DD complex. a) Solution structure bundle of the 20 energy-minimized conformers with the lowest CYANA target functions of the PaxA T<sub>1</sub>-<sup>C</sup>DD/PaxB <sup>N</sup>DD complex. The location of the serine residue, where the Ppant moiety is post-translationally added, is marked with an asterisk. b) Electrostatic surface potentials are mapped on the molecular surface of the complex. Negatively charged surface areas are colored in red, positively charged areas are colored in blue and white areas correspond to hydrophobic surfaces. c) Molecular architecture of the hydrophobic core of the <sup>C</sup>DD/<sup>N</sup>DD interface formed by the residues shown in stick representation.

a V shape and the two V shapes interlock in an antiparallel orientation (Figure 4). Helices  $\alpha 1$  to  $\alpha 3$  and the N-terminal half of  $\alpha 4$  from the  $T_1$  domain of PaxA are not involved in the interaction in agreement with the chemical shift perturbation data for complex formation (Figure S11a).

Overall, the fold of the  $T_1$  domain in the bound state closely corresponds to the classical right-handed up-and-down four-helical bundle fold observed in many other T domain structures in particular and carrier protein structures in general.<sup>[24–26]</sup> However, helix  $\alpha 4$  is notably longer than in many other previously reported T domain structures<sup>[28]</sup> (Figure S12) and the C-terminal helix extension is the part of the T domain that interacts with the PaxB<sup>NDD</sup>. Importantly, the docking interface between PaxA and PaxB is located on the site of the T domain that is opposite to the site for the attachment of the Ppant arm to residue S1027 (Figure 4a). Thus, the docking interaction does not interfere with substrate loading and transfer mediated by the Ppant arm of the  $T_1$  domain. Accordingly, the presence or absence of the Ppant moiety does not influence the affinity of the docking interaction in contrast to observations for the interaction of T domains with other NRPS domains which are often promoted by the presence of the Ppant arm.<sup>[26,29–31]</sup> The high affinity for the docking interaction between PaxA  $T_1$ - $CDD$  and PaxB<sup>NDD</sup> is the result of both electrostatic complementarity (Figure 4b) and favorable matching of extensive hydrophobic surfaces (Figure 4c, Figure S13). Salt bridges are formed for instance between the negatively charged residues E1060 and E1061 in helix  $\alpha 4$  of the PaxA  $T_1$  domain and the positively charged K15 and R14 residues in helix  $\alpha 2$  of the PaxB<sup>NDD</sup>, respectively, and between E1084 of the PaxA  $CDD$  and R16 in helix  $\alpha 2$  of the PaxB<sup>NDD</sup>. The side chain of Q1067 in helix  $\alpha 4$  of the PaxA  $T_1$  domain is involved in an additional polar interaction with the side chain of K22 in helix  $\alpha 2$  of the PaxB<sup>NDD</sup> which also forms a salt bridge with E996 in the  $T_1$  domain. Residues from both helices of the PaxB<sup>NDD</sup> (L6, T7, L9, L11, L17, L18, L20, A21; Figure S13a) show hydrophobic interactions with amino acids in the C-terminal part of helix  $\alpha 4$  and of helix  $\alpha 5$  in the PaxA  $T_1$  domain (I1064, L1068, L1069, F1071, A1073, L1076, I1079; Figure S13b). Residues L1076, I1079 and Y1080 from the PaxA  $CDD$  ( $\alpha 5$ ) add further hydrophobic interactions with both helices of the PaxB<sup>NDD</sup> (L6, L9, L17, L20, A24) to the docking interface (Figure 4c).

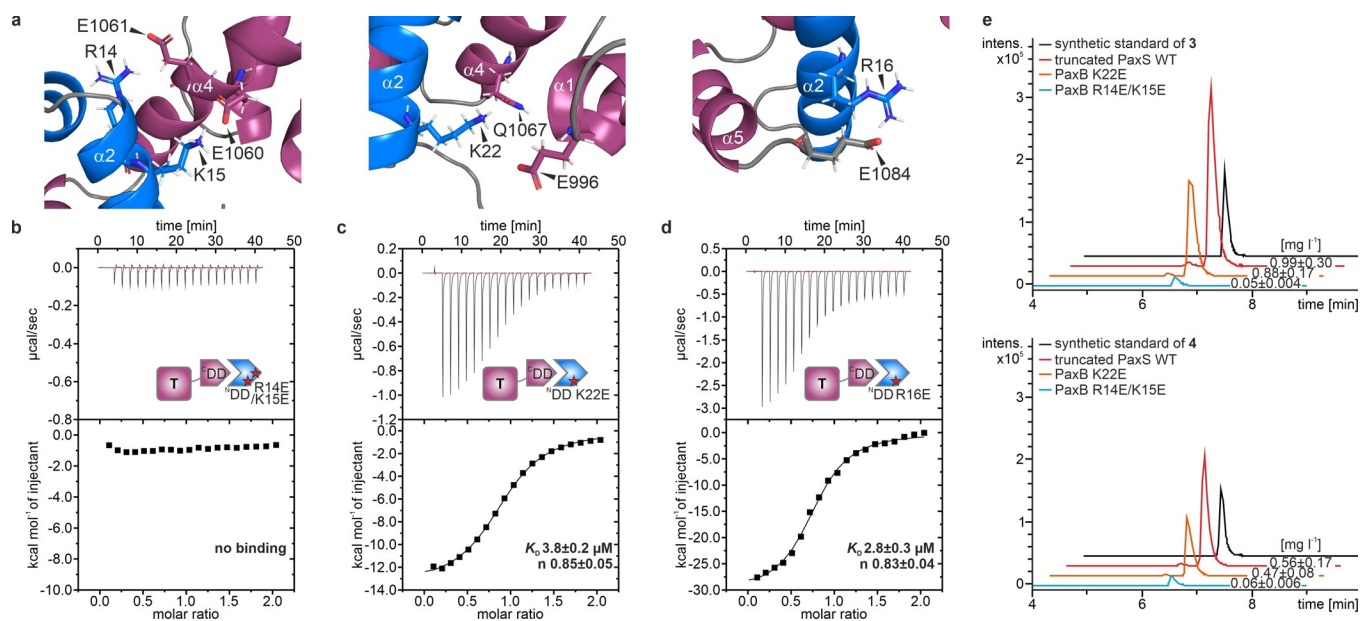
Our structure of the complex suggested that the extensive intermolecular contacts contributed by residues in helix  $\alpha 2$  in the PaxB<sup>NDD</sup> might be dominant for mediating the PaxA/PaxB interaction and helix  $\alpha 1$  might be less important. However, ITC titrations with a PaxB<sup>NDD</sup> variant where residues 1–8 were deleted showed that the absence of helix  $\alpha 1$  lead to a complete loss of the interaction (Figure S14a) suggesting that the full-length PaxB<sup>NDD</sup> is required for a high-affinity docking interaction. The equivalent deletion of helix  $\alpha 1$  in our truncated *X. bovienii* SS-2004 PaxS NRPS (PaxA, PaxB-TE<sub>PaxA</sub>C) overexpressed in *E. coli* strongly reduced product formation in vivo to about 12% for **3** and about 14% for **4** in comparison to the wild type (WT) (Figure S14b–d).

We also asked if the pronounced kink between helix  $\alpha 1$  and  $\alpha 2$  in the PaxB<sup>NDD</sup>, which is induced by the presence of P10 in the sequence and already preformed in the structure of the free PaxB<sup>NDD</sup>, is important for the binding affinity. Therefore, we replaced P10 by a leucine which promotes and stabilizes  $\alpha$ -helical conformations. According to the chemical shift index (<sup>1</sup>H, <sup>15</sup>N, C <sup>$\alpha$</sup> , C <sup>$\beta$</sup> , and C' shifts) the P10L variant adopts a conformation with a continuous straight  $\alpha$ -helix spanning residues L6–K25 (Figure S15a). Accordingly, the CD spectrum of the P10L mutant showed an increased ellipticity at circa 222 nm compared to the WT (Figure 5a).



**Figure 5.** Structural and functional characterization of the PaxB<sup>NDD</sup> P10L variant. a) Circular dichroism spectra of the wild-type PaxB<sup>NDD</sup> (blue) and the PaxB<sup>NDD</sup> P10L variant (purple). b) Exemplary ITC thermogram and the derived binding curve for a titration between the PaxA  $T_1$ - $CDD$  and the PaxB<sup>NDD</sup> P10L variant ( $n=3$ ). c) HPLC/MS data for the characterization of product formation by the modified truncated PaxS carrying the P10L mutation in the PaxB<sup>NDD</sup>, producing peptides **3** (intense EICs) and **4** (pale EICs).

The affinity of the P10L mutant for the PaxA  $T_1$ - $CDD$  di-domain construct in vitro is diminished approximately 16-fold ( $K_D = 3.9 \pm 0.1 \mu\text{M}$ ; Figure 5b, Figure S15b) compared to the WT. Thus, the preformed kink induced by P10 is important for high affinity binding. Interestingly, chemical shift comparisons between the WT complex and the P10L complex under saturating conditions show that the chemical shifts in both complexes are almost identical both for the bound PaxA  $T_1$ - $CDD$  di-domain (Figure S16) and the bound PaxB<sup>NDD</sup> (Figure S15c) despite the presence of the mutation. This suggests that the conformation of both binding partners in the WT and the mutant complex are very similar. In particular, the continuous  $\alpha$ -helix in the free mutant<sup>NDD</sup> P10L apparently breaks and the kinked conformation is adopted upon binding. In agreement with such a scenario, particularly large chemical shift changes are observed for residues T7–E13 upon binding of the PaxB<sup>NDD</sup> P10L mutant to the PaxA  $T_1$ - $CDD$  di-domain (Figure S15a). Thus, P10 preorganizes the conformation of the PaxB<sup>NDD</sup> in a binding-competent conformation. The conformational preorganization in the PaxB<sup>NDD</sup> is apparently also important in vivo for efficient product formation. The introduction of the equivalent P10L



**Figure 6.** The importance of salt bridges for the PaxA T<sub>1</sub>-<sup>C</sup>DD/PaxB <sup>N</sup>DD interaction. In vitro and in vivo analysis of PaxB <sup>N</sup>DD variants. a) The molecular architecture of salt bridge-forming residues that stabilize the cooperative docking domain interface. The PaxA T<sub>1</sub>-<sup>C</sup>DD (purple) and the PaxB <sup>N</sup>DD (blue) are shown in cartoon representation. Side chains of relevant residues are labeled and shown in stick representation. In vitro characterization of PaxB <sup>N</sup>DD b) R14E/K15E, c) K22E and d) R16E variants by ITC. Exemplary ITC thermograms and the derived binding curves for titration between PaxA T<sub>1</sub>-<sup>C</sup>DD and PaxB <sup>N</sup>DD variants. Resulting dissociation constants and binding stoichiometries are given ( $n = 3$ ). e) HPLC/MS data for the characterization of product formation by the modified truncated PaxS carrying the K22E or R14E/K15E mutation(s) in the PaxB <sup>N</sup>DD, producing peptides **3** and **4** (EICs).

mutation in our truncated *X. bovienii* SS-2004 PaxS NRPS (PaxA, PaxB-TE<sub>PaxC</sub>) lead to a clear reduction in product formation in vivo to about 35% for **3** and about 20% for **4** in comparison to the WT (Figure 5c, Figure S17).

The importance of the intermolecular salt bridges for the binding affinity in the PaxA T<sub>1</sub>-<sup>C</sup>DD/PaxB <sup>N</sup>DD complex was investigated in ITC titration experiments with charge reversing point mutations. A PaxB <sup>N</sup>DD R14E/K15E double mutant breaking the salt bridges to E1060 and E1061 in helix α4 of the PaxA T<sub>1</sub> domain led to a complete loss of binding (Figure 6a,b, Figure S18a). The PaxB <sup>N</sup>DD K22E mutant, which abolishes the salt bridge with E996 and the hydrogen bond with Q1067 of the PaxA T<sub>1</sub> domain, diminished the affinity approximately 15-fold ( $K_D = 3.8 \pm 0.2 \mu\text{M}$ ; Figure 6a,c, Figure S18b) while the R16E mutant, destroying the salt bridge with E1084 in the PaxA <sup>C</sup>DD, decreased the binding affinity approximately 11-fold ( $K_D = 2.8 \pm 0.3 \mu\text{M}$ ; Figure 6a,d, Figure S18c).

Importantly, all PaxB <sup>N</sup>DD mutants discussed here showed an α-helical content similar to the WT confirming that the mutations did not interfere with the structure of the unbound PaxB <sup>N</sup>DD (Figure S19). Overall, the pronounced effects of the salt bridge abrogating mutations on binding demonstrated that these interactions are important for establishing the high affinity interaction between PaxA and PaxB in vitro. The introduction of the PaxB <sup>N</sup>DD R14E/K15E double mutant also strongly inhibited product formation for the truncated *X. bovienii* SS-2004 PaxS NRPS (PaxA, PaxB-TE<sub>PaxC</sub>) in vivo. Only approximately 5%/10% of the amount of **3/4** formed by the WT were detected for this mutant (Figure 6e). The effect of introducing the K22E mutant in this

cluster on product formation in vivo was rather limited. The mutant NRPS still produced approximately 90%/85% of **3/4** compared to the WT (Figure 6e). However, the  $K_D$  for the K22E mutant in vitro is still in the low μM range ( $K_D = 3.8 \mu\text{M}$ ) and its affinity therefore higher than for many other previously described functional docking domain interactions.<sup>[8,12,14]</sup> This might explain why still significant amounts of product are formed particularly under conditions of heterologous protein overexpression.

The extended docking interface observed here for the PaxS NRPS between PaxA and PaxB where the T domain and a minimal <sup>C</sup>DD must cooperate to bind with high affinity to the PaxB <sup>N</sup>DD is not limited to PaxS-related NRPS systems from *Xenorhabdus* species. A simple BlastP search of NCBI's non-redundant protein sequence database<sup>[32]</sup> using the sequences of the PaxA T<sub>1</sub>-<sup>C</sup>DD and the PaxB <sup>N</sup>DD from *X. cabanillasii* and *X. bovienii* as input revealed a number of hits to putative multiprotein NRPS systems in different classes of the proteobacteria as well as in cyanobacteria (Figure S20). In these hits, the predicted C-terminal T domains feature long α4 helices (ca. 4-turns) in one protein chain of the NRPS gene cluster which precede short <sup>C</sup>DDs while a circa 35 amino acids long potential <sup>N</sup>DD with α-helical secondary structure propensity in another protein chain precedes a C domain. Furthermore, the α4 helix regions of the T domains and the putative <sup>C</sup>DD regions are in general rich in negatively charged residues whereas the predicted <sup>N</sup>DDs contain a number of positively charged residues (Figure S20a). Unfortunately, none of the putative NRPS systems found in this search were functionally characterized or have known peptide products (Figure S20b). For a hit from

the  $\beta$ -proteobacterium *Chromobacterium violaceum* Bergonzini we expressed and purified the PaxA T<sub>1</sub>-<sup>C</sup>DD di-domain analog and the PaxB <sup>C</sup>DD analog. In an NMR titration experiment the addition of the unlabeled PaxB <sup>N</sup>DD analog to the <sup>15</sup>N-labeled PaxA T<sub>1</sub>-<sup>C</sup>DD di-domain analog caused widespread chemical shift perturbations indicative of binding (Figure S21a). In contrast, NMR titration experiments of an isolated <sup>15</sup>N-labeled PaxA T<sub>1</sub> domain analog (Figure S21b) or an isolated PaxA <sup>C</sup>DD analog (Figure S21c) with the PaxB <sup>N</sup>DD analog showed no obvious evidence for an interaction. Thus, also in this system the T domain and the <sup>C</sup>DD must cooperate to bind to their cognate <sup>N</sup>DD suggesting that this type of composite docking interface is widespread in NRPS and NRPS-PKS systems from different classes of proteobacteria and also occurs in additional bacterial phyla such as the cyanobacteria (Figure S20).

## Conclusion

Previous structural and functional studies of docking domains in megasynthases have not only uncovered their rich structural diversity but also established them as interesting building blocks for megasynthase engineering since they are able to mediate non-covalent protein–protein interactions in multiprotein NRPS or PKS systems independently from other domains and can be transplanted between systems without loss of function.<sup>[9–11]</sup> While the docking domain interactions are specific for cognate pairs of N-terminal and C-terminal docking domains the affinity of these interactions is often limited to  $K_D$  values in the mid- and low-micromolar range. Here, we have identified a novel type of docking interaction in a multiprotein NRPS system where the C-terminal T domain and a short C-terminal docking domain of one NRPS protein cooperate to create a binding interface for the N-terminal docking domain of another NRPS protein to non-covalently assemble a productive NRPS complex. In comparison to previously described docking interactions the affinity of this interaction is significantly higher with a  $K_D$  in the nanomolar range. The T domain takes part in the docking interface based on the presence of an elongated helix  $\alpha_4$  where the C-terminal part of this extended helix contributes to the docking interaction. Since this helix is part of the canonical carrier protein fold it is structurally pre-organized already prior to docking. The <sup>N</sup>DD also shows a high degree of structural preorganization in its free form and only the rather short <sup>C</sup>DD behaves as an IDP in contrast to previous observations in 4  $\alpha$ -helix bundle (4HB) docking domains.<sup>[14,33]</sup> This high degree of structural pre-organization is a likely reason for the high affinity of the interaction described here. Importantly, the involvement of helix  $\alpha_4$  in the docking interaction does not interfere with the other functions of the T domain in substrate shuttling and the structural elements often involved in interactions of T domains with other NRPS domains such as helices  $\alpha_2$  and  $\alpha_3$ <sup>[28,34]</sup> are still accessible upon binding the <sup>N</sup>DD.

Docking interactions involving T domains such as the one described here are apparently not limited to the PaxS NRPS but also occur in NRPS and NRPS-PKS systems from

organisms not related to *Xenorhabdus*. Due to the high affinities observed for these systems they might be attractive tools for mediating non-native intermolecular interactions in “designer” NRPS systems. However, their functional introduction into non-native contexts will be more challenging compared to free-standing docking domains since it will involve the re-engineering of the sequence and structure of helix  $\alpha_4$  of the T domain while avoiding interference with the T domain fold and its other functions. Fortunately, the contacts between helix  $\alpha_4$  of the T domain and the <sup>N</sup>DD are limited to its C-terminal extension, which does not interact with the core of the T domain. Thus, it might be possible to extend helix  $\alpha_4$  and add the <sup>C</sup>DD in more conventional T domains to create a suitable docking interface for the <sup>N</sup>DD in artificial NRPS. Most likely, the high affinity provided by this type of docking interface will be beneficial for the efficiency of the assembled NRPS systems.

## Acknowledgements

All NMR measurements were carried out at the Center for Biomolecular Magnetic Resonance (BMRZ) at the Goethe University Frankfurt which is generously supported by the State of Hesse. This work was further supported by the LOEWE program of the state of Hesse and was conducted within the framework of the MegaSyn Research Cluster in the laboratories of H.B.B. and J.W. Work in the Bode lab is also supported by an ERC Advanced Grant (grant agreement number 835108). Open access funding enabled and organized by Projekt DEAL.

## Conflict of interest

The authors declare no conflict of interest.

**Keywords:** communication-mediating domains · docking domains · non-ribosomal peptide synthetase · peptide-antimicrobial-*Xenorhabdus* peptide · thiolation domain

- [1] R. D. Süssmuth, A. Mainz, *Angew. Chem. Int. Ed.* **2017**, *56*, 3770; *Angew. Chem.* **2017**, *129*, 3824.
- [2] J. M. Reimer, A. S. Haque, M. J. Tarry, T. M. Schmeing, *Curr. Opin. Struct. Biol.* **2018**, *49*, 104.
- [3] H. B. Bode, A. O. Brachmann, K. B. Jadhav, L. Seyfarth, C. Dauth, S. W. Fuchs, M. Kaiser, N. R. Waterfield, H. Sack, S. H. Heinemann, et al., *Angew. Chem. Int. Ed.* **2015**, *54*, 10352; *Angew. Chem.* **2015**, *127*, 10493.
- [4] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *ChemBioChem* **2002**, *3*, 490.
- [5] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15585.
- [6] R. Gokhale, *Curr. Opin. Chem. Biol.* **2000**, *4*, 22.
- [7] R. Broadhurst, D. Nietlispach, M. P. Wheatcroft, P. F. Leadlay, K. J. Weissman, *Chem. Biol.* **2003**, *10*, 723.
- [8] C. Hacker, X. Cai, C. Kegler, L. Zhao, A. K. Weickhmann, J. P. Wurm, H. B. Bode, J. Wöhnert, *Nat. Commun.* **2018**, *9*, 4366.
- [9] C. Kegler, H. B. Bode, *Angew. Chem. Int. Ed.* **2020**, *59*, 13463; *Angew. Chem.* **2020**, *132*, 13565.
- [10] X. Cai, L. Zhao, H. B. Bode, *Org. Lett.* **2019**, *21*, 2116.

- [11] J. R. Whicher, S. S. Smaga, D. A. Hansen, W. C. Brown, W. H. Gerwick, D. H. Sherman, J. L. Smith, *Chem. Biol.* **2013**, *20*, 1340.
- [12] J. Watzel, C. Hacker, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *ACS Chem. Biol.* **2020**, *15*, 982.
- [13] T. J. Buchholz, T. W. Geders, F. E. Bartley, K. A. Reynolds, J. L. Smith, D. H. Sherman, *ACS Chem. Biol.* **2009**, *4*, 41.
- [14] J. Dorival, T. Annaval, F. Risser, S. Collin, P. Roblin, C. Jacob, A. Gruez, B. Chagot, K. J. Weissman, *J. Am. Chem. Soc.* **2016**, *138*, 4155.
- [15] S. Kosol, A. Gallo, D. Griffiths, T. R. Valentic, J. Masschelein, M. Jenner, E. L. C. de Los Santos, L. Manzi, P. K. Sydor, D. Rea, et al., *Nat. Chem.* **2019**, *11*, 913.
- [16] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 275.
- [17] F. Risser, S. Collin, R. Dos Santos-Morais, A. Gruez, B. Chagot, K. J. Weissman, *J. Struct. Biol.* **2020**, *212*, 107581.
- [18] S. W. Fuchs, A. Proschak, T. W. Jaskolla, M. Karas, H. B. Bode, *Org. Biomol. Chem.* **2011**, *9*, 3130.
- [19] T. D. Vo, C. Spahn, M. Heilemann, H. B. Bode, *ACS Chem. Biol.* **2021**, *16*, 447–451.
- [20] J. M. Chaston, G. Suen, S. L. Tucker, A. W. Andersen, A. Bhasin, E. Bode, H. B. Bode, A. O. Brachmann, C. E. Cowles, K. N. Cowles, et al., *PLoS One* **2011**, *6*, e27909.
- [21] O. Schimming, F. Fleischhacker, F. I. Nollmann, H. B. Bode, *ChemBioChem* **2014**, *15*, 1290.
- [22] J. Watzel, S. Sarawi, E. Duchardt-Ferner, H. B. Bode, J. Wöhnert, *Biomol. NMR Assignments* **2021**, *15*, 229.
- [23] Y. Shen, A. Bax, *J. Biomol. NMR* **2013**, *56*, 227.
- [24] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407.
- [25] A. C. Goodrich, B. J. Harden, D. P. Frueh, *J. Am. Chem. Soc.* **2015**, *137*, 12100.
- [26] C. A. Mitchell, C. Shi, C. C. Aldrich, A. M. Gulick, *Biochemistry* **2012**, *51*, 3252.
- [27] M. Schubert, D. Labudde, H. Oschkinat, P. Schmieder, *J. Biomol. NMR* **2002**, *24*, 149.
- [28] J. R. Lohman, M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips, B. Shen, *Proteins Struct. Funct. Bioinf.* **2014**, *82*, 1210.
- [29] Y. Liu, T. Zheng, S. D. Bruner, *Chem. Biol.* **2011**, *18*, 1482.
- [30] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. L. Allen, G. Skinotitis, C. C. Aldrich, A. M. Gulick, *Nature* **2016**, *529*, 235.
- [31] K. Haslinger, C. Brieke, S. Uhlmann, L. Sieverling, R. D. Süßmuth, M. J. Cryle, *Angew. Chem. Int. Ed.* **2014**, *53*, 8518; *Angew. Chem.* **2014**, *126*, 8658.
- [32] S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D. J. Lipman, *Nucleic Acids Res.* **1997**, *25*, 3389.
- [33] J. Zeng, D. T. Wagner, Z. Zhang, L. Moretto, J. D. Addison, A. T. Keatinge-Clay, *ACS Chem. Biol.* **2016**, *11*, 2466.
- [34] J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, *Chem. Biol.* **2012**, *19*, 188.

Manuscript received: March 10, 2021  
Accepted manuscript online: April 19, 2021  
Version of record online: May 14, 2021