



**HAL**  
open science

## **Crystal structure of the transcriptional repressor DdrO: insight into the metalloprotease/repressor-controlled radiation response in Deinococcus**

Arjan de Groot, Marina I. Siponen, Romaric Magerand, Nicolas Eugenie,  
Raquel Martin-Arevalillo, Jade Doloy, David Lemaire, Géraldine Brandelet,  
François Parcy, Renaud Dumas, et al.

### ► To cite this version:

Arjan de Groot, Marina I. Siponen, Romaric Magerand, Nicolas Eugenie, Raquel Martin-Arevalillo, et al.. Crystal structure of the transcriptional repressor DdrO: insight into the metalloprotease/repressor-controlled radiation response in Deinococcus. *Nucleic Acids Research*, 2019, 47 (21), pp.11403-11417. 10.1093/nar/gkz883 . cea-02315882

**HAL Id: cea-02315882**

**<https://cea.hal.science/cea-02315882>**

Submitted on 5 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Crystal structure of the transcriptional repressor DdrO: insight into the metalloprotease/repressor-controlled radiation response in *Deinococcus*

Arjan de Groot<sup>1</sup>, Marina I. Siponen<sup>1</sup>, Romaric Magerand<sup>1</sup>, Nicolas Eugénie<sup>2</sup>, Raquel Martin-Arevalillo<sup>3</sup>, Jade Doloy<sup>1</sup>, David Lemaire<sup>4</sup>, Géraldine Brandelet<sup>1</sup>, François Parcy<sup>3</sup>, Renaud Dumas<sup>3</sup>, Philippe Roche<sup>5</sup>, Pascale Servant<sup>2</sup>, Fabrice Confalonieri<sup>2</sup>, Pascal Arnoux<sup>1</sup>, David Pignol<sup>1</sup> and Laurence Blanchard<sup>1,\*</sup>

<sup>1</sup>Aix Marseille Univ, CEA, CNRS, BIAM, Molecular and Environmental Microbiology Team, Saint Paul-Lez-Durance, F-13108, France, <sup>2</sup>Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Univ. Paris-Sud, Univ. Paris-Saclay, Gif-sur-Yvette cedex, F-91198, France, <sup>3</sup>Univ. Grenoble Alpes, CNRS, CEA, INRA, IRIG-DBSCI-LPCV, Grenoble, F-38000, France, <sup>4</sup>Aix Marseille Univ, CEA, CNRS, BIAM, Interaction Protein Metal Team, Saint Paul-Lez-Durance, F-13108, France and <sup>5</sup>Aix Marseille Univ, CNRS, Inserm, Institut Paoli Calmettes, CRCM, Marseille CEDEX 09, F-13273, France

Received June 08, 2019; Revised September 25, 2019; Editorial Decision September 26, 2019; Accepted October 01, 2019

## ABSTRACT

Exposure to harmful conditions such as radiation and desiccation induce oxidative stress and DNA damage. In radiation-resistant *Deinococcus* bacteria, the radiation/desiccation response is controlled by two proteins: the XRE family transcriptional repressor DdrO and the COG2856 metalloprotease IrrE. The latter cleaves and inactivates DdrO. Here, we report the biochemical characterization and crystal structure of DdrO, which is the first structure of a XRE protein targeted by a COG2856 protein. DdrO is composed of two domains that fold independently and are separated by a flexible linker. The N-terminal domain corresponds to the DNA-binding domain. The C-terminal domain, containing three alpha helices arranged in a novel fold, is required for DdrO dimerization. Cleavage by IrrE occurs in the loop between the last two helices of DdrO and abolishes dimerization and DNA binding. The cleavage site is hidden in the DdrO dimer structure, indicating that IrrE cleaves DdrO monomers or that the interaction with IrrE induces a structural change rendering accessible the cleavage site. Predicted COG2856/XRE regulatory protein pairs are found in many bacteria, and available data suggest two different molecular mechanisms for stress-induced gene expression: COG2856

protein-mediated cleavage or inhibition of oligomerization without cleavage of the XRE repressor.

## INTRODUCTION

Characterization of bacterial stress response mechanisms at a molecular level is crucial to increase our knowledge of the diverse bacterial defence strategies against different stresses and damaging agents, and may lead to new therapeutic and industrial applications. Exposure to harmful conditions like radiation and oxidative stress provoke damage to DNA and other molecules. DNA damage and oxidative stress responses have been studied most extensively in *Escherichia coli*, but there is considerable variation in regulation and molecular mechanisms of these responses across bacterial species (1,2). Regarding regulation of DNA repair genes, many characterized bacterial species including *E. coli* use the RecA/LexA-regulated SOS response, where DNA damage stimulates the RecA-dependent autocleavage of repressor LexA, which can then no longer form a stable DNA-binding dimer (3). The SOS response is known in particular for induction of DNA repair genes, but SOS-induced generation of antibiotic-resistant persister cells (mediated by the toxin of one or more toxin-antitoxin systems) as well as LexA-regulated apoptosis-like death (ALD) following irreversible DNA damage has also been demonstrated in some bacteria (1,4).

*Deinococcus* bacteria are extremely resistant to radiation, desiccation, and other DNA-damage- and oxidative stress-

\*To whom correspondence should be addressed. Tel: +33 442 252 431; Fax: +33 442 254 701; Email: laurence.blanchard@cea.fr

generating conditions. The genomic DNA of *Deinococcus* is not protected since exposure to high doses of ionizing radiation causes massive DNA damage, including hundreds of DNA double-strand breaks, but these bacteria have the capacity to repair such DNA damage within hours after irradiation or after prolonged desiccation. Many studies have indicated that this extreme resistance results from a combination of multiple factors and mechanisms, including limitation of oxidative protein damage (5–7). An efficient radiation/desiccation response (RDR) mechanism for induced expression of various DNA repair and other genes is also crucial for radiation resistance (8,9). This SOS-independent RDR mechanism controls expression of genes that contain, in their promoter regions around the transcription start site, a 17-base-pair palindromic motif designated radiation/desiccation response motif (RDRM) (9,10). Depending on the *Deinococcus* species, the predicted RDR regulon consists of at least 14–24 genes, including DNA repair genes such as *recA*, *ssb*, *gyrA*, *gyrB*, *uvrA* and *uvrB* and *Deinococcus*-specific genes such as *ddrB* and *ddrC* (11).

We previously showed that expression of the RDR regulon genes is controlled by two proteins, DdrO and IrrE (12). DdrO is a XRE family transcriptional regulator that binds the RDRM and functions as the repressor of RDR genes (including *ddrO* itself) (11–14). IrrE, also called PprI, is a COG2856 domain-containing metalloprotease that cleaves DdrO when the cells are exposed to radiation, resulting in gene induction through an original mechanism that differs from the RecA/LexA-controlled SOS response (12). Mutants lacking *irrE* grow normally under standard conditions, but the cells are very sensitive to radiation and RDR gene induction is abolished (11,15–17). The *ddrO* gene is essential for cell viability and, interestingly, its prolonged depletion induces ALD (membrane blebbing, DNA fragmentation), suggesting that DNA repair and ALD might be controlled by the same regulators, IrrE and DdrO (13). The genes/proteins that provoke this ALD are currently unknown.

COG2856/XRE protein pairs have been predicted in *Meiothermus* spp. and some other species belonging to the phylum *Deinococcus-Thermus* (12) and also in bacteria unrelated to *Deinococcus* including human pathogens (18), indicating that a regulatory mechanism involving such protein pair is more widespread than currently recognized. Genes encoding COG2856/XRE pairs are often located in mobile genetic elements (bacteriophages, transposons) and in predicted operons with genes encoding toxins that may be involved in formation of antibiotic-resistant persister cells (18). Although present in many bacteria, the majority of these IrrE/DdrO-like putative metalloprotease/repressor proteins have not been studied. For one such pair, ImmA/ImmR of *Bacillus subtilis*, cleavage of the repressor (ImmR) by the metalloprotease (ImmA) has been demonstrated, and shown to control horizontal transfer of the mobile genetic element on which the *immA* and *immR* genes are located (19).

Concerning the COG2856/XRE pairs, the 3D structure of only one protein has been solved previously, the metallopeptidase IrrE of *Deinococcus deserti* (17). Here, to move a step forward in the structural characterization of the ge-

netic switch involving COG2856/XRE pairs, we have characterized IrrE's substrate, the repressor protein DdrO. The N-terminal half of DdrO contains the predicted helix-turn-helix (HTH) motif-containing DNA-binding domain of the XRE family, whereas the C-terminal half does not show similarity to any conserved structural or functional domain. DdrO was studied using a 'domain dissection approach' associated to a combination of bacterial two-hybrid experiments, size exclusion chromatography (SEC), SEC-MALLS (multiple angle laser light scattering), native mass spectrometry, electrophoretic mobility shift assay (EMSA), X-ray crystallography, analysis of point mutants and *in vivo* complementation assays. We show that DdrO is a modular protein composed of two domains: the N-terminal HTH\_XRE DNA-binding domain associated to a novel structural domain at the C terminus. The N- and C-terminal domains can fold independently and are separated by a flexible linker. The C-terminal domain of DdrO is required for protein dimerization and for DNA binding of entire DdrO, and is constituted of three helices arranged in a new fold. Cleavage of the scissile peptide bond, located within a positively charged loop between the two last helices of DdrO, triggers a quaternary structural rearrangement. Our DdrO 3D structures also give insight into how DdrO interacts with IrrE.

## MATERIALS AND METHODS

### Bacterial strains and growth conditions

The bacterial strains used in the study are listed in Supplementary Table S1. *Escherichia coli* strains TOP10 or DH5 $\alpha$  were used for cloning, strain SCS110 to propagate plasmids prior to introduction into *Deinococcus radiodurans* via transformation (20), strain BL21 (AI) for overexpression of recombinant proteins, and strain BTH101 for bacterial two-hybrid experiments. *D. radiodurans* GY14125 was used for transformation with DdrO-expressing plasmids, and strains R1 and GY14127 for controls in experiments testing mitomycin C (MMC) resistance. *E. coli* strains were grown in LB medium at 37°C unless indicated otherwise. *D. radiodurans* was grown at 30°C in TGY2X (1% tryptone, 0.2% dextrose, 0.6% yeast extract) or on TGY1X plates. When required, antibiotics were added to the growth media at the following concentrations for *E. coli*: ampicillin, 100  $\mu$ g/ml; kanamycin, 50  $\mu$ g/ml; spectinomycin, 40  $\mu$ g/ml. For *D. radiodurans*, 75  $\mu$ g/ml spectinomycin and 3.5  $\mu$ g/ml chloramphenicol was used.

### Plasmid construction and point mutagenesis

Plasmids and primers used are listed in Supplementary Tables S2 and S3, respectively. To obtain the plasmids for bacterial two-hybrid assays, DNA fragments were amplified by PCR using primers that included restriction sites, and cloned in pCR4Blunt-TOPO prior to recloning in the EcoRI and XhoI sites of pEB354 and pEB355. To obtain plasmids for expressing proteins containing an N-terminal polyhistidine-SUMO tag, DNA fragments were amplified and cloned in pET SUMO as described by the manufacturer. Point mutations were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene).

For complementation experiments, *ddrO* genes were cloned in p13841 or p17235 that replicate in *D. radiodurans*. DNA sequencing was performed to verify absence of potential cloning or PCR errors in the obtained plasmids.

### Two-hybrid in *E. coli*

Interaction between DdrO proteins (full-length or domains) was tested using the Bacterial Adenylate Cyclase Two Hybrid (BACTH) system in *E. coli* as described (21). Briefly, DdrO proteins or domains were fused to the catalytic domains T18 and T25 using the appropriate plasmids. Production of the different fusion proteins was checked by Western blot using polyclonal antibodies raised against DdrO<sub>C</sub> from *D. deserti*. Reporter strain BTH101 was co-transformed with two plasmids, one encoding the T18 fusion and the other the T25 fusion (or plasmids not encoding fusion proteins for controls). The transformation plates were incubated 24 or 48 h at 30°C. Several clones from the transformation plates were inoculated in 3 ml LB containing ampicillin, kanamycin and 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG), and incubated overnight at 30°C at 150 rpm. The next day 5 μl of each culture were spotted on MacConkey plates (Difco™ MacConkey Agar Base, ref.: 281810) supplemented with ampicillin, kanamycin and 1% lactose. The MacConkey plates were then incubated at 30°C for 16 or 24 h.

### Complementation of the *ddrO* deletion in *D. radiodurans*

Following transformation of *D. radiodurans* strain GY14125 with plasmids expressing *D. deserti ddrO* genes (or *D. radiodurans ddrO* as control), the transformants were re-streaked several times on plates supplemented with chloramphenicol and spectinomycin. Then the complete deletion from all chromosome copies of native *D. radiodurans ddrO* was analysed by diagnostic PCR as described previously (13).

### Protein expression and purification and *in vitro* cleavage assay

IrrE was produced in *E. coli* from the pET-TEV plasmid as an N-terminal His-tagged fusion protein (12), and DdrO from pET SUMO as an N-terminal polyhistidine-SUMO fusion protein (11). For purification, 20 ml cell suspensions (in buffer 1: 500 mM NaCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4) containing the recombinant fusion protein were thawed and 100 μl of 10 mg/ml DNase I [DNase from bovine pancreas in 50 mM MgCl<sub>2</sub> (Sigma)] and 25 μl of anti-protease cocktail (Sigma P8849) were added to each tube. All purification steps were done at 4°C. Cells were broken by a cell disruptor (One Shot Model, Constant Cell Disruption Systems) at 2 kbar and the soluble extracts were then recovered after centrifugation at 10 000 g for 10 min at 4°C and ultracentrifugation at 150 000 g for 45 min at 4°C. The supernatant was injected at a speed of 1 ml/min onto HisTrap™ HP columns (1 ml) (GE Healthcare), previously equilibrated in Buffer 1 supplemented with 20 mM imidazole. A step gradient of imidazole (20, 40, 100 and 500 mM in Buffer 1) was used for elution (20 ml was used for

each fraction except for the 500 mM fraction where 4 ml was used). Eluted fractions were analysed by SDS-PAGE to verify presence and purity of the protein. IrrE proteins were recovered in the 40 and 100 mM fractions and DdrO in the 500 mM imidazole fraction. Removal of imidazole from the fractions containing IrrE or DdrO was done on a PD-10 desalting column (GE Healthcare) equilibrated with Buffer 1. To remove the polyhistidine tag of IrrE, purified protein at a concentration of 1 mg/ml was incubated with polyhistidine-tagged TEV protease overnight at room temperature (TEV protease:IrrE mass ratio of 1:80). The polyHis-SUMO tag of DdrO was removed from the fusion protein by incubation with SUMO protease. The pET28b derivative encoding His-tagged SUMO protease was kindly provided by Mossessova and Lima (22). The incubation mixture contained one unit of SUMO protease for 20 μg protein, and proteolysis proceeded overnight at room temperature. The cleavage reactions were then loaded on a nickel affinity column to separate untagged IrrE and untagged DdrO (in the flow-through fractions) from uncleaved fusion proteins and TEV protease or SUMO protease (both bound to the column). The fraction containing purified IrrE or DdrO was concentrated using an Amicon Ultra-15 Centrifugal filter unit (MWCO 10 000, Millipore) to a 5 ml volume which was then loaded on a gel filtration column [HiLoad 26/200 Superdex 200 prep grade (GE Healthcare)] equilibrated and eluted with 50 mM Tris-HCl pH 7.4 supplemented with either 0.15 M NaCl for IrrE or 0.4 M NaCl for DdrO (load at 1 ml/min, run at 2 ml/min). DdrO domains and DdrO point mutants were expressed and purified (including tag removal) as performed for full-length wild-type DdrO. SDS-PAGE was used to follow each purification step and to verify the purity of the proteins and mutants. The mass of the purified proteins was verified by electrospray ionization mass spectrometry as described previously (12). All protein samples were checked for the absence of undesired nucleic acid contaminants by UV-visible spectroscopy showing a 260/280 nm absorbance ratio close to 0.57 (0.57–0.61 for our purified proteins) that corresponds to an uncontaminated protein sample (23), and by native polyacrylamide gels stained with ethidium bromide showing no DNA. The *in vitro* cleavage assays were performed as previously described (11,12).

### Production of selenomethionine-labeled DdrO

Selenomethionine (SeMet)-labelled DdrO A70M I84M was produced in *E. coli* BL21 (AI) following standard procedures inhibiting methionine biosynthesis (24,25) and the protein was purified using the same protocol as used for the native protein.

### SEC-MALLS

SEC-MALLS experiments for molecular weight determination of DdrO and its domains was carried out with an analytical Superdex-200 increase column (GE Healthcare) equilibrated with 50 mM Tris-HCl pH 7.4, 150 mM NaCl. Runs were performed at 25°C with a flow rate of 0.5 ml/min. Elutions were monitored by using a DAWN-EOS detector with a laser emitting at 690 nm for online MALLS measurement (Wyatt Technology Corp., Santa Barbara, CA,

USA) and with a RI2000 detector for online refractive index measurements (Schambeck SFD). Molecular weight calculations were performed using the ASTRA software (version 5.3.4.20) using a refractive index increment ( $dn/dc$ ) of 0.185 ml/g.

## SEC

Additional SEC experiments were performed to investigate oligomerization of DdrO and its domains using a Superdex 75 10/300 GL column (GE Healthcare) at 4°C equilibrated with 50 mM Tris-HCl pH 7.4, 150 mM NaCl, and using different protein and NaCl concentrations. SEC analyses of DdrO point mutants were performed at 4°C on a Superdex 200 Increase 10/300 GL equilibrated with 50 mM Tris-HCl pH 7.4, 150 mM NaCl.

## Electrophoretic mobility shift assay (EMSA)

Gel shift experiments with DNA fragments containing the promoters of the RDR regulon genes *ddrA* and *ddrO<sub>C</sub>-ddrQ* were performed as described previously (11). The 270-base pair promoter region of *ddrA* (Deide\_09150) or the 270-base pair intergenic region of the divergent *ddrO<sub>C</sub>* (Deide\_20570) and *ddrQ* (Deide\_20580) genes was amplified by PCR (for primers, see Supplementary Table S3). DNA binding reactions were performed for 30 min at room temperature in a volume of 20  $\mu$ l containing 200 ng of DNA fragments and 1.7  $\mu$ M (final concentration) of either purified DdrO (full-length), NTD, CTD or DdrO<sub>(1-106)</sub> in 10 mM Tris-HCl pH 7.4, 60 mM NaCl, 1 mM DTT, 5% glycerol. EMSA was also done with 500 ng of a 23-base pair ds-DNA fragment called ddrB23, containing the RDRM of *ddrB* (Deide\_02990). This fragment was obtained by annealing primers ddrB23for and ddrB23rev (Supplementary Table S3). After adding 2  $\mu$ l of 10X Orange Loading Dye, the samples were loaded onto a prerun 5% polyacrylamide gel and run for 2 h at 70 V in TAE buffer at 4°C. The gel was then incubated for 30 min in ethidium bromide (0.5  $\mu$ g/ml) and DNA was visualized using UV.

## Electrospray ionization mass spectrometry of proteins and protein-DNA complexes

Mass spectrometry (MS) analyses were performed on a MicroTOF-Q (Bruker, Wissembourg, France) with an electrospray ionization source.

All purified protein samples (full-length DdrO, domains, mutants) used in SEC, SEC-MALLS, EMSA and for X-ray structure determination were analysed by mass spectrometry. In denaturing conditions, protein concentrations were 1  $\mu$ M in CH<sub>3</sub>CN/H<sub>2</sub>O (1/1, v/v), 0.2% formic acid (Sigma). Samples were continuously infused at a flow rate of 3  $\mu$ l/min. Mass spectra were recorded in the 50–7000 mass-to-charge ( $m/z$ ) range. MS experiments were carried out with a capillary voltage set at 4.5 kV and an end plate off set voltage at 500 V. The gas nebulizer (N<sub>2</sub>) pressure was set at 0.4 bar and the dry gas flow (N<sub>2</sub>) at 4 l/min at a temperature of 190°C.

For analysis of protein-DNA complexes, 2.1 ml of DdrO at 3.5 mg/ml (500 nmol in 50 mM Tris-HCl pH 7.4, 120

mM NaCl) and 250  $\mu$ l of ddrB23 at 1 mM (250 nmol in 10 mM Tris-HCl pH 7.4, 50 mM NaCl) were mixed to have a 2:1 stoichiometry. The mix was left 30 min at room temperature then loaded onto a HiLoad 26/600 Superdex 75 pg column equilibrated with 50 mM Tris-HCl pH 7.4, 120 mM NaCl. The run was done at 4°C at 2 ml/min and fractions of 4 ml were collected. Native mass spectrometry measurements were performed to determine the stoichiometry of the non-covalent protein/DNA complexes. The samples at a concentration of 10  $\mu$ M in 50 mM ammonium acetate were continuously infused at a flow rate of 7  $\mu$ l/min. The mass spectra were recorded in the 1000–8500 mass-to-charge ( $m/z$ ) range. The gas nebulizer (N<sub>2</sub>) pressure was set at 3 bar and the dry gas flow (N<sub>2</sub>) at 4 l/min at a temperature of 200°C. Data were acquired in the positive mode and calibration was performed using a calibrating solution of ESI Tune Mix (Agilent) in CH<sub>3</sub>CN/H<sub>2</sub>O (95/5, v/v). The system was controlled with the software package MicroTOF Control 2.2 and data were processed with DataAnalysis 3.4. Identification of different species, for example different protein oligomerization states or DNA-binding species, are based on the fact that a species can be detected, if and only if several consecutive charge states forming the charge state envelope of the species in question are present on the mass spectrum.

## Protein crystallization, data collection and structure determination

For all proteins, crystallization screenings were performed using 96-well plates in a sitting drop setup using a Freedom EVO 100 pipetting robot (Tecan).

For the SeMet-substituted A70M I84M mutant of DdrO, screening was done by mixing 0.5  $\mu$ l of protein at 10 mg/ml in 20 mM Tris-HCl pH 7.4, 280 mM NaCl with 0.5  $\mu$ l of reservoir solution from various commercial screens. Hits (spherulites) were initially obtained in position G2 of the JCSG-Plus screen (0.1 M Bis-Tris pH 5.5, 2 M NaCl) and were used as a seed in the subsequent optimisation conditions. Diffracting crystals were obtained by mixing 0.5  $\mu$ l of 10 mg/ml of protein with an equal volume of reservoir solution containing 0.1 M Bis-Tris propane pH 7, 3.5 M ammonium chloride (condition 7 (A7) SaltRx HT (Hampton Research)) and 0.1  $\mu$ l of seed crystal. For native full-length DdrO, crystals were obtained by mixing 0.5  $\mu$ l of 10 mg/ml of protein in 20 mM Tris-HCl pH 7.4, 280 mM NaCl with an equal volume of reservoir solution containing 0.09 M Bis-Tris pH 5.5, 0.2 M Na-thiocyanate, 2.3 M NaCl and 0.1  $\mu$ l of seed crystal. Diffracting crystals of NTD were obtained in a sitting drop setup by mixing 0.5  $\mu$ l of 6.5 mg/ml of protein (20 mM Tris-HCl pH 7.4, 200 mM NaCl) with an equal volume of reservoir solution (0.1 M Bis-Tris propane pH 7, 2.5 M ammonium sulfate (condition 65 (F5), SaltRx HT (Hampton Research))). Diffracting crystals of CTD were obtained by mixing 0.5  $\mu$ l of 7 mg/ml of protein (20 mM Tris-HCl pH 7.4, 200 mM NaCl) with an equal volume of reservoir solution (0.1 M sodium citrate tribasic dihydrate pH 4.5, 0.2 M lithium sulfate, 46% w/v PEG 400, (optimization of condition A1 of JCSG-Plus (Molecular dimensions))). All diffracting crystals appeared

within one or two days and reached their final size within a few days.

For cryoprotection, all crystals were transferred in a drop made of mother liquor supplemented with 20–30% (w/v) glycerol before flash freezing directly in liquid nitrogen. X-ray data were collected at different beam-lines at the ESRF (Grenoble) and data integration was performed with Mosflm from the CCP4 program suite (26). The structure of the full-length protein was solved using the SeMet-substituted mutant of the protein (A70SeM I84SeM) and the SAD phasing method (using Crank2 from the CCP4 suite). The structures of native DdrO, NTD and CTD were subsequently solved by molecular replacement (using Molrep from the CCP4 suite). All structures were refined with Refmac5 and phenix.refine with structure validations performed using Procheck from the CCP4 suite or MolProbity within the Phenix suite (27,28). In the case of TLS refinement, TLS groups were defined as per the TLS Motion Determination homepage (<http://skuld.bmsc.washington.edu/~tlsmd/>). CCP4mg software was also used to generate figures (26).

### Molecular docking calculations

Two docking webservers, HawkDock (<http://cadd.zju.edu.cn/hawkdock/>) (29) and HADDOCK (<https://haddock.science.uu.nl/>) (30), were used to generate structural models of IrrE in complex with the C-terminal domain of DdrO (either monomeric or dimeric). The PDB files uploaded for the calculations were 3DTE for the 3D structure of IrrE and 6RO6 for the CTD of DdrO (chain A or chains AC used for the CTD monomer or dimer respectively). For HawkDock calculations, no restraints were used to drive the docking. The Top 100 models generated between IrrE and the CTD monomer or the CTD dimer, without re-ranking using MM/GBSA, were analysed after repositioning Zn in the active site of IrrE and renumbering the residues. Distances between Zn in the IrrE 3D structure and peptide bond atoms C and N of respectively L106 and R107 of DdrO were calculated for each model. For HADDOCK calculations, active site residues 82, 83, 86 and 113 of IrrE and L106 and R107 of DdrO were used to drive the docking.

## RESULTS

### Dimerization of repressor DdrO

Deinococcal DdrO proteins are highly conserved over their entire length (Supplementary Figure S1). Most *Deinococcus* species possess a single *ddrO* gene. *D. deserti* produces two different DdrO proteins sharing 84% identity, DdrO<sub>C</sub> (Deide\_20570) and DdrO<sub>P3</sub> (Deide\_3p02170), encoded by the chromosome and plasmid P3, respectively. DdrO<sub>C</sub> is most similar to DdrO proteins from other species (e.g. 96% identity with *D. radiodurans* DdrO), and is characterized by several complementary methods in this work. From here on DdrO is used as a synonym for DdrO<sub>C</sub>.

DdrO from *D. deserti* is composed of 129 amino acid residues and has a theoretical molecular weight (MW) of 14.737 kDa. The N-terminal region of DdrO shows the presence of an HTH domain of the XRE family (Figure 1A). HTH\_XRE-type DNA-binding domains are present

in numerous transcriptional regulators, for example in *B. subtilis* prophage PBSX repressor protein Xre, in regulatory proteins Cro and C1 of several bacteriophages, and in the biofilm formation regulator protein SinR of *B. subtilis*. No similarity to a domain in the conserved domain database was found for the C-terminal region of DdrO, but this region does contain the site where DdrO is cleaved by metalloprotease IrrE.

Using a bacterial two-hybrid system, we previously showed (12) and confirmed here (Figure 1B) that DdrO interacts with itself, indicating that DdrO forms at least dimers. To determine the oligomeric state of DdrO in solution, SEC-MALLS experiments were performed. At low DdrO concentration, monomeric DdrO and a complex with a MW of 27.5 kDa corresponding to a dimer were observed (Figure 1C and Table 1). At higher concentration, SEC-MALLS indicates a complex of 38.4 kDa that is close to the theoretical MW of a trimer (Figure 1D and Table 1), but which likely corresponds to a fast dimer-tetramer equilibrium that prevents the separation of the two species in the SEC column (see native mass spectrometry and crystal structures below). Stable tetramers or higher order of DdrO multimers were not observed when increasing the DdrO concentration (Supplementary Figure S2).

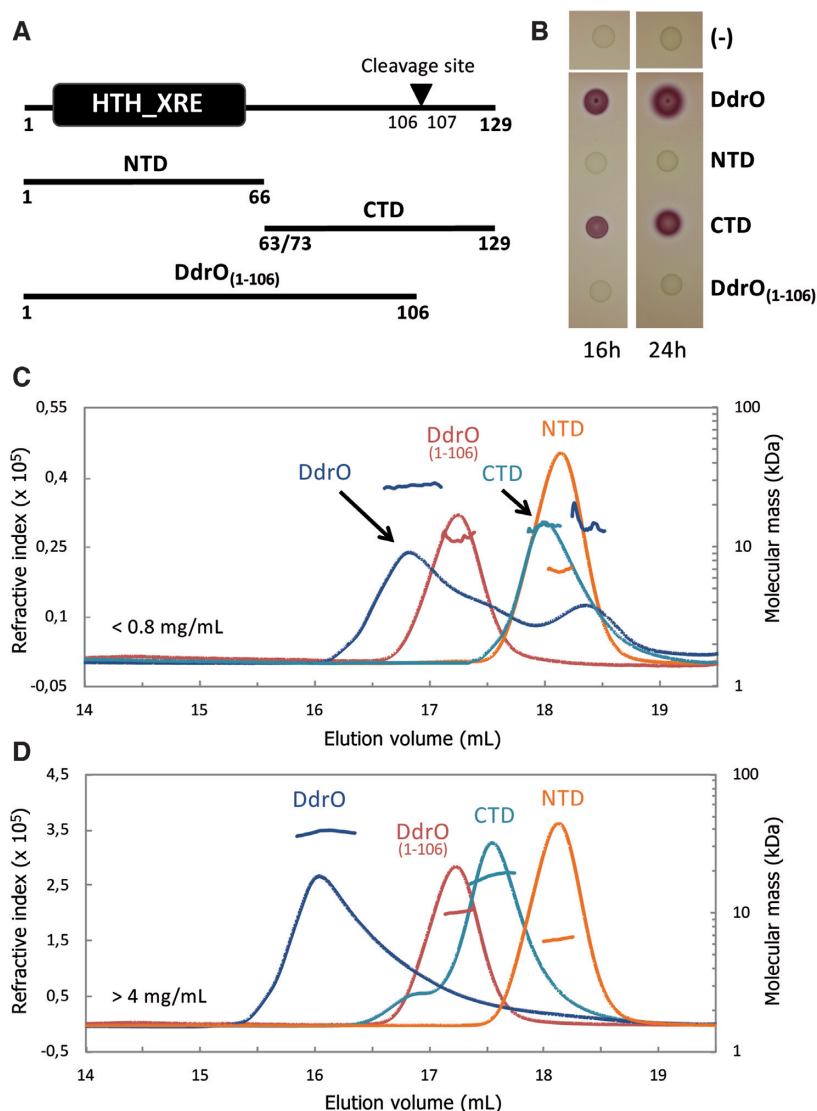
### DdrO dimerization is mediated by its C-terminal region

To investigate which part(s) of DdrO is involved in its oligomerization, a domain dissection approach was applied in association to several complementary methods for characterization. Constructs expressing fusion proteins for bacterial two-hybrid experiments or for protein purification containing the following DdrO parts were designed and prepared: (i) the N-terminal domain NTD that contains the predicted DNA-binding domain, (ii) the C-terminal domain CTD that contains the site where DdrO is cleaved by IrrE, and (iii) DdrO until residue 106, DdrO<sub>(1-106)</sub>, corresponding to DdrO until the cleavage site (i.e. DdrO lacking the C-terminal 23 residues) (Figure 1A, Supplementary Figures S3 and S4).

Bacterial two-hybrid experiments indicated an interaction between the CTDs of DdrO, but not between the NTDs and the DdrO<sub>(1-106)</sub> molecules (Figure 1B, Supplementary Figure S3). These results are in agreement with SEC-MALLS experiments for which the NTD and DdrO<sub>(1-106)</sub> are only observed as monomers (Figures 1C and D, Supplementary Figure S5, Table 1). The CTD recapitulated the full-length DdrO as it formed dimers at low protein concentrations, progressively shifting to a probable dimer-tetramer equilibrium at increasing concentrations (Figures 1C and D, Supplementary Figure S5, Table 1).

### DNA binding of DdrO

The oligomeric state of DdrO was also investigated in presence of short, RDRM-containing DNA fragments using native mass spectrometry of non-covalent complexes. The majority of the observed complexes are composed of two DdrO molecules and one double-stranded DNA molecule (Figure 2, Supplementary Figure S6), indicating that DdrO binds as a dimer to the RDRM. Complexes



**Figure 1.** Dimerization of DdrO is mediated by its C-terminal region. (A) Schematic representation of DdrO. The N-terminal DNA-binding domain of the XRE family and the position where metalloproteinase IrrE cleaves DdrO are indicated. Besides entire DdrO (top), DdrO fragments corresponding to the N-terminal domain (NTD), the C-terminal region (CTD) and DdrO until the cleavage site (DdrO<sub>(1-106)</sub>) were used in this work. The CTD starts with residue 63 in the bacterial two-hybrid experiments (which were performed before the structure of full-length DdrO was solved) and with residue 73 in all other experiments. (B) Bacterial two-hybrid analysis. The DdrO protein or DdrO fragments fused to the T18 and T25 domain of bacterial adenylate cyclase are indicated. (-) Control with plasmids that do not produce fusion proteins. Positive interactions were detected on MacConkey/lactose plates by the appearance of a red colour. (C and D) SEC-MALLS analysis of DdrO and DdrO fragments at low and high protein concentrations. Elution profiles were monitored by excess refractive index (left ordinate axis). The line under/above each elution peak shows the molecular mass distribution (right ordinate axis). Observed molecular masses are reported in Table 1. The same elution volume range is shown in (C) and (D), allowing comparison of the peak elution volumes at the low and high protein concentrations.

containing four DdrO molecules and two double-stranded DNA molecules were also observed, indicating that the sample also contains DdrO tetramer bound to two DNA molecules. Efficient binding of DdrO to the short DNA fragment was also observed using EMSA (Supplementary Figure S7).

Gel shift experiments were also used to test binding of the different DdrO domains to RDRM-containing DNA fragments. DNA binding was detected for entire DdrO, but not for NTD, CTD and DdrO<sub>(1-106)</sub> (Figure 3). Together, the data indicate that dimerization and possible tetramerization of DdrO is mediated by its C-terminal region, and

that cleavage by IrrE abolishes DdrO oligomerization and, as a consequence, stable DNA binding.

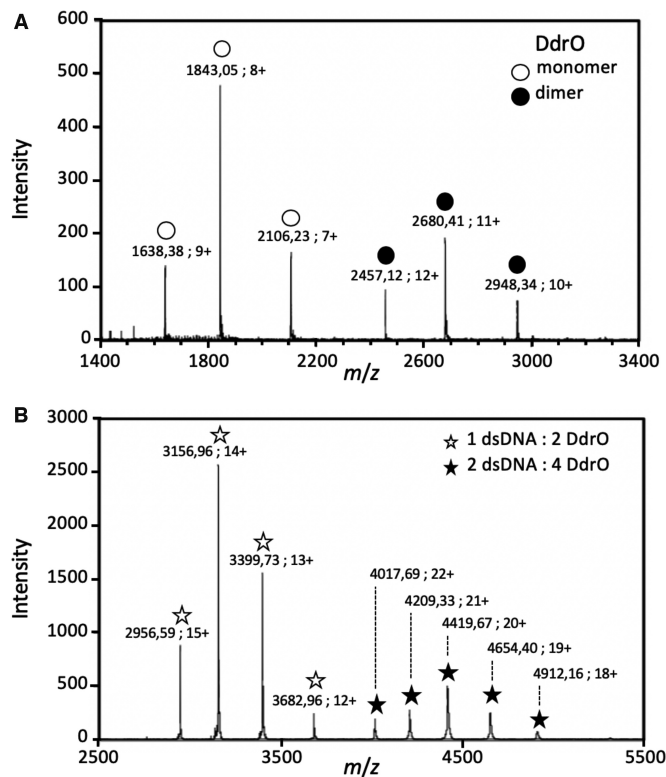
### Overall crystal structure of DdrO

In line with the conserved domain search for DdrO, the structure prediction tool Phyre2 (31) showed an N-terminal DNA-binding domain of the XRE family with good confidence, but no 3D structure was predicted for the C-terminal region of DdrO. To understand the structural basis for DdrO dimerization and function, the DdrO structure was determined using X-ray crystallography. The struc-

**Table 1.** Molecular mass determination by SEC-MALLS of full-length DdrO and of DdrO fragments at two concentrations

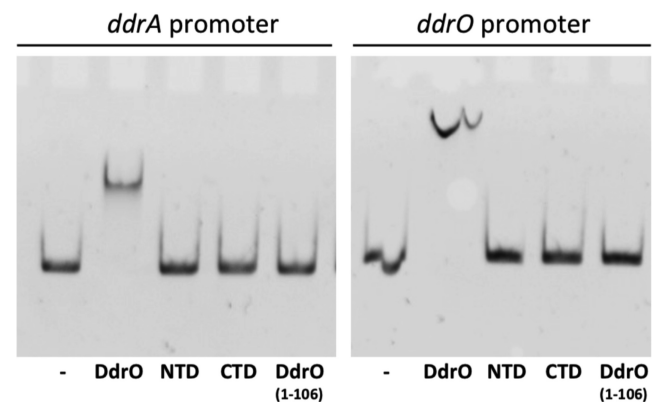
Protein	Theoretical molecular mass (Da)	Concentration		Observed molecular mass (Da) <sup>a</sup>	Calculated number of monomers
		mg/ml	mM		
DdrO	14736.8	0.74	0.05	27500	1.87
		7.78	0.53	14800	1.00
NTD	7501.5	0.47	0.063	38400	2.61
		5.4	0.72	6800	0.91
CTD	6757.9	0.41	0.061	6500	0.87
		6.38	0.94	14100	2.09
DdrO <sub>(1-106)</sub>	11783.4	0.49	0.042	18500	2.74
		4.27	0.36	11600	0.98
				10200	0.87

<sup>a</sup>Representative results of at least three experiments for each protein.



**Figure 2.** Native electrospray mass spectrometry showing non-covalent DdrO-DNA complexes with 2:1 and 4:2 stoichiometry. (A) DdrO alone showing both monomeric and dimeric state. (B) Following purification of complexes formed between DdrO and a 23-base pair DNA fragment containing the RDRM motif found upstream of *ddrB*, native mass spectrometry revealed complexes containing two DdrO molecules and one DNA molecule, or four DdrO molecules and two DNA molecules. Same results were obtained with a 21-base pair DNA fragment with the *ddrB* RDRM (data not shown).

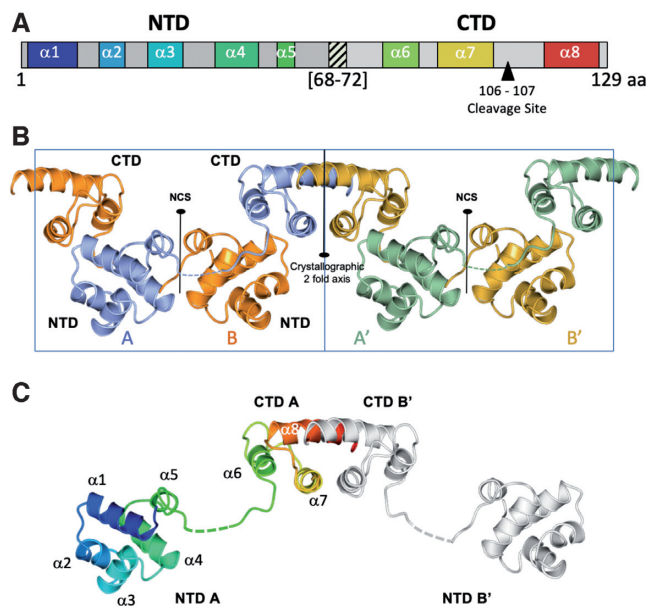
ture of the full-length DdrO was first solved using a SeMet-substituted mutant of the protein. Like most DdrO homologs, *D. deserti* Deide\_20570 lacks methionine residues (except for the N-terminus). Two methionine residues are present in DdrO (Deipe\_1111) from *Deinococcus peraridilitoris* (Supplementary Figure S1) and therefore the corresponding residues of *D. deserti* DdrO were changed to Met by site directed mutagenesis (A70M and I84M) for produc-



**Figure 3.** Binding of DdrO to RDRM-containing DNA requires a full-length protein. Gel shift experiments (EMSA) after incubation of DdrO or DdrO domains (1.7  $\mu$ M) with DNA fragments (57 nM) containing the promoter of *ddrA* or the intergenic region of the divergent *ddrO* (*Deide\_20570*) and *Deide\_20580* genes.

tion of SeMet DdrO for X-ray structure determination by the SAD method. The 3D structure of native DdrO was subsequently solved by molecular replacement. SeMet and native DdrO structure were solved at 3.1 and 2.84 Å resolution, respectively. Statistics on data collections and structure refinements are presented in Table 2 and the composite omit map for native DdrO in Supplementary Figure S8. Although crystallizing in two different space groups, the three-dimensional structures (SeMet and native) are similar with an overall rms deviation of 1.57 Å for 239 C $\alpha$  positions (Supplementary Figure S8). Remarkably, these crystal structures show a DdrO polymer, of which four monomers are depicted in Figure 4. Each DdrO monomer is composed of two folded domains corresponding to the NTD and CTD, containing respectively five and three alpha helices (Figure 4, Supplementary Figure S1). These domains are separated by three to five residues not visible in the electron density maps (residues 68–72 in the SeMet structure; 68–71 and 70–72 in respectively chain A and B of the native DdrO structure), indicating the presence of a flexible linker. The asymmetric unit is constituted of two monomers of which the NTDs form a dimer related by a non-crystallography symmetry (NCS) axis (Figure 4). In the asymmetric unit the CTDs do not interact with each other. Nevertheless, CTD dimers are present, but these are formed





**Figure 4.** Crystal structure of *D. deserti* DdrO. (A) Schematic representation of the positions of the alpha helices in DdrO as observed in the crystal structure (colour-coded as in panel C). Residues 68–72 between the NTD and CTD were not visible in the structure (residues 68–72 in the SeMet structure; 68–71 and 70–72 in respectively chain A and B of the native DdrO structure). The site where DdrO is cleaved by IrrE is located in the loop between the last two alpha helices. (B) DdrO formed a polymer in the crystal packing. Four monomers, each of them made of two folded domains (NTD and CTD), separated by a few residues not visible in the 3D structure (dotted line), are shown. The asymmetric unit (boxed with blue lines) is constituted of two monomers related by a NCS axis, with the NTDs forming a dimer. In this asymmetric unit, the C-terminal domains do not interact with each other. However, CTD dimers are observed between two DdrO monomers belonging to two different asymmetric units (only the CTD A-CTD B' interaction is shown; CTD B and CTD A' form dimers with CTDs from other DdrO monomers in the polymer). (C) View of a DdrO dimer mediated through the CTD dimerization. One monomer is colored from blue (N-terminus) to red (C-terminus) and the other in grey. This view highlights the length and flexibility of the loops connecting the NTDs to the CTDs and shows that dimerization of the NTD may easily occur in the dimer. Panel B and C show the native DdrO structure.

between the CTDs of DdrO monomers belonging to two different asymmetric units related by a two-fold symmetry axis (Figure 4). This peculiar DdrO polymer formation in the crystal is most likely a crystallization artefact (dimers-tetramers at most were observed in solution), made possible by the flexible linker between the NTD and CTD and the selection of a protein conformation that allows crystallization. The length of the linker between the NTD and CTD indicates that a DdrO dimer containing the NTD and CTD dimer structures can easily exist (Figure 4C). In the crystal packing, the linear polymers of DdrO are also interacting via the CTD dimers, making a tetramer interface involving the C-terminal helix ( $\alpha 8$ ) coming from four CTDs (Supplementary Figure S9). Considering all interfaces for a DdrO molecule in the polymer, the PISA server gives a total buried surface of  $3261 \text{ \AA}^2$  in the crystal packing (Supplementary Figure S9).

The crystal structure of entire DdrO suggests that the NTD and CTD can fold independently. To verify this, the NTD and CTD were expressed and purified separately, and

used for crystallization (Supplementary Figure S4). Crystals of both NTD and CTD were obtained more easily compared to crystals of entire DdrO for which seeding to grow crystals was necessary, probably due to the inherent flexibility of the entire protein. The 3D structures of both NTD and CTD alone were solved at 1.35 and 1.41 Å resolution, respectively (Table 2). The asymmetric unit of the NTD crystal structure contains two monomers forming a dimer that is very similar to the NTD dimer found in the structure of full-length DdrO (rms deviation 0.2 Å with native DdrO) (Supplementary Figure S10). Analysis of CTD crystal data indicated a pseudo-translation present in these crystals. However, the CTD crystal structure was solved containing eight monomers in the asymmetric unit that show the same fold as in the full length DdrO (rms deviation 0.4 Å for the CTD chain A versus the CTD of full-length native DdrO chain A) (Supplementary Figure S11). The same CTD dimers and tetramers (within the asymmetric unit or between monomers from adjacent asymmetric units) are observed as in the structure of entire DdrO (Supplementary Figures S9 and S11, Supplementary Table S4).

### The N-terminal DNA-binding domain of DdrO

The 3D structure of the N-terminal domain (NTD) of DdrO is very similar to the DNA-binding domains of previously obtained structures of some XRE family regulator proteins, notably those of SinR, the master transcriptional regulator involved in control of sporulation and biofilm formation in *B. subtilis*, and the lysis/lysogeny-regulating repressor proteins CI and Cro of bacteriophage 434 (32). The NTDs of these proteins are constituted of five alpha helices. The helix-turn-helix (HTH) motif involved in DNA recognition and binding comprises helices 2 and 3, corresponding to the scaffolding and recognition helix, respectively. Like the NTD of DdrO, the separately produced N-terminal HTH\_XRE domain of SinR is monomeric in solution but a dimer in the crystal structure (33). Moreover, SinR NTD dimer structure and interface are very similar in DNA-bound SinR and isolated SinR structures (rms deviation 1.28 Å; PDB codes 3ZKC and 3QQ6, respectively) (33,34). The DdrO NTD dimer structure superposes well with that of DNA-bound SinR (rms deviation 1.87 Å) (Figure 5, Supplementary Figure S10). However, subtle differences are observed. The NTD-NTD interface area of DdrO is  $313 \text{ \AA}^2$  and shows only one reciprocal hydrogen bond between the side chain of H59 and the main carbonyl atom of V65 indicating a weak interaction (Supplementary Figure S10). According to the PISA webserver this interface has a Complex Formation Significance Score (CSS) of 0, suggesting that it is not a dimeric interface, in accordance with SEC-MALLS and BACTH results. The SinR NTD interface is also relatively small ( $520 \text{ \AA}^2$ ) and is formed exclusively by the  $\alpha 3$ - $\alpha 4$  loop and the N-terminal half of  $\alpha 4$  showing four hydrogen bonds (between T40-Q45, N41-Q45, N41-S43 and P42-I44) (34) (Supplementary Figure S10). Four of the five residues of SinR involved in specific interaction with its 16-base pair palindromic target motif (34) are not conserved in DdrO (Figure 5). All together, DdrO presents at the N-terminus a classical HTH DNA-binding domain and the differences with SinR are expected to allow

**Table 2.** Data collection, phasing and refinement statistics

Data set	DdrO-2Met	DdrO-Wt	DdrO-NTD	DdrO-CTD
Protein	SeMet	Native	Native	Native
PDB Code	6RMQ	6RNX	6RNZ	6RO6
<b>Data collection</b>				
Beamlines (ESRF)	BM30	ID23-2	ID30A-1	ID30A-1
Space group	$P2_122_1$	$C222_1$	$P2_1$	$P2_12_12_1$
Cell dimensions				
a, b, c (Å)	34.4, 77.0, 131.2	32.9, 130.2, 154.8	38.2, 43.2, 45.9	51.9, 76.8, 103.3
$\gamma$ (°)			106.7	
Wavelength (Å)	0.9797	0.8726	0.9660	0.9660
Resolution Range (Å)	76–3.1	50–2.84	44–1.35	60–1.41
High Resolution Range (Å)	3.3–3.1	2.94–2.84	1.4–1.35	1.43–1.41
Total observations	49 095	79 276	87 582	359 298
Unique reflections	7531	8234	31 092	73 609
Completeness (%) <sup>a</sup>	100 (100)	99 (96)	98.3 (96.4)	92.4 (99.3)
I/ $\sigma$ (I)	19.5 (8.6)	8.6 (1.7)	10.9 (1.8)	17.5 (2.1)
$R_{\text{sym}}$ (%) <sup>a,b</sup>	8.9 (20.8)	15.9 (95.5)	5.7 (56.6)	3.7 (68.3)
CC <sub>1/2</sub>	99.7 (98.1)	99.6 (77.2)	99.8 (69.8)	99.9 (49.5)
CC <sub>1/2ano</sub>	99 (98)			
No. of Se sites	2			
<b>Refinement</b>				
$R_{\text{work}}/R_{\text{free}}$ (%)	19.7/28.4	22.2/29.6	12.7/17.5	19.5/22.4
No. of atoms	2025	2044	1397	3890
Protein	1982	2028	1107	3730
Water	43	16	272	110
<b>RMSD</b>				
Bonds (Å)	0.013	0.010	0.028	0.018
Angles (°)	2.0	1.194	2.3	1.5
<b>Average B factors (Å<sup>2</sup>)</b>				
Protein	26.5	68.2	15.9	35.8
Water	15.6	54.9	39.4	31.9
<b>Ramachandran plot (%)</b>				
Favored regions	93	95.04	100	98.6
Additionally allowed regions	4	4.96	0	1.4
Outliers	3	0	0	0

<sup>a</sup>Values in parentheses are for the highest resolution shell.

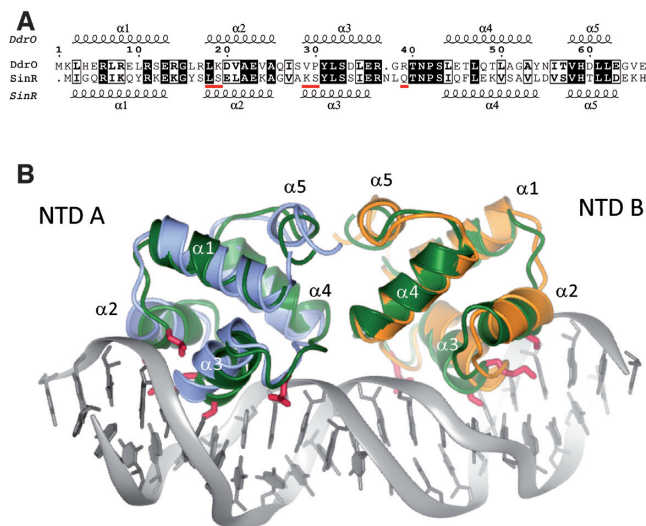
$$^b R_{\text{sym}} = R_{\text{merge}} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \bar{I}(hkl)|}{\sum_{hkl} \sum_i I_i(hkl)}$$

specific recognition of its target 17-base pair RDR palindromic motif (RDRM).

### The C-terminal dimerization domain of DdrO

The C-terminal domain (CTD) of DdrO is composed of three alpha helices arranged in a new fold (with DALI Z-scores below 5 against the PDB; Figures 4 and 6). The cleavage site, between L106 and R107, is located in the loop connecting the two last helices ( $\alpha 7$ - $\alpha 8$ ). The interface of the dimer formed by the CTD involves reciprocal ionic interactions between three residue pairs (D96-R107, R103-E105, E119-H123; Figure 6, Supplementary Table S4). The hydrogen bonds and salt bridges between D96 and R107 and between R103 and E105 involve residues close to the cleavage site. D96 and R103 are located in  $\alpha 7$ , and E105 and R107 in the  $\alpha 7$ - $\alpha 8$  loop. E119 and H123 are located in the C-terminal helix,  $\alpha 8$  (the two helices  $\alpha 8$  are oriented head to tail in the dimer; Figure 6, Supplementary Table S4). The structure thus shows a crucial role for  $\alpha 8$  and the  $\alpha 7$ - $\alpha 8$  loop in DdrO dimer formation. Especially, analysis of this dimeric interface by PISA gives a CSS of 1, indicating a clear dimeric interface that is supported by our

SEC-MALLS studies. Cleavage of DdrO by IrrE between L106 and R107 releases  $\alpha 8$  and a large part of the  $\alpha 7$ - $\alpha 8$  loop from the remainder of DdrO, largely explaining the monomeric form of cleaved DdrO (DdrO<sub>(1-106)</sub>) (Figures 1B–D). Moreover,  $\alpha 8$  contains three aromatic residues (W117, Y118 and Y121) that contribute to a hydrophobic core that is likely important for structural integrity of the CTD (Figure 6D). The side chain of Y121 points towards  $\alpha 7$  making a CH- $\pi$  interaction with W97 and together with other aromatic interactions they contribute to the correct positioning of  $\alpha 8$  with respect to the other helices in the CTD. DdrO<sub>(1-106)</sub> is monomeric in solution but at high concentration *in vitro* it shows a tendency to aggregate, which might be explained by an unstable CTD lacking  $\alpha 8$  (Supplementary Figure S5). *In vivo*, we previously observed that the two generated DdrO fragments after stress exposure (DdrO<sub>(1-106)</sub> and the 23-residue C-terminal fragment) are either faintly visible or not visible at all in Western blot experiments (11,12), probably due to a rapid degradation within the cells of these protein fragments. Within the crystals of full-length DdrO and of the CTD, an interface between two CTD dimers is present, involving the ionic interactions R126-D116 and R126-E119 (Supplementary Figure



**Figure 5.** Comparison of the N-terminal DNA-binding domains of DdrO and *B. subtilis* SinR. (A) Sequence alignment of the NTDs of DdrO and SinR. Residue numbering refers to the DdrO sequence. The five SinR residues involved in sequence-specific interaction of SinR with its target DNA are underlined (red). (B) Superposition of NTD of DdrO (monomer A in blue, monomer B in orange) and DNA-bound NTD of SinR (green) (PDB: 3ZKC) (rms deviation 1.87 Å). The five SinR residues underlined in panel A are shown as sticks.

S9, Supplementary Table S4). These residues are located in  $\alpha 8$ , and therefore also these interactions cannot be formed in DdrO cleaved by IrrE. All CTD residues highlighted in this paragraph are highly conserved in deinococcal DdrO proteins (Supplementary Figure S1). Prediction of protein binding energy using PRODIGY (35) indicates a dissociation constant of  $6 \times 10^{-6}$  M for the full-length DdrO dimer interacting through its CTD (as in Figure 4C) and of  $2.6 \times 10^{-4}$  M after cleavage, supporting the data obtained in solution.

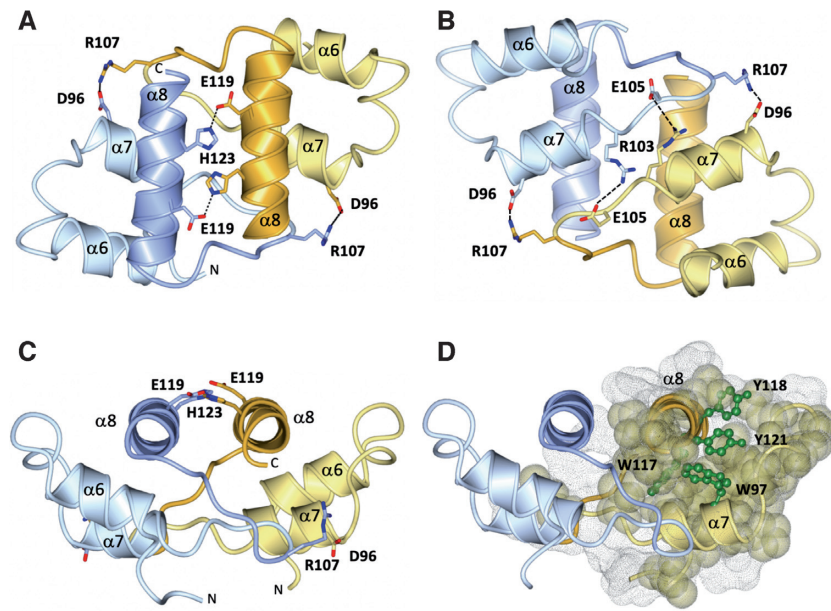
To investigate the importance of the ionic interactions involved in CTD oligomerization as observed in the crystal structures, various point mutants were constructed (Supplementary Figure S12). The DdrO mutants were produced, purified and first analysed *in vitro* by size exclusion chromatography (SEC). Almost all these DdrO mutants show less efficient oligomerization compared to wild-type DdrO (Supplementary Figure S13), with the strongest effect observed for the D96R mutant (Figure 7A), validating the involvement of the reciprocal ionic interactions in the dimerization/tetramerization state of DdrO. D96 of one monomer forms an ionic interaction with R107 of the other monomer. However, the R107D mutant is not affected in oligomerization (Figure 7A). A likely explanation is that the D96R mutation pushes away R107, whereas the D107 side chain of the R107D mutation is too short to be pushed away by D96. The important role of the D96–R107 interaction is further demonstrated by the fact that introduction of R107D in the D96R mutant restores the elution profile of wild-type DdrO (Figure 7A). Another point mutant was constructed, Y121A in  $\alpha 8$ . This mutant could also be purified but has a tendency to aggregate. Its behaviour on a SEC column is also different from native DdrO showing

two peaks with elution volumes similar to those of other mutants including D96R (Supplementary Figure S13), confirming the role of Y121 in stabilization of the 3D fold of the CTD.

The importance of some of these residues in DdrO function was also tested *in vivo* in *Deinococcus*. As *D. radiodurans* possesses only one *ddrO* gene, complementation experiments were performed in this species. The *ddrO* gene is essential for viability, and can be deleted from all chromosome copies of *D. radiodurans* only if a functional *ddrO* gene is present *in trans* on a plasmid (13). Therefore, genes encoding wild-type or mutant *D. deserti* DdrO proteins were cloned on a plasmid and introduced in *D. radiodurans*, and this was followed by attempts to delete native chromosomal *ddrO*. As expected because of the high percentage identity, wild-type *D. deserti* DdrO can substitute for *D. radiodurans* DdrO. The plasmid expressing *D. deserti* DdrO<sub>P3</sub> also allows deletion of the chromosomal *ddrO* in *D. radiodurans* (data not shown). Several DdrO mutants were tested (D96R, R107D, D96R R107D, E119A, E119A H123A and Y121A), all allowing deletion of the chromosomal *ddrO* except for the Y121A mutant that is thus not functional *in vivo* (data not shown). The *D. radiodurans* strains lacking its native *ddrO* but expressing *D. deserti* wild-type DdrO or DdrO mutants D96R, R107D or D96R R107D were analysed further. The genes encoding these proteins are cloned in fusion with the RDRM-containing promoter region of *D. radiodurans* *ddrO*, which means that these proteins repress their own expression when binding to the RDRM. A higher expression level of the D96R mutant protein compared to the other DdrO proteins was observed in these strains (Supplementary Figure S14), indicating that repression is less efficient with the D96R mutant than with the wild-type protein, in accordance with the *in vitro* data showing less efficient oligomerization (Figure 7A) and affected DNA binding (Supplementary Figure S14). The increased expression of the D96R protein possibly contributes to the formation of at least a fraction of the D96R protein pool functioning as repressor allowing cell viability. The strains were also analysed for sensitivity to the genotoxic agent MMC. Resistance to this agent requires cleavage of DdrO by IrrE, allowing induction of IrrE/DdrO-controlled DNA repair genes. R107 is located at the cleavage site, and therefore the R107D mutation is expected to have an effect on cleavage, as observed previously for the corresponding R107E mutant of *D. radiodurans* DdrO (14). Indeed, the results show an increased sensitivity for the strain with the R107D or D96R R107D mutant protein (Figure 7B), in accordance with the strongly reduced IrrE-mediated cleavage *in vitro* (Figure 7C) and *in vivo* (Supplementary Figure S14). The *in vitro* cleavage assays also show increased cleavage efficiency for the DdrO D96R mutant protein (Figure 7C, Supplementary Figure S15).

#### Cleavage of repressor DdrO by metalloprotease IrrE

Previously, we observed that IrrE does not bind to and cleave DdrO when the latter is bound to its palindromic target DNA, but that cleavage occurs with free DdrO (11). In this way, the level of repression of RDR regulon genes will correlate with the binding affinity of the repressor



**Figure 6.** 3D structure of DdrO CTD dimer and ionic interactions involved in the dimer interface. (A–C) CTD dimer shown in three orientations and indicating the three reciprocal ionic interactions (D96-R107, R103-E105, E119-H123). The parts of the protein located before and after the cleavage site targeted by IrrE are distinguished by lighter and darker colour, respectively. (D) The ribbon representation shows the DdrO dimer as depicted in (C). The hydrophobic core of the molecule is represented by (i) aromatic residues (Tyr and Trp) as dark green ball and sticks and (ii) other hydrophobic residues (Ile, Leu, Pro, Val, Ala, Gly) as khaki spheres. The overall surface of DdrO is represented by a dotted grey mesh surface.

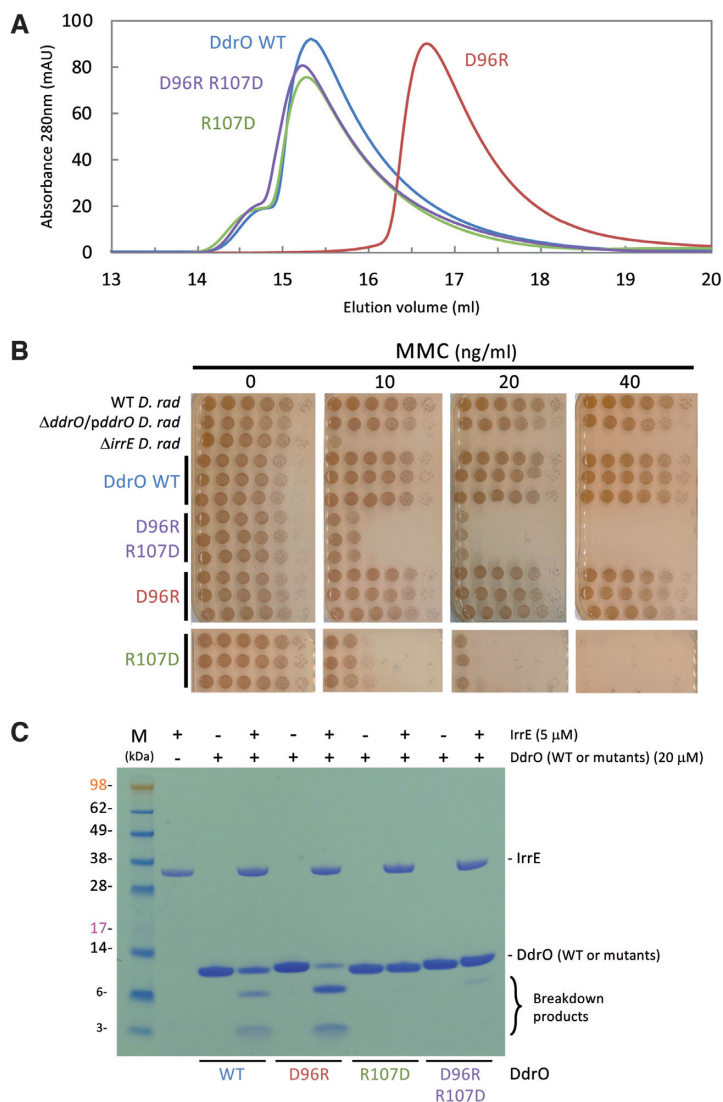
for the different target sites. DdrO is cleaved by IrrE between L106 and R107 in the 10-residue loop I104-D113 (IEL/RGKRPRD) located between the last two helices. This loop is highly basic (pI 10.74), which is compatible with the large negatively charged active-site cleft of *D. deserti* IrrE (17). Interestingly, however, this loop forms part of the CTD dimer interface and is highly buried in the CTD dimer (Figure 8, Supplementary Figure S12), suggesting that a monomer of DdrO in which this loop is exposed can be more efficiently cleaved than a dimer. This is in agreement with the increased cleavage efficiency of the DdrO D96R mutant protein (Figure 7C, Supplementary Figure S15), for which the strongest effect on oligomerization was observed (Figure 7A) and in which the cleavage site may be more accessible for IrrE if the D96R mutation pushes away R107 and the loop (Figures 6 and 8). DdrO binds as a dimer to its palindromic target DNA (Figures 1C, D, 2 and 3), and the DdrO structure solved here thus provides a rationale for why IrrE cleaves free DdrO but not when bound to DNA. The interaction of IrrE with DdrO CTD monomer or dimer was modeled using two different docking tools. Although this molecular modelling did not give a single preferential model, several IrrE-CTD monomer structural models were predicted in which the active site of IrrE is in close proximity of the cleavage site L106/R107 in DdrO (Supplementary Figure S16). Modeling of the IrrE-CTD dimer complex did not predict a satisfying complex where the scissile peptide bond of DdrO would be close enough to the active site of IrrE, supporting the suggestion that IrrE preferentially cleaves DdrO monomers. Besides the reduced accessibility of the cleavage site upon dimerization of the CTD, the presence of the NTD dimer in DNA-bound DdrO probably contributes to preventing cleavage of DNA-bound DdrO

(e.g. see Figure 6C for the position of the NTD with respect to the CTD and cleavage site). In the cells under standard conditions, DdrO exists as a dynamic equilibrium between DNA-bound and free dimers (and possibly tetramers) and free monomers (including freshly synthesized monomers). IrrE-mediated cleavage of DdrO monomers under stress condition will result in a shift of this equilibrium and a decrease of the amount of free DdrO oligomers and thus also a decrease of DNA-bound DdrO, and as a consequence in the induction of gene expression.

## DISCUSSION

We recently reported a novel stress response mechanism in *Deinococcus*, in which the COG2856 metalloprotease IrrE cleaves and inactivates DdrO, the XRE family transcriptional repressor of the radiation/desiccation response regulon (RDR) (12). The crystal structure of *D. deserti* IrrE has been solved previously (17). Here, using several complementary methods, we have structurally characterized *D. deserti* DdrO. Its 3D structure revealed a novel C-terminal domain connected to the DNA-binding domain through a flexible linker.

The N-terminal domain (NTD) of DdrO is a classical HTH\_XRE DNA-binding domain for which numerous 3D structures have already been obtained. Although the base/phosphate contacts differ in each case, the HTH modules adopt similar binding geometries with respect to their specific DNA operator sequences upstream of the regulated genes. Operator sites, like the RDR motif (RDRM) located upstream of DdrO-controlled genes, generally consist of palindromic half-sites separated by a few bases, with one monomer of a dimeric regulator binding to each half-

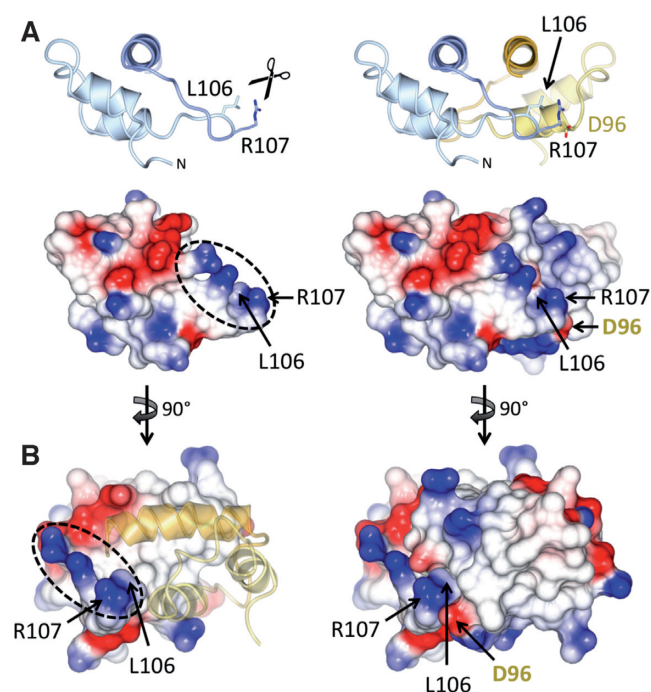


**Figure 7.** Characterization of DdrO point mutants. (A) SEC analysis of DdrO mutants compared to wild-type (WT) DdrO. Experiments were performed on a Superdex 200 Increase 10/300 GL. 500  $\mu$ L at 2 mg/ml of proteins were used. (B) MMC resistance of *D. radiodurans*  $\Delta ddrO$  mutant expressing *in trans* *D. deserti* DdrO. Eight  $\mu$ l of serial dilutions (up to  $10^{-5}$ ) of cultures were spotted (in triplicate for the cells expressing *D. deserti* wild-type or mutant DdrO) on plates containing the indicated MMC concentration. *D. deserti ddrO* genes on the plasmids were cloned in fusion with the *D. radiodurans ddrO* promoter to have regulation of *ddrO* expression similar to that in wild-type *D. radiodurans*. *D. radiodurans* wild-type (WT) and  $\Delta irrE$  strains, as well as a strain having *D. radiodurans ddrO* on the plasmid ( $\Delta ddrO/pddrO$  *D. rad*), were included as controls. (C) *In vitro* IrrE-dependent cleavage test of DdrO mutants compared to WT DdrO. Purified proteins were present in the samples (20  $\mu$ l total volume) as indicated at the top. Samples were incubated for 15 min at 37°C and then analysed by SDS-PAGE. Lane M contains molecular weight marker proteins (masses, in kDa, are indicated).

site (36,37). Accordingly, we observed that the DdrO NTD dimer structure superposes well with that of the NTD dimer of repressor SinR from *B. subtilis* bound to its palindromic target sequence. We also showed by mass spectrometry that DdrO interacts with its target RDRM as a dimer.

DdrO's C-terminal domain (CTD), which contains the cleavage site recognized by IrrE, is constituted of three alpha helices arranged in a new fold. We showed that this domain is responsible for dimerization (and possible tetramerization) of the entire DdrO mainly via the C-terminal alpha helix ( $\alpha 8$ ) and the  $\alpha 7$ - $\alpha 8$  loop. Previously, we showed that IrrE-mediated cleavage only occurs when DdrO is in solution but not when it is bound to DNA (11). The DdrO 3D structure obtained here gives structural insights into

this mechanism. DdrO binds as a dimer to its target DNA, and this requires an intact CTD. The structure showed that the cleavage site is hidden in the CTD dimer, indicating that IrrE may rather interact with and cleave DdrO monomer instead of dimer. Cleavage of DdrO monomers unbound to DNA will shift the dynamic equilibrium between monomers and DNA-bound and free dimers toward cleavable monomers, and will result in reduced amount of DdrO and thus gene induction. In this scenario, interconversion between dimeric and monomeric DdrO orchestrates the *Deinococcus* radiation/desiccation response. Another possibility is that IrrE may interact with DdrO dimer, provoking a conformational change of the  $\alpha 7$ - $\alpha 8$  loop and/or dimer dissociation rendering the cleavage site accessible.



**Figure 8.** Location of the scissile peptide bond within DdrO CTD monomer and dimer. (A) Only one monomer of the dimer is depicted at the left, showing the highly basic protruding loop (dashed lines) between  $\alpha 7$  and  $\alpha 8$  containing the scissile peptide bond between L106 and R107. The surface representation displays atoms color-coded according to the surface electrostatic potential from red (negative) to blue (positive). On the right, the dimer is depicted, showing that the loop and the cleavage site between L106 and R107 are less accessible. (B)  $90^\circ$  rotation of the molecules compared to (A) showing exposure of the loop and of the scissile peptide bond in the monomer at the left, where the peptide chain of the second monomer is indicated in transparency. In the dimer at the right the scissile peptide bond is buried with notably the presence of D96 that makes an electrostatic interaction with R107.

The scissile peptide bond between L106 and R107 may be even more inaccessible in DNA-bound DdrO dimer due to the proximity of the NTD dimer. In free DdrO dimers, the NTDs are expected to be mostly monomeric and also, because of the flexible linker, in different orientations with respect to the CTD dimer. In the crystals of full-length DdrO, both NTD and CTD dimers were formed, but the position of these dimers relative to each other in a DdrO dimer was not found because the NTDs of a CTD dimer not dimerize with each other but with an NTD of another CTD dimer. As observed previously, structural studies of flexible modular proteins such as transcriptional repressors can be hampered due to their inherent flexibility (38). The DdrO polymer obtained in the crystallization conditions highlights once more the technical challenge associated with structural studies of such flexible multi-domain proteins.

For some systems, transcription regulation is achieved by cooperative, high affinity binding of a multimeric regulator to several operator sites, with concomitant looping of the promoter DNA. In *Deinococcus*, the RDR regulon genes are located all over the genome and are generally preceded by only one predicted RDR motif, suggesting regulation by DdrO dimers. However, we do not exclude that additional DdrO target sites exist that have not

been found bioinformatically. Two RDRM sites were found in the intergenic region between *ddrO* (Deide\_20570) and *ddrQ* (Deide\_20580), and these genes might be regulated by a DdrO tetramer instead of two independent dimers. Although SEC-MALLS experiments did not show stable DdrO tetramers, some DdrO tetramers were found in presence of DNA by native mass spectrometry and a tetramer interface involving two CTD dimers was observed in the 3D structures. The tetrameric DdrO in these DdrO-DNA complexes might thus be composed of a dimer of dimers. Alternatively, a tetramer might be formed containing two NTD and two CTD dimers but with the NTDs of a CTD dimer not dimerizing with each other but with an NTD of the other CTD dimer (Supplementary Figure S17). Such alternative tetramers have been described for the repressor of a temperate *Salmonella* phage (39).

After this work was submitted and the referees' reports received, an article reporting a crystal structure of full-length DdrO from *Deinococcus geothermalis* was published by Lu *et al.* (42). Their structure is essentially identical to the full-length *D. deserti* DdrO structure presented here, also with the NTD dimer contained in the asymmetric unit (as in Figure 4B). They conclude that the dimer present in the asymmetric unit corresponds to the biologically relevant DdrO dimer, and that cleavage by IrrE impairs the formation of this dimer because of destabilization of the hydrophobic core of the CTD. Surprisingly, DdrO dimerization via its CTD was not considered and investigated in Lu *et al.* (42), although the same CTD-CTD interfaces and DdrO polymer as found for *D. deserti* DdrO are present in their crystal structure (PDB code 6JQ1). Our work includes a domain dissection approach and a number of *in vitro*, *in vivo* and *in silico* experimental data that support the model that DdrO dimerization is mediated by its CTD and that cleavage by IrrE prevents this dimerization mode.

The 3D structures of *Deinococcus* IrrE and DdrO are the only structures of widespread COG2856/XRE protein pairs that have been experimentally solved. Also, protease activity has been demonstrated for only two COG2856 domain-containing proteins, IrrE and *B. subtilis* ImmA (12,19). ImmA and IrrE share less than 15% sequence identity, mainly located around the peptidase signature HEXXH (Supplementary Figure S18). Besides the COG2856 domain, IrrE possesses a C-terminal GAF-like structural domain of unknown function. ImmA is 112 residues shorter than IrrE and, as shown in its Phyre2 model, lacks such additional domain (Supplementary Figures S18 and S19). ImmA cleaves repressor ImmR between F95 and M96 that are located in a 15-residue loop between the two last predicted alpha helices of ImmR (Supplementary Figure S18). In contrast to the basic  $\alpha 7$ - $\alpha 8$  loop of DdrO, this loop in ImmR is acidic. The protein pairs IrrE/DdrO and ImmA/ImmR are a good illustration of co-evolution of COG2856/XRE proteins involved in regulation of a stress-induced response using a similar mechanism.

Recently, the COG2856/XRE pair Rir/Crh encoded by temperate *Streptococcus thermophilus* phage TP-J34 has been studied. The lytic cycle of this phage is stimulated in stress-exposed host cells. Rir is required for prophage induction by interacting with and inhibiting oligomerization of

repressor Crh. Interestingly, this inhibition occurred without cleavage of Crh (40). Therefore, despite predicted structural similarity with the protease domain of IrrE (Supplementary Figure S19) and the presence of the HEXXH motif (Supplementary Figure S18), Rir may not be a metalloprotease, although we do not exclude that such activity remained undetected because of the experimental conditions. For the MA metallopeptidases clan that share the HEXXH motif, a larger signature has been described including a few more residues before and after HEXXH (41). The larger motif is conserved in IrrE and ImmA, but not entirely in Rir (Supplementary Figure S18), which may correlate with the absence of protease activity for the latter.

Of the many predicted COG2856/XRE pairs, the proteins IrrE/DdrO (induction of DNA repair genes in *Deinococcus*), ImmA/ImmR (induction of conjugative transposon excision and transfer in *Bacillus*) and Rir/Crh (prophage induction accompanied with host cell lysis in *Streptococcus*) have been studied experimentally. The obtained data indicate that COG2856/XRE protein pairs have co-evolved into at least two different regulatory mechanisms in which the COG2856 protein inhibits the function of the XRE repressor, either involving proteolytic cleavage or disruption of oligomerization through competitive binding without proteolytic cleavage. The results are a new illustration of the unsuspected diversity of bacterial stress response mechanisms and open the way to characterize stress-induced responses mediated by COG2856/XRE proteins in diverse bacteria.

## DATA AVAILABILITY

The coordinates and structure factors for DdrO SeMet and native DdrO, DdrO N-terminal domain (NTD, 1-66) and C-terminal domain (CTD, M72-129) have been deposited in the RCSB Protein Data Bank with the accession codes 6RMQ, 6RNX, 6RNZ, 6RO6, respectively.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Stéphanie Blangy and Nicolas Bremond for the SeMet protocol, Léa Bulteau for protein purification, Didier Nurizzo for help in data collection at the ESRF, Juan Fontecilla-Camps and Yvain Nicolet for helpful discussions and Christophe Carles for constant support.

## FUNDING

Transverse Division n°4 (Radiobiology) of the French Alternative Energies and Atomic Energy Commission (Segment n°4 Radiobiologie); Electricité De France [RB 2017-02]; Agence Nationale de la Recherche (ANR project NOVOREP). R.M. was supported by a Phare Ph.D. fellowship from the French Alternative Energies and Atomic Energy Commission and N.E. by a Ph.D. fellowship from the French Ministry of Education. Funding for open access charge: CEA.

*Conflict of interest statement.* None declared.

## REFERENCES

- Kreuzer, K.N. (2013) DNA damage responses in prokaryotes: regulating gene expression, modulating growth patterns, and manipulating replication forks. *Cold Spring Harb. Perspect. Biol.*, **5**, a012674.
- Imlay, J.A. (2015) Transcription factors that defend bacteria against reactive oxygen species. *Annu. Rev. Microbiol.*, **69**, 93–108.
- Baharoglu, Z. and Mazel, D. (2014) SOS, the formidable strategy of bacteria against aggressions. *FEMS Microbiol. Rev.*, **38**, 1126–1145.
- Bayles, K.W. (2014) Bacterial programmed cell death: making sense of a paradox. *Nat. Rev. Microbiol.*, **12**, 63–69.
- Slade, D. and Radman, M. (2011) Oxidative stress resistance in *Deinococcus radiodurans*. *Microbiol. Mol. Biol. Rev.*, **75**, 133–191.
- Daly, M.J. (2012) Death by protein damage in irradiated cells. *DNA Repair Amst.*, **11**, 12–21.
- Lim, S., Jung, J.-H., Blanchard, L. and de Groot, A. (2019) Conservation and diversity of radiation and oxidative stress resistance mechanisms in *Deinococcus* species. *FEMS Microbiol. Rev.*, **43**, 19–52.
- Tanaka, M., Earl, A.M., Howell, H.A., Park, M.J., Eisen, J.A., Peterson, S.N. and Battista, J.R. (2004) Analysis of *Deinococcus radiodurans*'s transcriptional response to ionizing radiation and desiccation reveals novel proteins that contribute to extreme radioresistance. *Genetics*, **168**, 21–33.
- Makarova, K. S., Omelchenko, M.V., Gaidamakova, E.K., Matrosova, V.Y., Vasilenko, A., Zhai, M., Lapidus, A., Copeland, A., Kim, E., Land, M. et al. (2007) *Deinococcus geothermalis*: the pool of extreme radiation resistance genes shrinks. *PLoS One*, **2**, e955.
- de Groot, A., Roche, D., Fernandez, B., Ludanyi, M., Cruveiller, S., Pignol, D., Vallenet, D., Armengaud, J. and Blanchard, L. (2014) RNA sequencing and proteogenomics reveal the importance of leaderless mRNAs in the radiation-tolerant bacterium *Deinococcus deserti*. *Genome Biol. Evol.*, **6**, 932–948.
- Blanchard, L., Guerin, P., Roche, D., Cruveiller, S., Pignol, D., Vallenet, D., Armengaud, J. and de Groot, A. (2017) Conservation and diversity of the IrrE/DdrO-controlled radiation response in radiation-resistant *Deinococcus* bacteria. *Microbiologyopen*, **6**, e477.
- Ludanyi, M., Blanchard, L., Dulermo, R., Brandelet, G., Bellanger, L., Pignol, D., Lemaire, D. and de Groot, A. (2014) Radiation response in *Deinococcus deserti*: IrrE is a metalloprotease that cleaves repressor protein DdrO. *Mol. Microbiol.*, **94**, 434–449.
- Devigne, A., Ithurbide, S., Bouthier de la Tour, C., Passot, F., Mathieu, M., Sommer, S. and Servant, P. (2015) DdrO is an essential protein that regulates the radiation desiccation response and the apoptotic-like cell death in the radioresistant *Deinococcus radiodurans* bacterium. *Mol. Microbiol.*, **96**, 1069–1084.
- Wang, Y., Xu, Q., Lu, H., Lin, L., Wang, L., Xu, H., Cui, X., Zhang, H., Li, T. and Hua, Y. (2015) Protease activity of PprI facilitates DNA damage response: Mn(2+)-dependence and substrate sequence-specificity of the proteolytic reaction. *PLoS One*, **10**, e0122071.
- Earl, A.M., Mohundro, M.M., Mian, I.S. and Battista, J.R. (2002) The IrrE protein of *Deinococcus radiodurans* R1 is a novel regulator of *recA* expression. *J. Bacteriol.*, **184**, 6216–6224.
- Hua, Y., Narumi, I., Gao, G., Tian, B., Satoh, K., Kitayama, S. and Shen, B. (2003) PprI: a general switch responsible for extreme radioresistance of *Deinococcus radiodurans*. *Biochem. Biophys. Res. Commun.*, **306**, 354–360.
- Vujicic-Zagar, A., Dulermo, R., Le Gorrec, M., Vannier, F., Servant, P., Sommer, S., de Groot, A. and Serre, L. (2009) Crystal structure of the IrrE protein, a central regulator of DNA damage repair in *Deinococcaceae*. *J. Mol. Biol.*, **386**, 704–716.
- Makarova, K.S., Wolf, Y.I. and Koonin, E.V. (2009) Comprehensive comparative-genomic analysis of type 2 toxin-antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct*, **4**, 19.
- Bose, B., Auchtung, J.M., Lee, C.A. and Grossman, A.D. (2008) A conserved anti-repressor controls horizontal gene transfer by proteolysis. *Mol. Microbiol.*, **70**, 570–582.
- Meima, R. and Lidstrom, M.E. (2000) Characterization of the minimal replicon of a cryptic *Deinococcus radiodurans* SARK

- plasmid and development of versatile *Escherichia coli*-*D. radiodurans* shuttle vectors. *Appl Environ. Microbiol.*, **66**, 3856–3867.
21. Battesti, A. and Bouveret, E. (2012) The bacterial two-hybrid system based on adenylate cyclase reconstitution in *Escherichia coli*. *Methods*, **58**, 325–334.
  22. Mossessova, E. and Lima, C.D. (2000) Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell*, **5**, 865–876.
  23. Raynal, B., Lenormand, P., Baron, B., Hoos, S. and England, P. (2014) Quality assessment and optimization of purified protein samples: why and how? *Microb. Cell Factor*, **13**, 180.
  24. Doublé, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.*, **276**, 523–530.
  25. Walden, H. (2010) Selenium incorporation using recombinant techniques. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 352–357.
  26. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A. et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67**, 235–242.
  27. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.-W., Kapral, G.J., Grosse-Kunstleve, R.W. et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.
  28. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B. et al. (2018) MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci. Publ. Protein Soc.*, **27**, 293–315.
  29. Weng, G., Wang, E., Wang, Z., Liu, H., Zhu, F., Li, D. and Hou, T. (2019) HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res.*, **47**, W322–W330.
  30. van Zundert, G.C.P., Rodrigues, J.P.G.L.M., Trellet, M., Schmitz, C., Kastiris, P.L., Karaca, E., Melquiond, A.S.J., van Dijk, M., de Vries, S.J. and Bonvin, A.M.J.J. (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725.
  31. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J.E. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
  32. Lewis, R.J., Brannigan, J.A., Offen, W.A., Smith, I. and Wilkinson, A.J. (1998) An evolutionary link between sporulation and prophage induction in the structure of a repressor:anti-repressor complex. *J. Mol. Biol.*, **283**, 907–912.
  33. Colledge, V.L., Fogg, M.J., Levdkov, V.M., Leech, A., Dodson, E.J. and Wilkinson, A.J. (2011) Structure and organisation of SinR, the master regulator of biofilm formation in *Bacillus subtilis*. *J. Mol. Biol.*, **411**, 597–613.
  34. Newman, J.A., Rodrigues, C. and Lewis, R.J. (2013) Molecular basis of the activity of SinR protein, the master regulator of biofilm formation in *Bacillus subtilis*. *J. Biol. Chem.*, **288**, 10766–10778.
  35. Vangone, A. and Bonvin, A.M. (2015) Contacts-based prediction of binding affinity in protein-protein complexes. *eLife*, **4**, e07454.
  36. Brennan, R.G. and Matthews, B.W. (1989) Structural basis of DNA-protein recognition. *Trends Biochem. Sci.*, **14**, 286–290.
  37. Huffman, J.L. and Brennan, R.G. (2002) Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.*, **12**, 98–106.
  38. Rasmussen, K.K., Frandsen, K.E.H., Boeri Erba, E., Pedersen, M., Varming, A.K., Hammer, K., Kilstrup, M., Thulstrup, P.W., Blackledge, M., Jensen, M.R. et al. (2016) Structural and dynamics studies of a truncated variant of CI repressor from bacteriophage TP901-1. *Sci. Rep.*, **6**, 29574.
  39. Kim, M., Kim, H.J., Son, S.H., Yoon, H.J., Lim, Y., Lee, J.W., Seok, Y.-J., Jin, K.S., Yu, Y.G., Kim, S.K. et al. (2016) Noncanonical DNA-binding mode of repressor and its disassembly by antirepressor. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E2480–E2488.
  40. Koberg, S., Mohamed, M.D.A., Faulhaber, K., Neve, H. and Heller, K.J. (2015) Identification and characterization of *cis*- and *trans*-acting elements involved in prophage induction in *Streptococcus thermophilus* J34. *Mol. Microbiol.*, **98**, 535–552.
  41. Jongeneel, C.V., Bouvier, J. and Bairoch, A. (1989) A unique signature identifies a family of zinc-dependent metalloproteases. *FEBS Lett.*, **242**, 211–214.
  42. Lu, H., Wang, L., Li, S., Pan, C., Cheng, K., Luo, Y., Xu, H., Tian, B., Zhao, Y. and Hua, Y. (2019) Structure and DNA damage-dependent derepression mechanism for the XRE family member DG-DdrO. *Nucleic Acids Res.*, **47**, 9925–9933.