



# DNA-binding and protein structure of nuclear factors likely acting in genetic information processing in the *Paulinella* chromatophore

Luis Macorano<sup>a</sup> , Taniya M. Binny<sup>a</sup> , Tobias Spiegl<sup>a</sup> , Victoria Klimenko<sup>a</sup> , Anna Singer<sup>a</sup>, Linda Oberleitner<sup>a</sup>, Violetta Applegate<sup>b</sup> , Sarah Seyffert<sup>a</sup>, Anja Stefanski<sup>c</sup> , Lothar Gremer<sup>d,e</sup> , Christoph G. W. Gertzen<sup>b,f</sup> , Astrid Höppner<sup>b</sup>, Sander H. J. Smits<sup>b,g</sup> , and Eva C. M. Nowack<sup>a,1</sup>

Edited by Éva Kondorosi, Hungarian Academy of Sciences, Biological Research Centre, Szeged, Hungary; received December 20, 2022; accepted May 25, 2023

The chromatophores in *Paulinella* are evolutionary-early-stage photosynthetic organelles. Biological processes in chromatophores depend on a combination of chromatophore and nucleus-encoded proteins. Interestingly, besides proteins carrying chromatophore-targeting signals, a large arsenal of short chromatophore-targeted proteins (sCTPs; <90 amino acids) without recognizable targeting signals were found in chromatophores. This situation resembles endosymbionts in plants and insects that are manipulated by host-derived antimicrobial peptides. Previously, we identified an expanded family of sCTPs of unknown function, named here “DNA-binding (DB)-sCTPs”. DB-sCTPs contain a ~45 amino acid motif that is conserved in some bacterial proteins with predicted functions in DNA processing. Here, we explored antimicrobial activity, DNA-binding capacity, and structures of three purified recombinant DB-sCTPs. All three proteins exhibited antimicrobial activity against bacteria involving membrane permeabilization, and bound to bacterial lipids in vitro. A combination of in vitro assays demonstrated binding of recombinant DB-sCTPs to chromatophore-derived genomic DNA sequences with an affinity in the low nM range. Additionally, we report the 1.2 Å crystal structure of one DB-sCTP. In silico docking studies suggest that helix  $\alpha 2$  inserts into the DNA major groove and the exposed residues, that are highly variable between different DB-sCTPs, confer interaction with the DNA bases. Identification of photosystem II subunit CP43 as a potential interaction partner of one DB-sCTP, suggests DB-sCTPs to be involved in more complex regulatory mechanisms. We hypothesize that membrane binding of DB-sCTPs is related to their import into chromatophores. Once inside, they interact with the chromatophore genome potentially providing nuclear control over genetic information processing.

endosymbiosis | effector proteins | antimicrobial peptides | DNA-binding protein | host control

The cercozoan amoeba *Paulinella chromatophora* contains evolutionary-early-stage photosynthetic organelles of cyanobacterial origin termed “chromatophores” (1, 2). Despite their relatively recent origin (~100 Mya) compared to primary plastids that evolved ~1.6 billion years ago in the Archaeplastida, the chromatophores reached an astonishing level of cellular integration. Chromatophores are ~140 times larger in volume compared to their free-living relatives, and each *Paulinella* cell contains strictly two chromatophores of which one segregates to the daughter cell upon host cell division and duplicates shortly after (3, 4). The functional replacement of chromatophore genes by nuclear genes resulted in metabolic pathways and multiprotein complexes of dual genetic origin in the chromatophore (e.g., photosystem I contains two nucleus-encoded subunits, PsaE and PsaK) (5, 6). The tight interconnection of host and chromatophore biological networks suggests coordination between replication of their genomes and gene expression, especially of nuclear and chromatophore genes encoding proteins that are involved in the same biological process or multiprotein complex.

The chromatophore is delimited by two membranes and a peptidoglycan wall. Whereas the cytoplasmic membrane is of cyanobacterial origin, the outer membrane is likely host-derived. Nucleus-encoded proteins that translocate across this “symbiotic interface” fall into two classes. Whereas long chromatophore-targeted proteins [ICTPs; ~>250 amino acids (aa)] contain a conserved bipartite N-terminal targeting signal that likely mediates protein trafficking via the secretory pathway to the chromatophore, short chromatophore-targeted proteins (sCTPs; <90 aa) lack recognizable targeting signals (6, 7). The mechanism underlying import of sCTPs is not well understood. Despite the lack of N-terminal signal peptides (SPs) that typically mediate cotranslational sorting of proteins into the secretory pathway, immuno

## Significance

The chromatophore of *Paulinella* represents a unique model to study early-stages in the complex transformation process from bacterial endosymbiont to semiautonomous organelle. Despite their relatively recent origin ~100 Mya, chromatophores reached an astonishing level of cellular integration. Chromatophore division is coordinated with the host cell cycle, and biological processes in the chromatophore depend on chimeric metabolic pathways and multiprotein complexes composed of proteins of dual genetic origin. This suggests that mechanisms evolved that synchronize genome replication and gene expression between chromatophore and nucleus. Here, we identified an expanded protein family of short chromatophore-targeted proteins, comprising ~200 members, that represent promising nuclear factors that might be involved in nuclear control of genetic information processing in the evolving organelle.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2023 the Author(s). Published by PNAS. This article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](#).

<sup>1</sup>To whom correspondence may be addressed. Email: e.nowack@hhu.de.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221595120/-DCSupplemental>.

Published June 26, 2023.

gold analyses suggest vesicular transport through the Golgi to be involved in translocating sCTPs across the outer chromatophore membrane (5).

Although some sCTPs have functions related to photosynthesis (e.g., PsaE and PsaK), most sCTPs are orphan proteins with unknown cellular functions. Interestingly, short secreted proteins that are specifically expressed in bacteriocytes (i.e., the symbiont-housing cells) have been reported in several other unrelated endosymbiotic associations. These include the expanded arsenal of several hundred nodule-specific cysteine-rich (NCR) peptides in some legumes (8–10), the bacteriocyte-specific cysteine-rich (BCR) peptides in aphids (11), defensin-like peptides in the *Alnus* tree, and coleopteracin-A in the grain weevil *Sitophilus* (12). These short proteins apparently are targeted to the symbiotic bacteria and, at high concentrations, show antimicrobial activity against diverse bacteria in *in vitro* assays (12–15). Therefore, these short symbiont-targeted proteins have been collectively referred to as “symbiotic antimicrobial peptides” (16). Their precise functions within the symbiotic associations (at sublethal concentrations) are only beginning to be characterized and seem to range from inhibition of endosymbiont division resulting in gigantism and polyploidy of the endosymbionts (12, 17, 18), spatial confinement of the bacteria to the symbiotic organ (12), modulating permeability of endosymbiont membranes (13), metabolic interference (19), to changing expression profiles in the target bacteria (20).

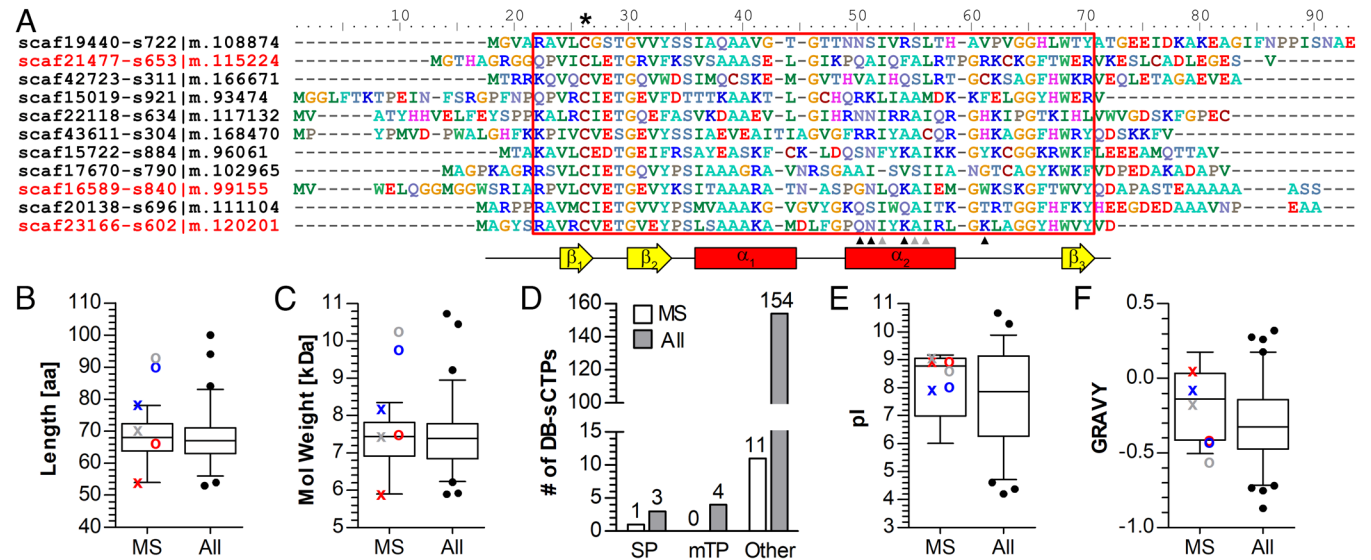
Scrutiny of sCTPs in *P. chromatophora* revealed that many of these short proteins cluster into four extended protein groups (21). Similar to NCRs and BCRs, members of sCTP “group 3” and “group 4” (as defined in ref. 21) are characterized by the occurrence of multiple repetitive Cys motifs (CxxC or CxxxxC). Most members of sCTP “group 2” that are the focus of this study, contain one Cys in a conserved position close to the N terminus, and sometimes one or more additional Cys in variable positions but all feature a conserved ~45-aa-long sequence motif (Fig. 1A

and *SI Appendix, Fig. S1A*). Interestingly, this motif is also found in a number of bacterial and viral proteins that contain other domains related to DNA processing (e.g., domains annotated as “group I intron endonuclease domains” or “Superfamily II DNA or RNA helicase domains”) (21), suggesting that the conserved motif could function in protein–DNA interaction.

Here we show that, similar to previously described symbiotic AMPs, sCTPs display antimicrobial activity against *Escherichia coli* and additionally, bind *in vitro* to bacterial membrane lipids. However, the bacterial membrane does not seem to be the main site of function of sCTPs. Instead, revealing the structure of one representative group 2 sCTP by X-ray crystallography at atomic resolution and different *in vitro* assays characterize group 2 sCTPs as proteins that are capable to bind to the chromatophore genome – likely in a sequence-specific manner. Thus, we name these proteins DNA-binding (DB-)sCTPs here and hypothesize that they represent nuclear factors involved in genetic information processing in the chromatophore.

Results

**In Silico Analysis of DB-sCTPs.** Previous protein mass spectrometric (MS) analyses detected 12 DB-sCTPs in chromatophores (Fig. 1A) (21). Screening the available *P. chromatophora* transcriptome dataset (GenBank: GEZN00000000.1; ref. 22) revealed DB-sCTPs to form an expanded protein family containing at least 204 members (21, 22). For 161 of these proteins full-length sequence information is available (*SI Appendix, Fig. S1B* and *Dataset S1*). MS-identified DB-sCTPs range in size from 54 to 78 aa (Fig. 1B) corresponding to 5.9–8.4 kDa (Fig. 1C). The whole DB-sCTP family including all 161 full-length members shows overall a similar size range (53 to 84 aa and 5.9 to 9.2 kDa); only two proteins of 94 and 100 aa length exceed the previously observed length cutoff of <90 aa for sCTPs. These two outliers show an N-terminal



**Fig. 1.** Amino acid sequence alignment and predicted characteristics of DB-sCTPs. (A) Clustal-Ω alignment of 11 out of 12 MS-identified DB-sCTPs. (Scaffold16209-size858 | m.97815 was excluded from the alignment due to an unusual 9-aa insertion within the conserved motif.) The three experimentally studied DB-sCTPs have red sequence identifiers. Red frame, conserved motif (see also *SI Appendix, Fig. S1A*). The asterisk highlights the conserved Cys. Secondary structural elements identified by X-ray crystallography (see text below) are mapped on the alignment (yellow,  $\beta$ -strand; red,  $\alpha$ -helix). Black arrowheads, exposed residues of helix  $\alpha_2$  predicted to interact with the DNA; gray arrowheads, more conserved hydrophobic residues of  $\alpha_2$  facing inward. (B–F) Box plots and bar graph representing distribution of protein lengths (B), molecular weights (C), subcellular localization predicted by TargetP (SP, signal peptide; mTP, mitochondrial transit peptide; O, other localization) (D); isoelectric points (pI) (E), and GRAVY scores (i.e., sum of hydrophathy values of all aa divided by the protein length) (F) for the 12 MS-identified DB-sCTPs (MS) and all 161 full-length DB-sCTPs (All). Whiskers in box plots represent the 2.5 and 97.5 percentile. Outliers are represented by dots. Values for the natural and recombinant versions of the experimentally studied DB-sCTPs sCTP-16589, sCTP-21477, and sCTP-23166 are indicated by blue, gray, and red crosses (natural) and circles (recombinant), respectively.

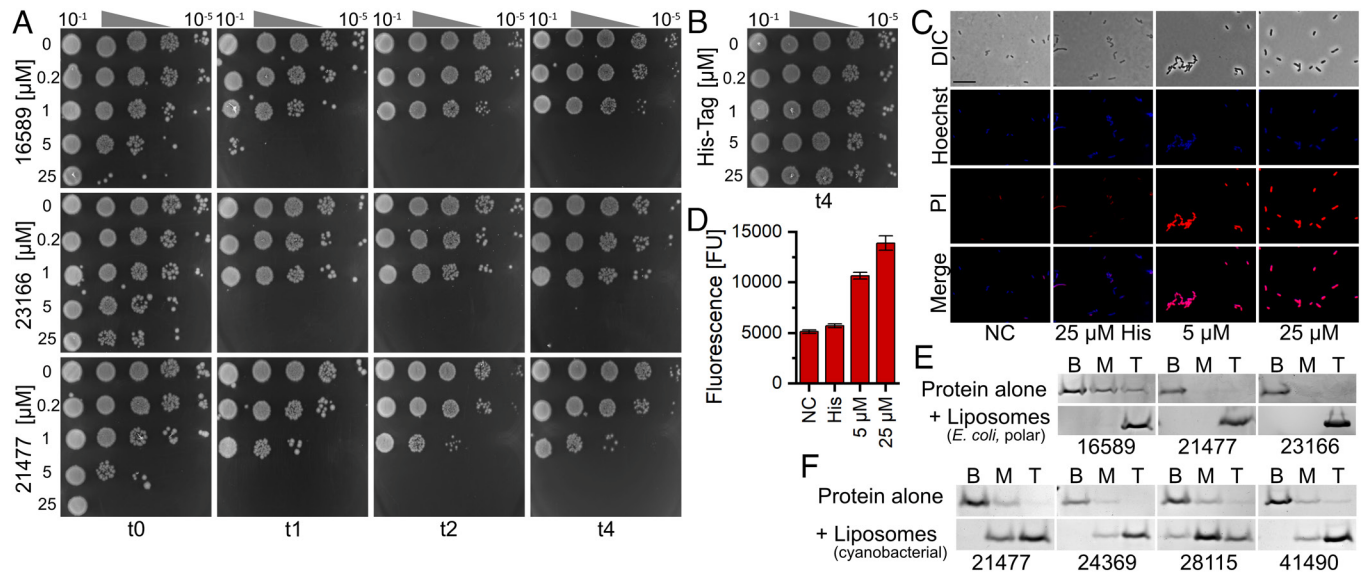
extension of ~30 aa (SI Appendix, Fig. S1B). As noted before for sCTPs in general, predicted signal peptides (SPs) or mitochondrial transit peptides (mTPs) are largely missing in DB-sCTPs. Only for 7 out of 161 predicted DB-sCTPs, mostly weak predictions of mTPs or SPs were obtained with TargetP (23) of which almost all reach well into the conserved motif, suggesting that they represent false-positive predictions (Fig. 1D and SI Appendix, Fig. S1B). Most MS-identified as well as predicted DB-sCTPs are cationic (75% and 60%, respectively;  $pI > 7.5$ ); however, both groups encompass also neutral ( $6.5 \leq pI \leq 7.5$ ) or anionic proteins ( $pI < 6.5$ ) (Fig. 1E). The grand average hydropathy of most DB-sCTPs ranges from -0.45 to 0 characterizing DB-sCTPs as overall hydrophilic (Fig. 1F).

**DB-sCTPs Show Antimicrobial Activity against *E. coli* and Bind to Bacterial Lipids In Vitro.** As *P. chromatophora* is as of yet not amenable to genetic manipulations, we aimed to characterize DB-sCTPs in in vitro assays. To this end, three representative DB-sCTPs (scaffold16589-size840|m.99155, scaffold21477-size653|m.11522, and scaffold23166-size602|m.12020; from here on sCTP-16589, sCTP-21477, and sCTP-23166, respectively; see Fig. 1) were heterologously expressed as recombinant proteins with an N-terminal decahistidin tag (His<sub>10</sub>) in *E. coli*. The recombinant, purified proteins (Material and Methods and SI Appendix, Fig. S2A) were used for all further experiments. If not indicated otherwise, 50 mM MES pH 5.5 was used as buffer system in these experiments, in which all three purified proteins were stable (SI Appendix, Fig. S2B). To test whether DB-sCTPs exhibit antimicrobial activity under conditions similar to those previously used for symbiotic AMPs (e.g., refs. 14, 18, 24, and 25), the three recombinant DB-sCTPs were tested in confrontation assays with *E. coli*. At concentrations in the low  $\mu$ M range, all three DB-sCTPs showed a time- and dose-dependent antimicrobial effect in a drop-plate assay and efficiently killed the cells at concentrations of 5  $\mu$ M (corresponding to

~100 pg protein per *E. coli* cell); sCTP-21477 and sCTP-16589 already markedly reduced colony formation at concentrations of 1  $\mu$ M (Fig. 2A). Importantly, 25  $\mu$ M of the His-tag alone that was used as a control, did not show any antimicrobial activity under the same experimental conditions after 4-h incubation time (Fig. 2B). However, the observed antimicrobial activity does not seem to be a unique feature of DB-sCTPs as similar activities were seen for other soluble sCTPs used in the same confrontation assays [sCTP-28115 (member of sCTP group 3), sCTP-24369 (member of sCTP group 4), and sCTP-41490 (annotated as 4-oxalocrotonate tautomerase; not member of any sCTP group)] (SI Appendix, Fig. S3A and B).

Fluorescence analyses of PI-stained DB-sCTP-treated *E. coli* cells revealed that antimicrobial activity was accompanied by the permeabilization of the cytoplasmic membrane (Fig. 2C and D). Whereas Hoechst 33342 stains both live and dead cells, PI is impermeable to the intact cytoplasmic membrane. Both, fluorescence microscopy as well as fluorescence quantification in a plate reader clearly showed an increase in membrane permeability caused by the treatment with 5  $\mu$ M or 25  $\mu$ M sCTP-16589. PI-stained *E. coli* cells treated with the His-tag alone showed fluorescence intensities comparable to nontreated cells. Similar results were obtained for the other two DB-sCTPs.

To explore whether sCTPs can directly bind to membrane lipids, sCTPs from different groups (sCTP-21477, sCTP-28115, sCTP-24369, and sCTP-41490; for group affiliation see above) were tested in protein-lipid overlay assays (SI Appendix, Fig. S3C). All four proteins bound to a very similar set of lipids with a clear preference to negatively charged lipids (phosphoinositides, phosphatidic acid, phosphatidylserine, and cardiolipin). However, binding-preferences appear to be more specific as the negatively charged phosphatidyl glycerol (PG) was not or only very weakly bound. The strong binding of all four proteins to cardiolipin suggested that sCTPs can bind to bacterial membranes, which



**Fig. 2.** Antimicrobial activity against *E. coli* and binding of DB-sCTPs to bacterial membrane lipids. (A) *E. coli* cells were incubated with DB-sCTPs at indicated concentrations for 0, 1, 2, and 4 h, spotted onto LB agar plates at ten-fold dilutions from OD<sub>600</sub> of 10<sup>-1</sup> to 10<sup>-5</sup> and then grown overnight at 37 °C. (B) Same assay as in A using the His-tag alone as a negative control. (C) Epifluorescence microscopic analysis of *E. coli* cells that were incubated with 25  $\mu$ M His-tag or 5  $\mu$ M or 25  $\mu$ M sCTP-16589 for 1 h and stained with Hoechst 33342 (Hoechst) and propidium iodide (PI). Mock-treated cells were used as a negative control (NC). DIC, differential interference contrast; PI, red channel; Hoechst, blue channel; Merge, merge of both fluorescence channels. The scale bar is 10  $\mu$ m. (D) Fluorescence in arbitrary fluorescence units [FU] of PI-stained *E. coli* cells treated with no protein, His-tag or sCTP-16589 as in C. Mean and SD of three independent samples are displayed. (E and F) Floation assays of sCTPs in sucrose gradients with and without liposomes. Following ultracentrifugation, the gradients were fractionated into bottom (B), middle (M), and top fraction (T) and protein recovered subjected to SDS-PAGE. The Coomassie-stained gels are displayed. (E) Binding of three different DB-sCTPs to liposomes of *E. coli* polar lipids. (F) Binding of sCTPs from different groups to liposomes of cyanobacterial lipids.

was confirmed by floatation assays. Whereas the proteins alone remained in the bottom fraction of the discontinuous sucrose gradient during ultracentrifugation, addition of liposomes, generated from *E. coli* polar lipid extracts, resulted in the floatation of the protein with the liposomes to the top fraction (Fig. 2E). Importantly, the same binding behavior was observed for diverse sCTPs using liposomes mimicking cyanobacterial membranes, which contain the three glycolipids monogalactosyl diacylglycerol (MGDG), digalactosyl diacylglycerol (DGDG), and sulfoquinovosyl diacylglycerol (SQDG), plus PG, but lack cardiolipin (26) (Fig. 2F).

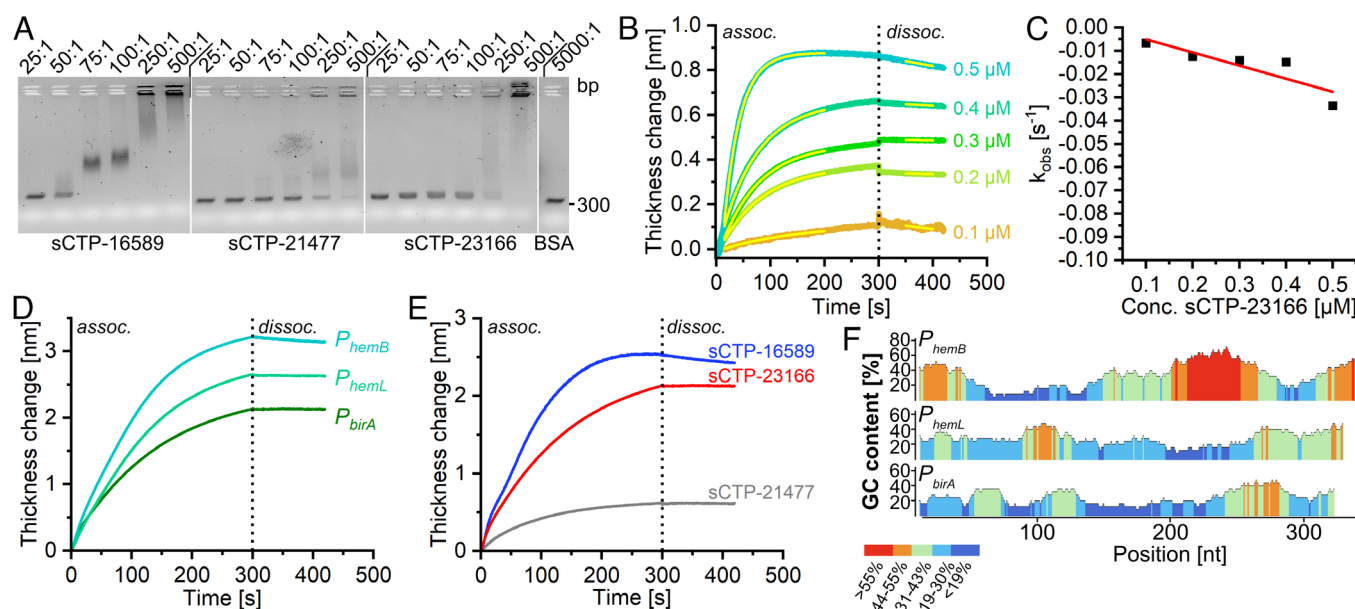
**DB-sCTPs Bind with High Affinity to Chromatophore Genomic DNA In Vitro.** Given the observation of a conserved sequence motif in DB-sCTPs (Fig. 1A) that also occurs in various DNA-processing proteins in bacteria (21), we hypothesized that the bacterial membrane was not the primary target of DB-sCTPs, but following membrane translocation, DB-sCTPs bind to DNA targets in the chromatophore. In line with this hypothesis, we could pull down DB-sCTPs from *P. chromatophora* lysate using  $P_{hemB}$ , a DNA fragment surrounding the promoter region for the chromatophore *hemB* gene (encoding the  $\delta$ -aminolevulinic acid dehydratase involved in chlorophyll biosynthesis) as a bait. MS analysis of the eluate identified among 41 proteins five DB-sCTPs, among them was the previously studied sCTP-23166. Other proteins that were identified included for instance several ribosomal proteins, the  $\alpha$ -subunit of the chromatophore-encoded RNA polymerase, a histone, and highly expressed enzymes of the central carbon metabolism of the *P. chromatophora* host cell (Dataset S2).

To confirm DNA-binding capacity of DB-sCTPs, we performed electrophoretic mobility shift assays (EMSAs) using the  $P_{hemB}$  fragment (Fig. 3A). At molar ratios of 75:1 (protein:DNA) for sCTP-16589 and 250:1 for sCTP-21477 and sCTP-23166, the

recombinant proteins bound to the DNA indicated by the shift of the band. Bovine serum albumin (BSA) that was used as a negative control, did not show DNA binding at a 5,000:1 molar ratio.

Since the hypothesized DNA-binding potential of all three DB-sCTPs analyzed was confirmed by EMSA, BLI experiments were developed to characterize the binding kinetics in more detail (for the optimization of the experimental procedure see SI Appendix, Fig. S4). The binding affinity of sCTP-23166 to the immobilized  $P_{hemB}$  was investigated by recording association and dissociation curves at protein concentrations ranging from saturation to close to the detection limit (0.5 to 0.1  $\mu$ M, respectively) (Fig. 3B and C). Determined association and dissociation rates ( $k_{off}$  of  $1.95 \times 10^{-4} \pm 0.87 \times 10^{-5} \text{ s}^{-1}$ ;  $k_{on}$  of  $56,380 \pm 17,910 \text{ M}^{-1} \text{ s}^{-1}$ ) (Fig. 3B and C) yielded a  $K_D$  of  $3.4 \pm 1.2 \text{ nM}$ , which identified sCTP-23166 as a high-affinity dsDNA-binding protein. Using further two promoter regions of the chromatophore genes *hemL* and *birA* (encoding a glutamate-1-semialdehyde 2,1-aminomutase involved in chlorophyll biosynthesis and a biotin-acetyl-CoA-carboxylase ligase involved in transcriptional control of biotin biosynthetic pathway enzymes, respectively) in BLI analyses (Fig. 3D–F) demonstrated that sCTP-23166 binds at a fixed concentration with different affinities to different promoter regions (Fig. 3D). Finally, the three different DB-sCTPs (sCTP-23166, sCTP-21477, and sCTP-16589) bind with markedly different affinities to the same promoter region (here:  $P_{birA}$ ; Fig. 3E).

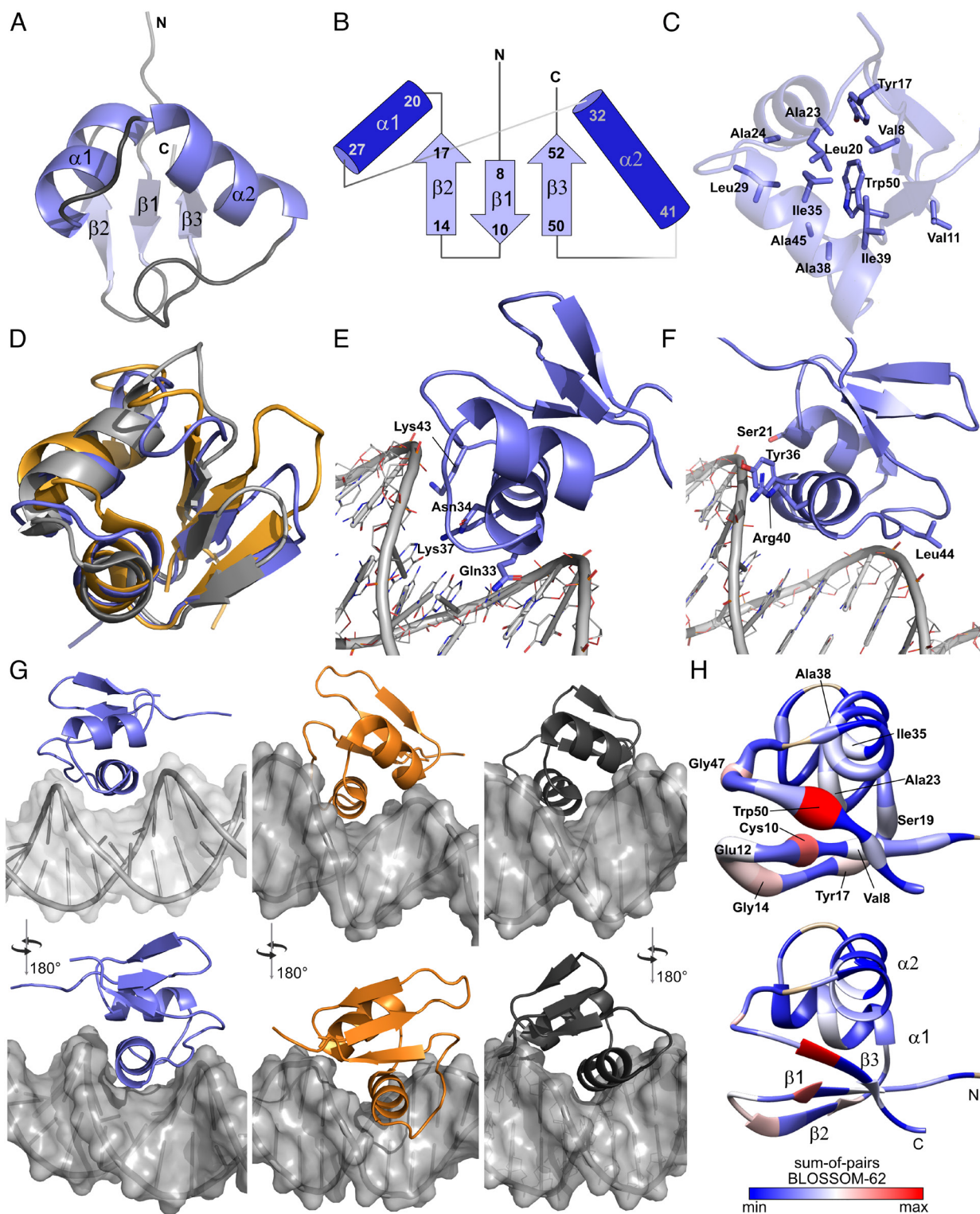
**3D Structure Determination by X-ray Crystallography.** To investigate the structural basis for the observed high binding affinity of sCTP-23166 to dsDNA, we determined the 3D structure of the protein by X-ray crystallography. We obtained a structure at 1.2 Å resolution (pdb-code: 8B6E) with two monomers present in the asymmetric unit, covering the whole protein sequence. Only the first two N-terminal aa show no clear electron density indicating that they are structurally flexible



**Fig. 3.** DB-sCTPs are DNA-binding proteins. (A) EMSA with the three DB-sCTPs and the  $P_{hemB}$  fragment at molar ratios from 25:1 to 500:1 (protein:DNA). BSA:  $P_{hemB}$  (5,000:1) was used as a control. (B) Binding of sCTP-23166 to  $P_{hemB}$  immobilized on SAX BLI sensors. The changes in thickness generated by the binding of sCTP-23166 at 0.1 to 0.5  $\mu$ M to  $P_{hemB}$  was recorded for 300 s; then dissociation was measured for 120 s. Fitting curves ( $k_{on}$ : nonlinear fit using the equation  $Y = Y_0 + Ae^{kt}$ ;  $k_{off}$ : linear fit) are shown in yellow for the fitted data range superimposed on the experimental curves. (C) Determined rate constants  $k_{obs}$  were plotted against the corresponding sCTP-23166 concentrations. The slope equals the negative association constant  $-k_{on}$ . The quotient  $k_{off}/k_{on}$  equals the  $K_D$ . (D) Binding of sCTP-23166 (0.5  $\mu$ M) to three different immobilized DNA probes covering promoter regions of the *hemB*, *hemL*, and *birA* genes immobilized on SAX2 sensors. (E) Binding of the three DB-sCTPs sCTP-16589, sCTP-23166, and sCTP-21477 to the  $P_{birA}$  DNA probe immobilized on SAX2 sensors. (F) GC profiles of DNA probes used for EMSA and BLI over a 22-nt sliding window.

(Fig. 4A). The structure was refined to a final  $R_{\text{work}}$  of 13.6% and  $R_{\text{free}}$  of 17.9%, respectively. All data and refinement statistics are given in *SI Appendix, Table S1*. The observed crystallographic dimer is due to the crystal packing since the interactions between

both monomers are only weak, with a maximum interface area of  $384 \text{ \AA}^2$  and a complexation significance score of 0.000 when calculated via the PDBePISA server ([https://www.ebi.ac.uk/msd-srv/prot\\_int/cgi-bin/piserver](https://www.ebi.ac.uk/msd-srv/prot_int/cgi-bin/piserver)). The two monomers have the



**Fig. 4.** X-ray crystallography-derived 3D structure of sCTP-23166 and predicted DNA interaction. (A) Cartoon representation of the sCTP-23166 structure (pdb-code: 8B6E). (B) 2D topology diagram of sCTP-23166. (C) Hydrophobic core of sCTP-23166. (D) Superposition of the structures of sCTP-23166 (violet) with 1U3E (orange) and 1I3J (gray). (E and F) Predicted protein:DNA interaction. (E) Helix  $\alpha 2$  deeply inserts into the DNA major groove where residues Gln33, Asn34, Lys37, and Lys43 interact with the DNA bases. (F) Residues Ser21, Tyr36, Arg40, and Leu44 contact the DNA backbone. (G) Comparison of protein:DNA interaction of sCTP-23166 (violet), 1U3E (orange), and 1I3J (gray). (H) Amino acid conservation calculated from the alignment of 161 DB-sCTPs (*SI Appendix, Fig. S1B*) mapped on the crystal structure of sCTP-23166. *Top*, worm model in which the thickness corresponds to sequence conservation; *Bottom*, ribbon model.

same overall fold as shown by the rmsd value of 0.41 Å (over 51 Cα atoms) when comparing both protein chains. Therefore, we describe only the monomeric structure in the following.

sCTP-23166 folds into a three-stranded antiparallel β-sheet [β1, residues (res.) 8 to 10; β2, res. 14 to 17; β3, res. 50 to 52] and two α-helices (α1, res. 20 to 27; α2, res. 32 to 41), adopting a β1-β2-α1-α2-β3 topology; an 8-residue loop links helix α2 with strand β3 (Figs. 1A and 4A and B). The two α-helices form a helix–turn–helix (HTH) structure that is packed on one side of the 3-stranded β-sheet forming together a hydrophobic core composed of residues Val8, Val11, Tyr17, Leu20, Ala23, Ala24, Leu29, Ile35, Ala38, Ile39, Ala45, and Trp50 (Fig. 4C). This fold is known for several DNA-binding proteins in terms of the overall topology and secondary structure composition [e.g., the Xis protein from the *Escherichia* virus Lambda (pdb-code: 2IEF; ref. 27) or the DNA-binding N terminus of GP44, a bacteriophage SPO1 protein described as an RNA polymerase inhibitor (pdb-code: 6L6V) ref. 28]. However, the spatial arrangement of the secondary structures (especially the orientation of the helices toward the three-stranded β-sheet) differ in these other proteins, as seen by the high rmsd values of 2.7 Å (over 35 Cα atoms) and 3.6 Å (over 44 Cα atoms), respectively, when overlaid with sCTP-23166 (SI Appendix, Fig. S5A).

By PDBFold (<https://www.ebi.ac.uk/msd-srv/ssm/>) and DALI (<http://ekhidna2.biocenter.helsinki.fi/dali/>) searches against the PDB database (using default settings), we found only two proteins – more precisely protein domains—which adopt an almost identical fold and arrangement of the secondary structure elements to each other despite low aa sequence similarity to sCTP-23166. These were the DNA-binding C-terminal HTH domains of the homing endonucleases I-HmuI from *Bacillus* virus SPO1 (pdb-code: 1U3E; ref. 29) and I-TevI from *Escherichia* virus T4 (pdb-code: 1I3J; ref. 30) with rmsd values of 1.8 Å (over 51 Cα atoms) and 1.4 Å (over 45 Cα atoms), respectively (Fig. 4D and SI Appendix, Fig. S5B).

A sequence identity of at least 35.9 % between sCTP-23166 and the remaining two experimentally studied DB-sCTPs, sCTP-16589 and sCTP-24177, suggests a high structural similarity (31). Hence using the 3D structure of sCTP-23166 to generate homology models of sCTP-16589 and sCTP-24177, resulted in structurally similar models (SI Appendix, Fig. S5C; rmsd values of 0.16 Å and 0.17 Å, respectively). In organisms outside of *Paulinella*, we never found the conserved domain as a standalone protein (using BlastP against the NCBI nr database). Solely a few phages contain the conserved HTH domain fused only to an N-terminal transmembrane domain (SI Appendix, Fig. S5D).

**An In Silico Docking Approach Reveals Putative DNA-Binding Mode of sCTP-23166.** We also aimed for the structure of the DNA-bound sCTP-23166 complex; however, no complex crystal of sufficient quality for X-ray crystallography could be obtained. Therefore, we performed molecular docking to gain additional insights into the DNA-binding mode. The results suggest that helix α2 inserts into the major groove of the DNA (Fig. 4E–G) where the side chains of Gln33, Asn34, and Lys37 (originating from the exposed side of α2) as well as of Lys43 (originating from the loop connecting α2 and β3) contact specific bases of the double helix (Fig. 4E). The polar side chains of Ser21, Tyr36, and particularly the positively charged Arg40 are in close proximity to the negatively charged phosphate backbones of the DNA strands enabling polar and salt–bridge interactions (Fig. 4F). As the optimal DNA target sequence is yet unknown, a random sequence (ATATGCGCGCATAT) was used for the docking. Docking the structure of sCTP-23166 as well as the homology models

of sCTP-16589 and sCTP-24177 to further random sequences (GGGGGGGG, GACTACGT, and AGTCAGTC) suggested that DNA-binding always occurs via helix α2, but not all DB-sCTPs bind to the same DNA sequence (SI Appendix, Table S2).

Mapping protein conservation over the whole alignment of 161 DB-sCTPs onto the sCTP-23166 protein structure reveals high conservation mostly for aa forming the hydrophobic core; in contrast, the predicted DNA proximal residues of helix α2 show particularly low aa conservation (Figs. 1A and 4H).

Although an unbiased docking via HADDOCK2.4 using random patches was performed, the resulting sCTP-23166:DNA complex is surprisingly similar to the DNA-binding modes of the HTH domains of both the homing endonucleases I-HmuI and I-TevI (both structures were determined in the DNA-bound state), where also the more C-terminal helices are deeply buried in the major groove (Fig. 4G). In I-HmuI (pdb-code: 1U3E) the side chains of Arg149, Lys151, Asp154, and Arg160 are forming hydrogen bonds and the side chain of Tyr153, a hydrophobic interaction with the bases of the dsDNA, whereas the side chains of Tyr138, Tyr148, Lys157, His162, His163, Lys164, and Tyr170 are in close proximity to the phosphate backbone of the DNA (29). In the HTH domain of I-TevI (pdb-code: 1I3J), interactions with the DNA bases occur only through one direct hydrophobic interaction between the side chain of Tyr231 and the C5-methyl group of a specific thymine, and one water-bridged hydrogen bond between the main chain nitrogen of Gly227 and another nucleotide in the DNA recognition site. Side chains of Arg220, Ser225, Thr230, Tyr231, Arg232, Lys237, and Tyr242 are in close contact with the DNA backbone (30).

Also in other DNA-binding proteins with similar topology, such as the Xis protein (pdb-code: 2IEF) or the N terminus of GP44 (pdb-code: 6L6V), helix α2 is inserted deeply into the DNA major groove (27, 28). However, for the Xis protein, for which the DNA-binding mode has been explored in detail, the DNA-binding mode differs slightly from the previously discussed proteins. Here the loop connecting two β-strands (called the “wing” motif) is buried in the adjacent minor groove. Furthermore, three monomers of Xis bind to the double-stranded DNA-helix cooperatively with all three monomers showing differences in the number of interacting residues with the DNA (27).

**Photosystem II Subunit CP43 Is a Potential Interaction Partner of sCTP-23166.** To test if sCTP-23166 interacts with chromatophore-localized proteins and to identify potential interaction partners, we performed an in vitro pull-down assay using recombinant sCTP-23166, and as a control, recombinant sCTP-41490 as bait proteins. Proteins pulled down from *P. chromatophora* lysate using both bait proteins were identified by LC-MS/MS and results from quadruplicates compared quantitatively. Besides sCTP-23166 itself, one protein appeared massively enriched in all four samples when sCTP-23166 was used as bait compared to the control. This protein was the chromatophore-encoded chlorophyll-binding photosystem II subunit CP43 (PsbC, PCC0034; Difference 8.4; –logP 6.0; see SI Appendix, Fig. S6 and Dataset S3). Only four other proteins appeared significantly enriched, however, are not considered as likely interaction partners. These proteins were enriched to much lower levels (Difference < 2), and three of these proteins apparently localize to the host cell, whereas for the last protein subcellular localization is unknown.

**DB-sCTP Orthologs in *Paulinella micropora* Show Differential Expression.** Following chromatophore establishment, photosynthetic members of *Paulinella* diverged into at least three species (32). For

the photosynthetic *P. micropora* strain KR01 the nuclear genome has been assembled and differential gene expression analyzed throughout the diurnal cycle and under high-light stress (33). BlastP using the MS-identified DB-sCTPs identified 48 proteins  $\leq 90$  aa among KR01 proteins that share the conserved sequence motif characteristic for DB-sCTPs in *P. chromatophora* (SI Appendix, Fig. S7). Of note, 29% of these proteins are differentially expressed (i.e.  $|\log_2[\text{fold change}]| > 1$ ;  $P = 0.05$ ) under high-light vs. control light conditions, and 20% are classified as “diurnal genes” showing rhythmic differential expression throughout the day; this compares to 9% and 4% of high-light-reactive and diurnal genes, respectively, among all nuclear-encoded proteins (33). Furthermore, a large majority (42/48) of these proteins have been assigned to 20 different (out of 71) distinct coexpression modules. Some of these modules were shown to be enriched in genes involved in specific pathways (such as nitrogen metabolism, arginine or monoterpenoid biosynthesis) or associated with specific GO terms (such as gene expression or RNA processing) (33).

## Discussion

**DB-sCTPs Are DNA-Binding Proteins.** In this work, we describe a class of DNA-binding proteins, the DB-sCTPs, that form an expanded protein family with >200 members in *P. chromatophora*. Only 12 of these proteins have been identified in chromatophores by protein MS so far. But likely more if not all of them localize to the chromatophore as no DB-sCTPs were confidently identified as chromatophore-depleted by MS (6, 21).

X-ray crystallography revealed that DB-sCTPs are almost entirely comprised of a standalone HTH domain that is composed of two  $\alpha$ -helices and three  $\beta$ -strands. The underlying aa sequence is conserved in a variety of bacterial and viral proteins [often annotated as GIY-YIG or HNH type endonucleases or proteins containing domains that are typical for such endonucleases (34)] (SI Appendix, Fig. S5B and ref. 21). However, outside of *Paulinella* this conserved HTH domain is typically combined with other domains in the same protein. Thus, the origin and exact cellular function of these standalone HTH proteins in *Paulinella* cannot be determined with certainty. Their origin might involve HGT from bacterial sources, possibly in relation to the invasion of the genome with a homing endonuclease. However, intact GIY-YIG or HNH type endonucleases were neither found in the transcriptome dataset available for *P. chromatophora* (22) nor the gene models for the nuclear genome of *P. micropora* KR01 (33).

Homing endonucleases that mediate the horizontal transfer of genetic elements to unoccupied integration points use a combination of several DNA-binding domains to recognize their 12- to 40-bp-long DNA target sequences (29, 30, 34). In I-Hmu and I-TevI, the structure and DNA-binding mode of these domains have been characterized in detail (29, 30). In I-HmuI, the HTH domain contacts the terminal eight base pairs of the target site forming seven direct contacts to five different DNA bases as well as several contacts with phosphate groups of the DNA backbone (29). Thus, the HTH domain is a major factor determining the binding specificity of I-HmuI. The HTH domain of I-TevI, directly contacts six nucleotides of the target site. However, here, contacts are made predominantly to the phosphate groups of the DNA backbone, thus, contributing less to the sequence specificity of the binding (30).

In line with its structural similarity to the DNA-binding HTH domains of I-Hmu and I-TevI, EMSA, pull-down from complex *P. chromatophora* lysates, and BLI analyses confirmed that DB-sCTPs generally have the capacity to bind to double-stranded DNA, in

particular fragments amplified from the chromatophore genome comprising putative promoter regions (Dataset S2 and Fig. 3).

**DNA-Binding Properties of DB-sCTPs.** Our unbiased in silico docking approach suggests that similar to the DNA-binding mode of HTH-containing endonucleases, helix  $\alpha 2$  of sCTP-23166 functions as “recognition helix” (Fig. 4G). Whereas some aa residues form unspecific contacts with the phosphate backbone (Fig. 4F), other side chains appear to contact specific DNA bases (Fig. 4E). Although the exact interaction of different DB-sCTPs with the DNA bases, can only be characterized once their cognate DNA target sequences are known, this arrangement suggests that the binding of DB-sCTPs to DNA has a sequence-specific component.

Over the expanded family of DB-sCTPs, the aa residues that are predicted to interact with the DNA show highest variation (Figs. 1A and 4H), which implies that the >200 DB-sCTPs have divergent sequence-dependent DNA-binding preferences. This assumption is supported by the fact that different DB-sCTPs show differential binding to different random DNA sequences in unbiased docking studies (SI Appendix, Table S2). Additionally, our BLI measurements confirmed that sCTP-23166 shows at a given concentration different affinities to different promoter regions (Fig. 3D). Vice versa, the three experimentally studied DB-sCTPs show, at a given concentration, different binding affinities to the same DNA fragment ( $P_{\text{birA}}$ ) (Fig. 3E).

As DNA fragments used in the different in vitro binding assays were chosen arbitrarily and varied in their sequence as well as GC content (Fig. 3F), the DNA-binding properties detected likely reflect mostly the unspecific binding component of DB-sCTPs plus some contribution of sequence elements similar to the optimal target that were present by chance in these DNA fragments. Nevertheless, the binding affinity determined for sCTP-23166 to  $P_{\text{hemB}}$  with a  $K_D$  in the low nM range is close to binding affinities reported for some cyanobacterial transcription factors (TFs) [e.g., a  $K_D$  of 27 nM for the nitrogen-control TF NtcA to its binding site in the *glnA* promoter (35)]. However, experimentally determined binding affinities of different TFs vary greatly depending on experimental conditions (pH, ion composition, presence/absence of ligands) and also  $K_D$  values that are several orders of magnitude lower have been reported [e.g., 1 to 0.1 pM for the TF NF-kB (36)].

We assume that the binding affinity of DB-sCTPs greatly enhances once the optimal sequence is presented and potential interaction partners or ligands are provided. Whether DB-sCTPs function as monomers with low sequence specificity (recognizing ~5 nucleotides) or show more stringent sequence specificity (e.g., by forming complexes with other DNA-binding proteins) is not known yet. The finding of the photosynthetic reaction center protein CP43 (PCC0034) as a potential interaction partner of sCTP-23166 suggests that possibly the latter is true. In chloroplasts, the expression of subunits of the photosynthetic machinery (incl. the *psbC* gene) is tightly controlled by nucleus-encoded factors (mostly at the posttranscriptional level) (37). An important process, termed control by epistasy of synthesis, ensures that efficient synthesis of one chloroplast-encoded subunit of a multiprotein complex only occurs when other subunits are present that are required for assembly of the complex. It is tempting to speculate that in the chromatophore similar control mechanisms evolved. Here, the interaction of chromatophore-encoded proteins with DB-sCTPs (e.g., when their proper assembly into a multiprotein complex is not possible owing to a deficiency of other subunits of the complex) could lead to the formation of a protein complex including two DB-sCTP monomers that can then bind

sequence-specifically to the DNA, blocking further transcription of the protein under control. Future experimentation will be required to test this hypothesis.

**Antimicrobial Activity and Membrane Lipid-Binding Characteristics of sCTPs.** We hypothesize that the observed antimicrobial activity of DB-sCTPs is not primarily related to their DNA-binding capacity, since very similar antimicrobial activities were observed for all sCTPs tested, irrespective of their affiliation to a specific group or predicted cellular function (Fig. 2A and *SI Appendix*, Fig. S3B). Furthermore, these activities resemble the antimicrobial activities that have been observed in similar confrontation assays for NCR proteins from legumes and BCR proteins from aphids where also protein concentrations in the low  $\mu\text{M}$  range resulted in bacterial membrane permeabilization and cell death (14, 15, 18).

We found that for sCTPs in *P. chromatophora*, the observed antimicrobial effect is accompanied by the ability to bind to bacterial (including cyanobacterial) membrane lipids. We assume that this membrane-binding capacity leads to a disturbance of the membrane barrier function at the high-protein concentrations applied in the confrontation assays, whereas at lower concentrations in the cellular context, binding of sCTPs to the chromatophore inner membrane would be nondisruptive. We hypothesize that this characteristic of the sCTPs is related to their import mechanism into the chromatophore. Binding to the chromatophore inner membrane following their release from Golgi-derived vesicles might facilitate translocation of sCTPs across the inner chromatophore membrane by a so far unknown mechanism. However, all sCTPs tested here were cationic and our lipid overlay assays demonstrated preferential binding to lipids with negatively charged head groups. Whether anionic and neutral sCTPs show a similar lipid binding behavior remains to be tested.

**A Possible Function of DB-sCTPs as Nuclear Regulators of Chromatophore Genetic Information Processing.** The coordination of host and chromatophore cell cycles and dependence of chromatophore functions on proteins of dual genetic origin underpins the need of coordinated genetic information processing between the nucleus and chromatophore. Chromatophore-encoded proteins seem to provide only limited capacity for autonomous transcriptional regulation [e.g., only 6 out of 35 genes encoding TFs identified in closely related free-living cyanobacterium *Synechococcus* sp. WH5701 survived reductive genome evolution in the chromatophore (38)]. With their capacity to bind to DNA, the expanded family of DB-sCTPs, thus, represents interesting candidates that could act as nuclear switches controlling aspects of genetic information processing in the chromatophore.

In line with this idea, in *P. micropora* KR01, many DB-sCTPs are differentially expressed under high-light stress or throughout the diurnal cycle (33) and thus, might be involved in translating environmental or diurnal signals into a suitable expressional response in the chromatophore. For other DB-sCTPs, expression or DNA-binding behavior could be modulated by other environmental or cellular triggers.

In sum, here we describe an expanded family of short nucleus-encoded proteins in *P. chromatophora* that are targeted to the chromatophore and have the capacity to bind with high affinity to double-stranded DNA. Exposed amino acid residues of helix  $\alpha 2$  that seems to function as recognition helix, are highly variable between the >200 members of the DB-sCTP family, apparently resulting in a differential, nucleotide sequence-specific binding behavior. We hypothesize that DB-sCTPs represent factors providing *P. chromatophora* with nuclear control over aspects of genetic information processing in the chromatophore and thus, might play

a central role in the integration of host and chromatophore biological networks during the process of organogenesis.

## Material and Methods

**Cultivation of *P. chromatophora* and Synthesis of Complementary DNA.** *P. chromatophora* CCAC0185 (axenic version) was grown and total RNA extracted as described before (22). cDNA was prepared as described in *SI Appendix*.

**Heterologous Expression and Purification of Recombinant Proteins.** His<sub>10</sub>-tagged *P. chromatophora* proteins were overexpressed in *E. coli*, purified on Ni-NTA columns followed by reverse-phase high-performance liquid chromatography, and eluates lyophilized. For all following experiments, the recombinant proteins were dissolved in 10 mM MES/NaOH pH 5.5. For details see *SI Appendix*.

**SDS-PAGE and Western Blot Analysis.** Proteins were analyzed by SDS-PAGE under denaturing conditions on 16% polyacrylamide Tris-Tricine gels (39) and by western blots using anti-His antibodies conjugated to horseradish peroxidase (HRP) in combination with SuperSignal West Pico PLUS HRP chemiluminescent substrate. For details see *SI Appendix*.

***E. coli* Confrontation Assay and PI Staining.** *E. coli* cells were harvested, resuspended in 10 mM MES/NaOH pH 5.5 to an OD<sub>600</sub> of 0.2, and incubated with proteins at the indicated concentrations at 37 °C for 0 to 4 h. Then, 5  $\mu\text{L}$  of serially diluted cells were spotted onto nutrient agar plates, and incubated overnight at 37 °C before documentation.

*E. coli* cells diluted to an OD<sub>600</sub> of 0.1 with 10 mM MES/NaOH pH 5.5 were incubated with proteins at indicated concentrations at 37 °C and 400 rpm for 30 min. Mock treated cells were used as a reference. Then, cells were stained with PI and Hoechst 33342 for further 30 min and analyzed by epifluorescence microscopy and in a plate reader. More details are provided in *SI Appendix*.

**Floatation Assays.** Liposomes either composed of 100% *E. coli* polar lipid extract, or 50% MGDG, 31% DGDG, 11% PG, and 8% SQDG (mimicking cyanobacterial membranes) were incubated with proteins at a molar ratio of 1:1,000 (protein:lipid) for 30 min at RT. The samples were mixed 1:1 with 60% sucrose (final conc. 30%), overlaid stepwise with 20% sucrose and 0% sucrose, and subjected to ultracentrifugation. Then, the gradient was fractionated and protein collected from top, middle and bottom fractions analyzed by SDS-PAGE. For details, see *SI Appendix*.

**DNA Affinity Purification.** MyOne™ Streptavidin C1 Dynabeads™ (ThermoFisher) were coated with a 351-nt-long biotinylated DNA fragment surrounding *P<sub>hemB</sub>* amplified from *P. chromatophora* gDNA. Negative control beads were coated with the biotinylated primer. Beads were used to pull down protein from *P. chromatophora* whole-cell lysate. Protein eluted from the washed beads was analyzed by MS-based protein fingerprinting. For details, see *SI Appendix*.

**EMSA.** The *P<sub>hemB</sub>* fragment was incubated with purified DB-sCTPs at the indicated molar ratios for 30 min at RT. Then, samples were analyzed by gel electrophoresis. For details, see *SI Appendix*.

**BLI Measurements.** Biotinylated DNA fragments amplified from *P. chromatophora* gDNA were loaded onto High Precision streptavidin-coated SAX or SAX2 biosensors and protein association and dissociation studied in 10 mM MES/NaOH pH 5.5 and 100 mM NaCl on a BLITZ platform. For details on the experimentation and evaluation procedures, see *SI Appendix*.

**Protein Crystallization and 3D Structure Determination by X-ray Crystallography.** sCTP-23166 was crystallized by sitting-drop vapor-diffusion. Diffraction data from obtained crystals were collected at beamline P13 (DESY, Hamburg, Germany). Details on crystallization conditions and data processing are provided in *SI Appendix*. The final structure was deposited in the World Wide Protein Data Bank (<https://www.rcsb.org>) under the accession code 8B6E. Figures were generated using PyMOL (Schrodinger LLC; [www.pymol.org](http://www.pymol.org))

**Protein-Nucleotide Docking and Homology Modeling.** Proteins and nucleotides were docked using HADDOCK 2.4 (40). Modeller-10.2 (41) was used for generation of homology models and Clustal  $\Omega$  (42) for sequence alignment. For details, see *SI Appendix*.

**Data, Materials, and Software Availability.** Protein structure data have been deposited in World Wide Protein Data Bank (<https://www.rcsb.org/structure/8B6E>) (43). All study data are included in the article and/or [supporting information](#).

**ACKNOWLEDGMENTS.** This study has been supported by the Deutsche Forschungsgemeinschaft CRC 1208 project B09 (to E.C.M.N.) and Deutsche Forschungsgemeinschaft grant NO 1090/1-1 (to E.C.M.N.). The Center for Structural Studies is funded by the Deutsche Forschungsgemeinschaft (Grant numbers 417919780 and INST 208/740-1 FUGG to S.H.J.S.). We acknowledge the European Synchrotron Radiation Facility (Grenoble, France) for providing assistance in using beamline ID30A-3 during crystal screening. We also thank the European Molecular Biology Laboratory outstation at Deutsches Elektronen-Synchrotron (Hamburg, Germany) for excellent support during data collection at beamline P13.

Author affiliations: <sup>a</sup>Institute of Microbial Cell Biology, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; <sup>b</sup>Center for Structural Studies, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; <sup>c</sup>Molecular Proteomics Laboratory, Medical Faculty and University Hospital, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; <sup>d</sup>Institute of Biological Information Processing (IBI-7 Structural Biochemistry) and JuStruct Jülich Center of Structural Biology, Forschungszentrum Jülich, 52428 Jülich, Germany; <sup>e</sup>Institute of Physical Biology, Department of Biology, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; <sup>f</sup>Institute for Pharmaceutical and Medicinal Chemistry, Department of Pharmacy, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany; and <sup>g</sup>Institute of Biochemistry, Department of Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

Author contributions: L.M. and E.C.M.N. designed research; L.M., T.M.B., T.S., V.K., A.S., L.O., V.A., S.S., A.S., C.G.W.G., A.H., and S.H.J.S. performed research; L.M., L.G., A.H., S.H.J.S., and E.C.M.N. analyzed data; and E.C.M.N. wrote the paper with contributions of all other authors.

1. A. Gabr, A. R. Grossman, D. Bhattacharya, *Paulinella*, a model for understanding plastid primary endosymbiosis. *J. Phycol.* **56**, 837–843 (2020).
2. E. C. M. Nowack, *Paulinella chromatophora*—Rethinking the transition from endosymbiont to organelle. *Acta Soc. Bot. Pol.* **83**, 387–397 (2014).
3. H. R. Hoogenraad, Zur Kenntnis der Fortpflanzung von *Paulinella chromatophora* Lauterb. *Zool. Anz.* **72**, 140–150 (1927).
4. L. Macorano, E. C. M. Nowack, Quick guide: *Paulinella chromatophora*. *Curr. Biol.* **31**, R1024–R1026 (2021).
5. E. C. M. Nowack, A. R. Grossman, Trafficking of protein into the recently established photosynthetic organelles of *Paulinella chromatophora*. *Proc. Natl. Acad. Sci. USA* **109**, 5340–5345 (2012).
6. A. Singer *et al.*, Massive protein import into the early evolutionary stage photosynthetic organelle of the amoeba *Paulinella chromatophora*. *Curr. Biol.* **27**, 2763–2773 (2017).
7. L. Oberleitner, A. Perrar, L. Macorano, P. F. Huesgen, E. C. M. Nowack, A bipartite chromatophore transit protein and N-terminal protein processing in the *Paulinella* chromatophore. *Plant Physiol.* **189**, 152–164 (2022).
8. B. Alunni *et al.*, Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. *Mol. Plant Microbe Interact.* **20**, 1138–1148 (2007).
9. G. Maróti, É. Kondorosi, Nitrogen-fixing Rhizobium-legume symbiosis: Are polyploidy and host peptide-governed symbiont differentiation general principles of endosymbiosis? *Front. Microbiol.* **5**, 326 (2014).
10. P. Mergaert *et al.*, A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiol.* **132**, 161–173 (2003).
11. S. Shigenobu, D. L. Stern, Aphids evolved novel secreted proteins for symbiosis with bacterial endosymbiont. *Proc. Biol. Sci.* **280**, 20121952 (2012).
12. F. H. Login *et al.*, Antimicrobial peptides keep insect endosymbionts under control. *Science* **334**, 362–365 (2011).
13. L. Carro *et al.*, *Alnus* peptides modify membrane porosity and induce the release of nitrogen-rich metabolites from nitrogen-fixing *Frankia*. *ISME J.* **9**, 1723–1733 (2015).
14. A. Farkas, G. Maróti, A. Kereszt, E. Kondorosi, Comparative analysis of the bacterial membrane disruption effect of two natural plant antimicrobial peptides. *Front. Microbiol.* **8**, 51 (2017).
15. N. Uchi *et al.*, Antimicrobial activities of cysteine-rich peptides specific to bacteriocytes of the pea aphid *Acyrtosiphon pisum*. *Microbes Environ.* **34**, 155–160 (2019).
16. P. Mergaert, Role of antimicrobial peptides in controlling symbiotic bacterial populations. *Nat. Prod. Rep.* **35**, 336–356 (2018).
17. P. Mergaert *et al.*, Eukaryotic control on bacterial cell cycle and differentiation in the Rhizobium-legume symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5230–5235 (2006).
18. W. van de Velde *et al.*, Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science* **327**, 1122–1126 (2010).
19. S. Sankari *et al.*, A haem-sequestering plant peptide promotes iron uptake in symbiotic bacteria. *Nature Microbiol.* **7**, 1453–1465 (2022).
20. J. Penterman *et al.*, Host plant peptides elicit a transcriptional response to control the *Sinorhizobium meliloti* cell cycle during symbiosis. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 3561–3566 (2014).
21. L. Oberleitner *et al.*, The puzzle of metabolite exchange and identification of putative octotricopeptide repeat expression regulators in the nascent photosynthetic organelles of *Paulinella chromatophora*. *Front. Microbiol.* **11**, 607182 (2020).
22. E. C. M. Nowack *et al.*, Gene transfers from diverse bacteria compensate for reductive genome evolution in the chromatophore of *Paulinella chromatophora*. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 12214–12219 (2016).
23. O. Emanuelsson, S. Brunak, G. von Heijne, H. Nielsen, Locating proteins in the cell using TargetP, SignalP and related tools. *Nat. Protoc.* **2**, 953–971 (2007).
24. A. F. Haag *et al.*, Protection of *Sinorhizobium* against host cysteine-rich antimicrobial peptides is critical for symbiosis. *PLoS Biol.* **9**, e1001169 (2011).
25. A. F. Haag *et al.*, Role of cysteine residues and disulfide bonds in the activity of a Legume root nodule-specific, cysteine-rich peptide. *J. Biol. Chem.* **287**, 10791–10798 (2012).
26. N. Sato, T. Yoshitomi, N. Mori-Moriyama, Characterization and biosynthesis of lipids in *Paulinella microphora* MYN1: Evidence for efficient integration of chromatophores into cellular lipid metabolism. *Plant Cell Physiol.* **61**, 869–881 (2020).
27. M. A. Abbani *et al.*, Structure of the cooperative Xis-DNA complex reveals a micronucleoprotein filament that regulates phage lambda intasome assembly. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2109–2114 (2007).
28. Z. Wang *et al.*, A bacteriophage DNA mimic protein employs a non-specific strategy to inhibit the Bacterial RNA Polymerase. *Front. Microbiol.* **12**, 692512 (2021).
29. B. W. Shen, M. Landthaler, D. A. Shub, B. L. Stoddard, DNA binding and cleavage by the HNH homing endonuclease I-Hmul. *J. Mol. Biol.* **342**, 43–56 (2004).
30. P. Van Roey, C. A. Waddling, K. M. Fox, M. Belfort, V. Derbyshire, Intertwined structure of the DNA-binding domain of intron endonuclease I-Tev with its substrate. *EMBO J.* **20**, 3631–3637 (2001).
31. L. R. Forrest, C. L. Tang, B. Honig, On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J.* **91**, 508–517 (2006).
32. D. Lee *et al.*, Diversity of the photosynthetic *Paulinella* species, with the description of *Paulinella microphora* sp. nov. and the chromatophore genome sequence for strain KR01. *Protist* **168**, 155–170 (2017).
33. D. Lee *et al.*, Amoeba genome reveals dominant host contribution to plastid endosymbiosis. *Mol. Biol. Evol.* **38**, 344–357 (2021).
34. E. Sitbon, S. Pietrokovski, New types of conserved sequence domains in DNA-binding regions of homing endonucleases. *Trends Biochem. Sci.* **28**, 473–477 (2003).
35. M. F. Vázquez-Bermúdez, E. Flores, A. Herrero, Analysis of binding sites for the nitrogen-control transcription factor NtcA in the promoters of *Synechococcus* nitrogen-regulated genes. *Biochim. Biophys. Acta* **1578**, 95–98 (2002).
36. U. Zabel, R. Schreck, P. A. Baeuerle, DNA binding of purified transcription factor NF- $\kappa$ B. Affinity, specificity, Zn<sup>2+</sup> dependence, and differential half-site recognition. *J. Biol. Chem.* **266**, 252–260 (1991).
37. F. A. Wollman, L. Minai, R. Nechushtai, The biogenesis and assembly of photosynthetic proteins in thylakoid membranes. *Biochim. Biophys. Acta* **1411**, 21–85 (1999).
38. E. C. M. Nowack, M. Melkonian, G. Glöckner, Chromatophore genome sequence of *Paulinella* sheds light on acquisition of photosynthesis by eukaryotes. *Curr. Biol.* **18**, 410–418 (2008).
39. H. Schägger, Tricine-SDS-PAGE. *Nat. Protoc.* **1**, 16–22 (2006).
40. R. V. Honorato *et al.*, Structural biology in the clouds: The WeNMR-EOSC Ecosystem. *Front. Mol. Biosci.* **8**, 729513 (2021).
41. N. Eswar *et al.*, Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinformatics* **Chapter 5**, Unit 5.6 (2006).
42. F. Madeira *et al.*, Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **50**, W276–W279 (2022).
43. L. Macorano *et al.*, Crystal structure of the DNA-binding short chromatophore-targeted protein sCTP-23166 from *Paulinella chromatophora*. RCSB Protein Data Bank. <https://www.rcsb.org/structure/8B6E>. Deposited 27 September 2022.