International Journal of Biological Macromolecules 254 (2024) 127935

Contents lists available at ScienceDirect

International Journal of Biological Macromolecules

journal homepage: www.elsevier.com/locate/ijbiomac

Structural characterization of PaaX, the main repressor of the phenylacetate degradation pathway in *Escherichia coli* W: A novel fold of transcription regulator proteins

Víctor M. Hernández-Rocamora [a,b,1], Rafael Molina [c,1], Alejandra Alba [c], César Carrasco-López [c], Alzoray Rojas-Altuve [c], Santosh Panjikar [d,e], Ana Medina [f], Isabel Usón [f,g], Carlos Alfonso [h], Beatriz Galán [h], Germán Rivas [h], Juan A. Hermoso [c,*], Jesús M. Sanz [h,i,*]

[a] *Instituto de Investigación, Desarrollo e Innovación en Biotecnología Sanitaria de Elche, Universidad Miguel Hernández, Av. Universidad, s/n, E-03202 Elche, Alicante, Spain*
[b] *Centre for Bacterial Cell Biology, Biosciences Institute, Newcastle University, Newcastle upon Tyne, United Kingdom*
[c] *Department of Crystallography and Structural Biology, Instituto de Química-Física "Blas Cabrera", Consejo Superior de Investigaciones Científicas, Serrano 119, 28006 Madrid, Spain*
[d] *Australian Synchrotron, ANSTO, Clayton, Australia*
[e] *Department of Molecular Biology and Biochemistry, Monash University, Melbourne, Australia*
[f] *Crystallographic Methods, Institute of Molecular Biology of Barcelona (IBMB-CSIC), Baldiri Reixach 15, 08028 Barcelona, Spain*
[g] *ICREA: Institució Catalana de Recerca i Estudis Avançats, Pg. Lluis Companys 23, 08010 Barcelona, Spain*
[h] *Centro de Investigaciones Biológicas Margarita Salas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28049 Madrid, Spain*
[i] *Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Instituto de Salud Carlos III, Madrid, Spain*

ARTICLE INFO

Keywords:
Phenylacetyl-coenzyme A
Aromatic biodegradation
Folding intermediates

ABSTRACT

PaaX is a transcriptional repressor of the phenylacetic acid (PAA) catabolic pathway, a central route for bacterial aerobic degradation of aromatic compounds. Induction of the route is achieved through the release of PaaX from its promoter sequences by the first compound of the pathway, phenylacetyl-coenzyme A (PA-CoA). We report the crystal structure of PaaX from *Escherichia coli* W. PaaX displays a novel type of fold for transcription regulators, showing a dimeric conformation where the monomers present a three-domain structure: an N-terminal winged helix-turn-helix domain, a dimerization domain similar to the Cas2 protein and a C-terminal domain without structural homologs. The domains are separated by a crevice amenable to harbour a PA-CoA molecule. The biophysical characterization of the protein in solution confirmed several hints predicted from the structure, i.e. its dimeric conformation, a modest importance of cysteines and a high dependence of solubility and thermostability on ionic strength. At a moderately acidic pH, the protein formed a stable folding intermediate with remaining α-helical structure, a disrupted tertiary structure and exposed hydrophobic patches. Our results provide valuable information to understand the stability and mechanism of PaaX and pave the way for further analysis of other regulators with similar structural configurations.

1. Introduction

Aromatic compounds are abundant in soil and water coming from different sources. They are highly resistant to biodegradation, and their biological transformation is usually carried out by bacteria and fungi [1,2]. In particular, the degradation of phenylacetic acid (PAA) and its derivatives can be carried out by several bacteria and fungi both in aerobic and anaerobic conditions (see [3,4] for extensive reviews). The aerobic degradation of PAA by bacteria starts with the esterification of PAA with coenzyme A (CoA) by a phenylacetyl-CoA (PA-CoA) ligase, such as the PaaK protein, a usual mechanism in anaerobic degradation pathways [3,5]. The route is subsequently continued by the enzymes encoded in the *paa* gene clusters [6]. PA-CoA is therefore the common intermediate of the large PA-CoA catabolon [3,5]. The importance of

* Corresponding authors.
E-mail addresses: xjuan@iqfr.csic.es (J.A. Hermoso), jmsanz@cib.csic.es (J.M. Sanz).
[1] Equally contributing authors.

https://doi.org/10.1016/j.ijbiomac.2023.127935
Received 20 July 2023; Received in revised form 30 October 2023; Accepted 5 November 2023
Available online 9 November 2023

these pathways relies on the fact that the PAA degradation routes constitute the main CoA-dependent metabolic pathway for aromatic compounds in bacteria [3], being present in around 16 % of the bacterial sequence genome [7]. However, despite its widespread presence in the microbial realm, the role of PAA is still largely unknown [3].

The GntR-type transcriptional regulator PaaX, which is encoded in one of the *paa* clusters, represses the expression of the PAA aerobic degradation pathway enzymes, binding to the *Pa*, *Px* and *Pz* operator sequences of the clusters in the absence of PA-CoA and releasing when the ligand is present, thus activating the route [6]. PaaX also inhibits the expression of the enzyme penicillin G acylase (PGA), a scavenger enzyme that hydrolyzes esters and amines of PAA [8,9] and is also involved in the repression of the enzymes of the styrene degradation pathway in *Pseudomonas* sp. strain Y2, which produces PAA as its final catabolite [10]. Therefore, PaaX is currently considered a central regulator in the aerobic degradation of aromatic compounds derived from PAA in several microorganisms [5].

The consensus DNA binding site for PaaX in *E. coli* W has been postulated to be a palindromic sequence composed of two moieties of six nucleotides separated by 27 base pairs [8]. The PaaX binding site is situated overlapping the RNA polymerase (RNAP) binding site or immediately downstream it, hence preventing the formation of the RNAP complex [11,12]. Conversely, the PaaX binding site is upstream the RNAP binding site of the PGA promoter and, in this case, PaaX attachment has been reported to prevent binding of cAMP receptor protein (CRP), which is necessary for the activation of the RNAP complex [8]. This last mechanism of transcriptional repression is uncommon among transcriptional regulators.

It has been shown that the accumulation of toxic intermediates of phenylacetate catabolism might also promote the virulence of pathogens [3] such as *Burkholderia cenocepacia* and *Mycobacterium abscessus*, which are involved in lung infections in immunocompromised patients or in cystic fibrosis cases [7,13]. Moreover, the regulation of the PAA route seems to be related to the response of *Acinetobacter baumannii* to antibiotics [14,15]. In addition, it has also been suggested that PAA catabolic routes by *Escherichia coli* and other gut bacteria could lead to the accumulation of phenols with incidence in some forms of leukemia [7,16]. Therefore, the importance of the PAA route and its regulation might encompass health issues such as bacterial pathogenicity as well as antimicrobial resistance [3,7].

Little is known about the structure of PaaX and other GntR-type regulators. The crystal structures of similar repressors from *Jannaschia* sp. (PDB: 3L09) and *Mycobacterium tuberculosis* (PDB: 3KFW) have been deposited in the Protein Data Bank database, but no further structural or biochemical studies have been carried out on these proteins [3,4]. In this study, we report the crystal structure of a histidine-tagged version of PaaX from *E. coli W* along with a characterization of protein stability, oligomerization and substrate binding using techniques such as analytical ultracentrifugation, fluorescence spectroscopy and circular dichroism. This structural analysis provides important cues in understanding the mechanisms by which PaaX binds to DNA and its inducer PA-CoA and thus represses or enables transcription.

2. Materials and methods

2.1. Bacterial strains, plasmids and growth conditions

E. coli MV1184 [17] was used to express a histidine-tagged PaaX fusion protein. *E. coli* AF141X is a reporter strain used to study the biological function of PaaX in vivo [18]. It is an *E. coli* W14 derivative [12] with the insertion in its chromosome of a mini-Tn5Km2 *Px::lacZ* transposon, which confers resistance to rifampicin. As a W14 strain derivative is auxotroph for vitamin B12 and has a deletion of the *paa* gene clusters and of the *lacZ* gene. The mini-Tn5Km2 *Px::lacZ* transposon contains a fusion of *lacZ* gene with the *paaXY* operon, *Px*. This reporting fusion is useful for studying the regulation of the *Px* operon by PaaX.

pX2BS is derived from pQE32 (QIAGEN) with the *E. coli* W *paaX* gene cloned in its poly-linker as a *BamHI/SacI* fragment obtained via PCR from pAAD plasmid [6] using X5-Bam and X3-Sac oligonucleotides (Table S1). pAFX2 is a derivative of pCK01 [19], a low copy number plasmid, and it expresses the *E. coli* W *paaX* gene under the control of a P_{lac} operon [6]. pAFK5 is a derivative of pUC18 [17] that expresses the *paaK* gene under the control of a Plac operon [6]. pAAD is also a derivative of pCK01 and contains the complete *paa* pathway genes [6].

E. coli AF141X and plasmids pX2BS, pAFK5, pAAD and pAFX2 were generously donated by Dr. José Luis García from the Centro de Investigaciones Biológicas Margarita Salas, Madrid, Spain.

E. coli AF141X was grown at 30 or 37 °C in M63 minimal medium [20] supplemented with vitamin B12 (1 μg/mL), glycerol (20 mM) and, when indicated, PAA (5 mM). When appropriate, antibiotics were added at the following concentrations: ampicillin (100 μg/mL), chloramphenicol (35 μg/mL), kanamycin (50 μg/mL) and tetracycline (20 μg/mL).

2.2. Site-directed mutagenesis

Site-directed mutagenesis to construct the cysteine to alanine mutants (C168A, C189A, C264A, C312A) was performed using the QuikChange Site-Directed Mutagenesis Kit (Agilent) using the complementary oligonucleotides displayed in Table S1.

2.3. Expression and purification of PaaX

E. coli MV1184 competent cells were transformed with pX2BS plasmid. An LB culture inoculated with a fresh single transformant colony was grown overnight at 37 °C at 200 rpm and then diluted 100 times in 1 L LB medium and grown at 30 °C at the same speed until the optical density at 600 nm was 0.5–0.6. Then IPTG was added to a final concentration of 1 mM. The culture was incubated overnight at 30 °C to allow the expression of PaaX.

Cells were pelleted at 4 °C at 5000 rpm for 15 min in a Beckman L8–70 centrifuge with a JA-10.000 rotor. Cells were resuspended in 40 mL of Im20 buffer (50 mM sodium phosphate pH 8, 300 mM NaCl and 20 mM imidazole) per liter of culture and sonicated in ice using 8 cycles of 15 s, separated by 30-sec intervals, in a Branson 250 sonicator. The sonicated cells were centrifuged at 4 °C and 9000 rpm for 20 min in a JA-20.000 rotor. The supernatant contained most of the expressed PaaX as deduced from SDS-PAGE analysis (15 % acrylamide) [21]. The supernatant was incubated with 5 mL chelating sepharose resin (Chelating Sepharose Fast Flow, Amersham Biosciences) loaded with Ni^{2+}, and gently shaken for 1 h at room temperature (RT). The resin was washed twice with Im20 buffer and twice with Im100 buffer (as Im20 but containing 100 mM imidazole instead). Finally, the resin was packed in a chromatographic column and eluted with Im400 buffer (50 mM sodium phosphate pH 8, 500 mM NaCl, 400 mM imidazole and 20 % glycerol). Glycerol was added at this stage to enhance the stability and solubility of the protein. Samples for analytical ultracentrifugation were eluted in 20 mM Tris-HCl pH 8 buffer with 500 mM NaCl and 400 mM imidazole. Fractions of 1.5 mL were collected. The purity of the eluted protein was checked by SDS-PAGE.

The most concentrated fractions according to SDS-PAGE were desalted to eliminate imidazole using HiTrap Desalting columns (Cytiva) as described by the manufacturer's instructions. UV spectroscopy was then used to measure protein concentration. Molar extinction coefficient at 280 nm (ε_{280nm}) was calculated with Eq. (1) [22]:

$$\varepsilon_{280nm}\left(M^{-1}cm^{-1}\right) = NW \times 5500 + NY \times 1490 + NC \times 125 \quad (1)$$

where NW, NY and NC represent the number of tryptophan, tyrosine and cysteine residues, respectively. This formula produces a ε_{280nm} for PaaX and PaaX-C4 of 46,553 M^{-1} cm^{-1}.

2.4. Determination of free cysteines

Ellman's reagent (5,5′-dithio-bis-(2-nitrobenzoic) acid or DTNB, Sigma) was used to determined free cysteines in PaaX and its mutants [23]. Briefly, a DTNB stock was freshly prepare at 40 mM in 100 mM phosphate buffer at pH 8. Then, 0.2 mg/mL protein, 1 % SDS, and 1 mM DTNB were mixed in a total volume of 600 μL. After a 30 min incubation at RT, absorption at 412 nm was measured to determine the amount of thionitrobenzoic acid (TNB) formed, using the reported extinction coefficient 13,600 M^{-1} cm^{-1}, which is proportional to the moles of free cysteines present in the sample. Experiments were repeated twice with different samples.

2.5. β-Galactosidase assay

β-Galactosidase activities were measured with permeabilized cells from cultures grown to mid-log phase, as described by Miller [20].

2.6. Electrophoretic mobility shift assays (EMSA)

A radioactively labelled 280-base-pair probe containing the *Pa* promoter was prepared by PCR using oligonucleotides PAP and ^{32}P-labelled PA5-1 (Table S1), and the plasmid pAAD as template [6]. Labelling of PA5-1 was performed using T4 polynucleotide kinase (Merck) and [γ-^{32}P]ATP (3000 Ci/mmol, Perkin Elmer).

The binding reaction contained the specified amount of purified PaaX or PaaX-C4 and 0.5 nM of ^{32}P-labelled probe in 20 mM Tris-HCl pH 7.5 with 10 % glycerol, 50 mM KCl and 2 mM 2-mercaptoethanol. Samples were incubated for 15 min at RT. Electrophoresis of the binding reaction was carried out in a 5 % polyacrylamide gel polimerized in 0.25× TBE (Tris-borate-EDTA buffer) run at 60 V for 2 h at 4 °C in 0.5× TBE (45 mM Tris borate, 1 mM EDTA). The gels were then dried onto Whatman 3MM paper and exposed to Hyperfilm MP (Cytiva Amersham).

2.7. Crystallization, X-ray data collection and processing

PaaX crystal production, X-ray data collection and processing were carried out as previously described [24].

2.8. Crystal structure solution, model building and refinement

PaaX crystal structure was solved by molecular replacement method, as implemented in MORDA [25] from the AutoRickshaw pipeline [26,27] using an ab initio model given by AlphaFold2 [28] as a search model. A similar model was also achieved combining the model given by AlphaFold2 and ARCIMBOLDO_SHREDDER pipeline [29–31]. In this case, the multicopy mode was activated to locate with Phaser four copies of the prioritized fragments [32]. After the multicopy step, density modification and auto-tracing with SHELXE [33] was performed on the best probes. The resulting trace was characterized by a 38 % CC for 729 residues. Additional model building with SHELXE [34] completed the model with side-chains, revealing that the structure only contained three of the four expected copies. The final solution had 804 residues out of the total 1308 and a CC of 46 %. A merged model derived from both pipeline was then initially subjected to iterative cycles of model building and refinement with Coot [35] and REFMAC [36], respectively. Final cycle refinements were performed with PHENIX [37] yielding the refinement and data collection statistics summarized in the Table 2. The PaaX final model has a R_{work}/R_{free} of 0.19/0.23 with 0.99 % of the residues in disallowed regions of the Ramachandran plot. Figures were generated using PyMOL [38] and UCSF ChimeraX [39]. Coordinates were deposited in the Protein Data Bank database (PDB code: 8A39).

A simulation movie representing the possible interdomain movements of PaaX was created with UCSF Chimera 1.16 program (developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311) [40], as a morphing sequence containing the three different monomer conformations found in the unit cell (https://saco.csic.es/index.php/s/TeYjq7293yYzZ8R, Supplementary movie M1).

2.9. Protein homology search

Methods used for sequence and structural homology search were HHPred [41], DALI [42] and Foldseek [43].

2.10. Analytical ultracentrifugation

Analytical ultracentrifugation of PaaX was performed at several protein concentrations, 2–35 μM for sedimentation equilibrium and 5–20 μM for sedimentation velocity. PaaX C-4 sedimentation velocity experiments were performed in a range of 5–38 μM. Experiments in the presence of PA-CoA were carried out at a concentration of 50 μM, more than ten-fold the dissociation constant (K_d) determined by fluorescence spectroscopy (2.5 μM). All samples were processed in 20 mM Tris-HCl at pH 8.0 with 500 mM NaCl, and 1 mM (tris(2-carboxyethyl-phosphine) (TCEP) in the case of wild-type PaaX. Glycerol was removed in analytical ultracentrifugation experiments because it increases the density and viscosity of the solutions, in addition to magnifying solvation contributions to the buoyancy of macromolecules. These modifications might interfere in the correct determination of sedimentation coefficient by sedimentation velocity and of the molecular weight by sedimentation equilibrium.

Sedimentation velocity runs were carried out at 48000 rpm and 20 °C in a XL-I analytical ultracentrifuge (Beckman-Coulter Inc.) with a UV–Vis optics detection system, using an An50Ti rotor and 12-mm double-sector centerpieces. Absorbance scans were collected at 280 nm. Additional wavelengths of 290 nm or 295 nm were used when PA-CoA or CoA were present. Sedimentation coefficient distributions were calculated by least-squares boundary modeling of sedimentation velocity data using the *c(s)* method [44], as implemented in the SEDFIT program. These s-values were corrected to standard conditions (water, 20 °C, and infinite dilution) using the SEDNTERP program [45] to obtain the corresponding standard s-values ($s_{20,w}$).

Sedimentation equilibrium studies were conducted to determine the state of association of PaaX. The sedimentation equilibrium runs were carried out at multiple speeds (9500, 11,000 and 13,000 rpm) and wavelengths (230, 250, 280 and 290 nm) with short columns (90–100 μL), using the same experimental conditions and instruments as for the sedimentation velocity experiments. After the equilibrium scans, a high-speed centrifugation run (48,000 rpm) was done to estimate the corresponding baseline offsets. The weight-average buoyant molecular weight of PaaX was determined by fitting data to the single species model using the Hetero Analysis program (retrieved from the FTP site of the Analytical Ultracentrifugation Facility of the University of Connecticut, Storrs, CT, ftp://spin6.mcb.uconn.edu). The molecular weight of the protein was determined from the experimental buoyant masses using 0.743 as the partial specific volume of PaaX as calculated from the amino acid composition using the SEDNTERP program [45].

2.11. In silico ligand docking methods

The PDB 3D model of PA-CoA was created with ChemDraw 22.2 and subjected to energy minimization using Chem3D 22.2 (Perkin Elmer Informatics). The molecule was then prepared for docking using UCSF Chimera 1.16 [40]. Both target (dimer extracted from PDB 8A39) and ligand files were submitted to the SwissDock server (http://www.swissdock.ch/) [46], and generated complexes with the lowest energies were selected for analysis.

2.12. Fluorescence spectroscopy

2.12.1. Fluorescence spectra

Emission scans were carried out at 20 °C on an Aminco SLM8000 spectrofluorometer using a 5 × 5 mm path length cuvette and a protein concentration of 0.1 mg/mL (2.73 μM). Tryptophan emission spectra were obtained using an excitation wavelength of 295 nm for the PA-CoA binding experiments, with excitation and emission slits of 4 nm and a scan rate of 60 nm/min. 8-anilinonaphthalene-1-sulfonic acid (ANS) emission spectra were acquired using an excitation wavelength of 380 nm, with excitation and emission slits of 2 nm and a scan rate of 60 nm/min.

2.12.2. Analysis of ligand binding

Binding to both CoA and PA-CoA was followed by monitoring tryptophan fluorescence emission from 300 to 400 nm, with excitation at 295 nm. For PA-CoA titrations, the protein concentration was 3 μM, and PA-CoA (Sigma) was added from a more concentrated stock up to a concentration of 100 μM. For CoA titrations, PaaX was at 2.72 μM, and CoA (Sigma) was added from a more concentrated stock up to a concentration of 10 mM. In both cases fluorescence emission intensity was measured and the normalized decrease of intensity at 340 nm was then plotted against PA-CoA or CoA concentrations to obtain binding curves.

Binding curves for PA-CoA were analyzed using the simplest model for one identical and independent binding site per monomer when ligand is not in excess of protein (Eq. (2)),

$$M + L \rightleftarrows ML \tag{2}$$

where M is the unbound protein monomer, L is the ligand and ML is the complex formed by the monomer and bound ligand. The equilibrium is governed by a binding constant (K_b)

$$K_b = \frac{[ML]}{[M][L]} \tag{3}$$

where K_b represents the binding constant. We can define the saturation fraction (θ) as the ratio between the concentration of the complex formed ($[ML]$) and the total concentration of monomer ($[M]_T$). Eq. (3) can then be expressed in these terms:

$$K_b = \frac{\theta}{(1-\theta)[M]_T\left([L]_T - \theta[M]_T\right)} \tag{4}$$

Rearranging Eq. (4) to express θ as a function of $[L]_T$:

$$\theta = \frac{\left(1 + \frac{[L]_T}{[M]_T} + \frac{1}{K_b[M]_T}\right) - \sqrt{\left(1 + \frac{[L]_T}{[M]_T} + \frac{1}{K_b[M]_T}\right)^2 - 4\frac{[L]_T}{[M]_T}}}{2} \tag{5}$$

Since the titrations reached saturation, we considered the normalized change in fluorescence ($\Delta F/\Delta F_{max}$) as a direct measure of the saturation fraction θ. Protein concentration was fixed and PA-CoA concentrations and the normalized change in fluorescence signal were then inputted as variables, in order to obtain K_b by using the SigmaPlot 10.0 software (Systat Software). Finally, dissociation constants (K_d) were calculated as the inverse of K_b.

Binding curves for CoA were analyzed by a simplified model for one binding site per monomer, assuming an excess of ligand compared to protein, so $[L] \approx [L]_T$. Therefore:

$$\theta = \frac{\Delta F}{\Delta F_{max}} = \frac{[ML]}{[M]_T} = \frac{[L]}{K_d + [L]} \approx \frac{[L]_T}{K_d + [L]_T} \tag{6}$$

so K_d was directly estimated with an hyperbolic function using SigmaPlot 10.0.

2.12.3. pH titration

Protein samples were prepared at 0.1 mg/mL and pH was adjusted using different buffers for each pH range: sodium phosphate buffer for pHs 6–8, acetate buffer for pHs 4–5.6, glycine-HCl for pHs 2.4–3.6 and KCl-HCl for pHs lower than 2.4. Buffer concentration was 100 mM and NaCl was kept at 500 mM. After measuring tryptophan fluorescence emission of each sample, pH was checked in situ with a Crison Basic-20 pH meter using a miniprobe.

The titration data were analyzed according to Eq. (7) [47]

$$F = \frac{F_{pH7} + F_{pH2}10^{-n(pH-pK)}}{1 + 10^{-n(pH-pK)}} \tag{7}$$

where F represents observed fluorescence signal, F_{pH2} and F_{pH7} are the fluorescence signal of the protein at pH 2 an pH 7, respectively, n is the number of protons involved in the transition and pK is the average pK of the groups involved in the transition.

Table 1
Functional activity and thermostability of cysteine to alanine PaaX mutants. Induction was measured as β-galactosidase activity in *E. coli* AF141X [pAFX2-derived mutants, pAFK5]. Values are presented relative to the wild-type protein. Averages of at least two experiments are shown. Errors were lower than 15 %. The thermal unfolding midpoint temperatures (t_m) were measured on purified hexahistine tagged proteins following the changes in ellipticity on increasing temperature (Fig. 7A). ND: Not determined.

Protein	Induction (%)	t_m (°C)
Wild-type	100	55.0 ± 0.1
C168A	34	ND
C189A	34	ND
C264A	168	ND
C312A	98	ND
C168A/C264A/C312A	34	51.2 ± 0.1
C189A/C264A/C312A	13	52.5 ± 0.1
C168A/C189A/C264A/C312A	23	52.2 ± 0.1

2.13. Circular dichroism

Circular dichroism (CD) experiments were performed in a Jasco J-810 spectropolarimeter equipped with a Peltier PTC-423S system. Isothermal wavelength spectra were acquired at a scan speed of 50 nm/min with a response time of 2 s and averaged over at least four scans at 20 °C. Protein concentration was 0.1 mg/mL (2.73 μM) and the cuvette path length was 0.1 or 0.2 cm. Buffer was 20 mM sodium phosphate pH 8, 500 mM NaCl and 20 % glycerol unless otherwise stated. Samples were centrifuged 5 min prior to CD measuring. Ellipticities ($[\theta]$) are expressed in units of (deg cm^2 dmol^{-1}), using the residue concentration of protein after centrifugation. For CD-monitored temperature-scanning denaturation experiments the sample was layered with mineral oil to avoid evaporation, and the heating rate was 60 °C h^{-1}. Since the transitions were not reversible, a thermodynamic analysis could not be carried out and the curves were only fitted to sigmoidal transitions in order to calculate the temperature midpoint (t_m).

3. Results and discussion

3.1. Role of cysteine residues and engineering of a crystallizable PaaX variant

Our first attempts to determine the three-dimensional structure of PaaX by X-ray crystallography were hampered by the poor solubility of the protein. It has been described that one of the possible sources of insolubility might involve free cysteine residues that usually behave as strong hydrophobic entities [48] or that may react with other thiol groups [49]. Additionally, free cysteine residues are known to difficult the ordered arrangement of the protein molecules during crystallization trials [50,51]. In this regard, PaaX contains four cysteine residues in its sequence, namely Cys-168, Cys-189, Cys-264 and Cys-312. We performed an Ellman assay on purified PaaX which yielded an average of

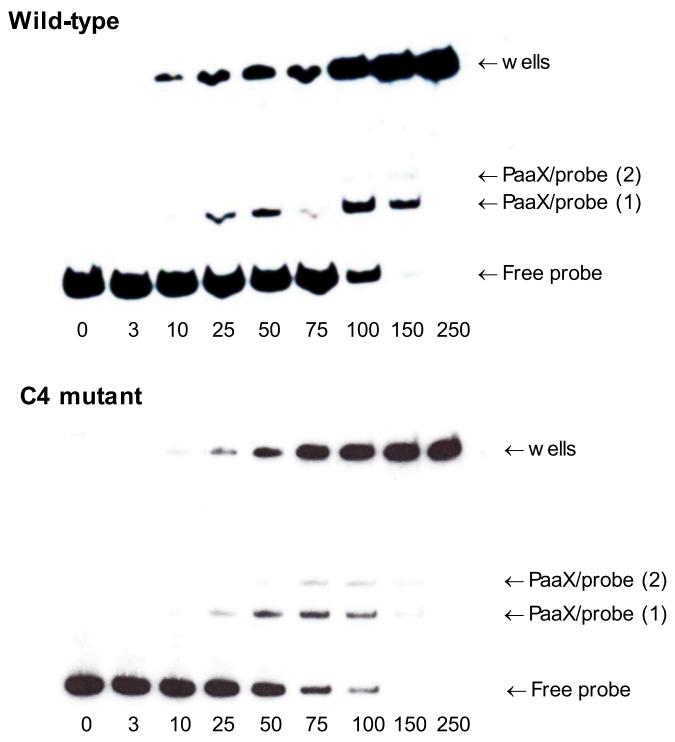

Fig. 1. Binding of PaaX and PaaX-C4 to DNA in vitro. Electronic mobility shift assays (EMSA) using a 280-bp probe containing the sequence of the PaaX binding site from the *Pa* promoter (from position −87 to +187 relative to the major transcription start). In both cases, each lane was loaded with 0.5 nM ^{32}P-labelled probe in 20 mM Tris-HCl pH 7.5 with 10 % glycerol, 2 mM 2-mercaptoethanol and 50 mM KCl. Purified PaaX or PaaX-C4 was added to the wells at the concentrations indicated under each lane (nanomolar). Aggregation was visible with both proteins, although complex bands are evident at protein concentrations between 10 and 150 nM.

3.9 ± 0.3 mol of thionitrobenzoic acid per mole of protein, therefore discarding the presence of disulfide bridges. We then decided to exchange each of the four cysteine residues in PaaX by alanine and analyze the role of these residues in the activity of the protein. With this aim, we employed an in vivo reporter assay on the repression of the *lacZ* gene under the control of the *Px* promoter in the *E. coli* strain AF141X [18] (Table 1). PaaX variants were expressed after site-directed mutagenesis of the *paaX* gene in plasmid pAFX2 [6]. In these assays, β-galactosidase activity was only detected when the strain was grown in the presence of both phenylacetic acid (PAA) and a second, compatible plasmid, pAFK5, which encodes the PaaK PA-CoA ligase from *E. coli* W [6]. In the absence of inducer, all the PaaX variants completely repressed *lacZ* expression, with no measurable β-galactosidase activity unless PAA was added to the culture medium (Table 1). The C312A mutant displayed a similar induction level as the wild-type protein, whereas the C264A mutant showed a 1.7-fold higher induction and C189A and C168A mutants exhibited 3-fold lower activations. Triple mutants C168A/C264A/C312A and C189A/C264A/C312A both displayed induction levels lower than the wild-type form, indicating that the hyperactivating C264A mutation could not overcome the effect of the C168A or C189A changes. The quadruple mutant with no cysteines at all (C168A/C189A/C264A/C312A) induced a roughly similar activity (23 %) than the single C168A and C189A mutants, as well as the triple mutants, indicating that the decrease in activity due to the C168A and C189A mutations was not additive. In any case, while the differences in activation may truly reflect the effect of conformational changes in the mutants as compared to the wild-type form, their non-additivity and moderate scope (less than one order of magnitude) might also reflect different levels of expression of the protein in the cell.

Table 2
X-ray crystallographic data collection and refinement statistics.

Data collection	
Space group	C2
Cell dimensions	
a, *b*, *c* (Å)	167.88, 106.23, 85.87
α, β, γ (°)	90, 108.33, 90
Wavelength (Å)	0.93340
Resolution (Å)	44.20–2.30 (2.40–2.30)[a]
R_{pim}	0.04 (0.40)
CC(1/2)	0.99 (0.99)
Mean I/σ(I)	20.9 (2.8)
Completeness (%)	99.9 (99.9)
Redundancy	3.7 (3.7)
Refinement	
Resolution (Å)	44.20–2.30
No. Reflections	63,235
R_{work}/R_{free}	0.19/0.23
Molecules per a.u.	3
No. Atoms	
Protein	7209
Ligands	112
Water	181
Ramachandran	
Favored/Allowed (%)	95.49/3.52
Disallowed (%)	0.99
R.m.s. deviations	
Bond lengths (Å)	0.011
Bond angles (°)	1.361
PDB code	8A39

[a] Values in parentheses are for highest-resolution shell. One crystal was used to solve the structure.

In any case, as the Cys → Ala mutants were able to both repress transcription and be regulated by PA-CoA, we overexpressed and purified a histidine-tagged variant of PaaX with all four cysteines changed to alanine, (PaaX-C4, hereafter) by immobilized metal affinity chromatography (IMAC), observing that the C4 variant turned out to be much more soluble than its wild-type counterpart at high protein concentrations. To confirm the integrity of the PaaX-C4 protein, an electrophoretic mobility shift assay (EMSA) was carried out to check its ability to bind to a DNA probe containing the PaaX binding site from the *Pa* promoter [6,8] (Fig. 1). Both PaaX versions formed two complexes with this probe, one appearing at ~50:1 protein:DNA molar ratio and a second, less abundant one with lower mobility at somewhat higher protein concentrations (150–200:1 protein:DNA molar ratio). Aggregation of both proteins was observed at the highest concentrations as the signal of the marked probe was visible in the loading wells (Fig. 1). Overall, these results indicate that both purified variants are active in vitro with similar binding efficiency. Therefore, we used PaaX for most biochemical assays described below, whereas we used its more soluble PaaX-C4 version for the crystallographic experiments.

3.2. Crystal structure of PaaX-C4

3.2.1. General overview

PaaX-C4 crystals yielded high diffraction-quality data to 2.3 Å resolution and the structure was solved in parallel by molecular replacement using the model provided by the AlphaFold ab initio structure prediction algorithm [28] implemented in Auto-Rickshaw [52] and by using ARCIMBOLDO_SHREDDER pipelines [29–31] (Table 2). The structure of PaaX is composed of an N-terminal winged helix-turn-helix (wHTH) domain (residues 1–99), a βαβαβ-fold middle domain (residues 105–173) showing, as described below, similarity with the Cas2 protein from *Enterococcus faecalis* (Cas2-like domain, hereafter) [53], and a C-terminal domain without any other previous similitude (residues 177–307) (Fig. 2A–B). The final eight residues (308–316) were not included in the model due to the lack of electron density for this region. Three independent molecules were found in the asymmetric unit showing a very similar overall fold (Fig. S1) but with slightly different

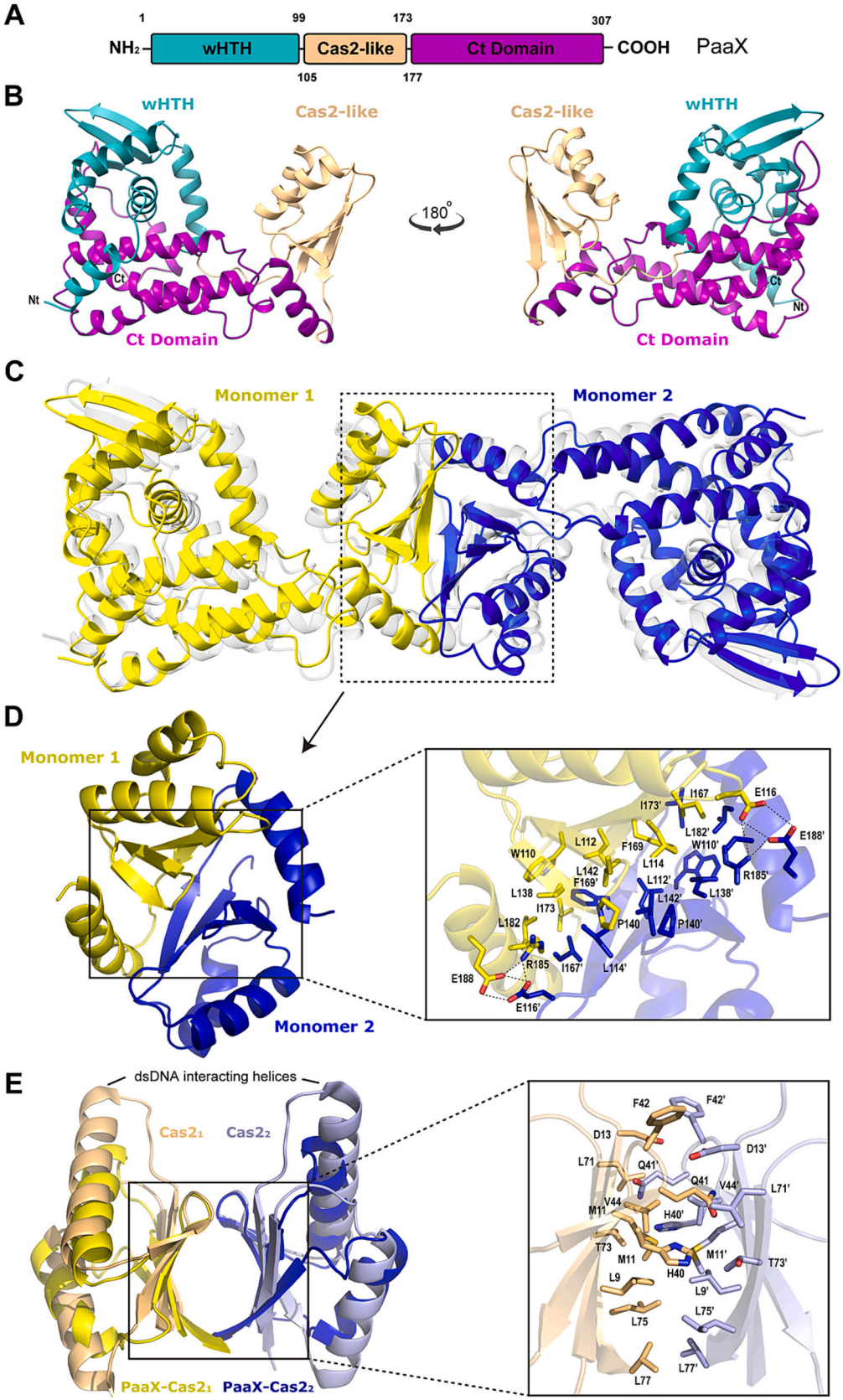

Fig. 2. Crystal structure of PaaX. (A) Domain organization of PaaX. (B) Cartoon model of the PaaX monomer. (C) A dimeric model of PaaX superimposing the crystal structure (yellow/blue) and the AlphaFold model (grey). (D) Dimer interface (left panel) displaying key residues for dimer interactions (right panel). (E) Superimposition of the Cas2-like dimer from PaaX and Cas2 dimer (left panel) and key interactions of the Cas2 dimer (right panel).

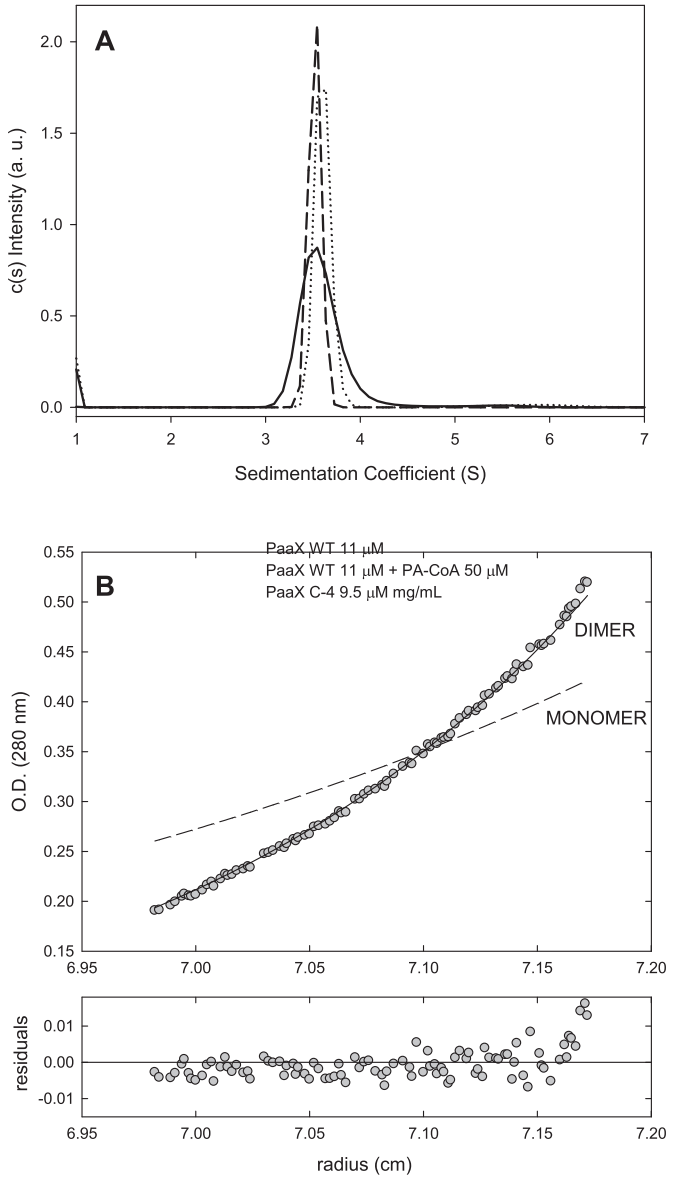

Fig. 3. Hydrodynamic experiments. (A) Distribution of sedimentation coefficients c(s) of 11 μM PaaX WT in the presence (solid line) or in the absence of 50 mM PA-CoA (dotted line), and of 9.5 μM PaaX-C4 in the absence of PA-CoA (dashed line). Experiments were carried out at 48000 rpm and monitored by absorbance at 280 nm. (B) Gradient at sedimentation equilibrium of 8.7 μM PaaX (grey circles) at 9500 rpm, monitored at 280 nm. The solid and dashed lines represent the theoretical gradients of a dimer and a monomer respectively. The residuals plot below corresponds to the residuals of the fit of the experimental data to the dimer model.

arrangements between the wHTH, the Cas2-like domain and the first helix of the C-terminal domain (residues 173–190), resulting in a rmsd ranging from 0.46 to 1.11 Å for Cα superimposition between monomers and reflecting some relative mobility between them (Fig. S1; Supplementary simulation movie M1). This intrinsic flexibility of the PaaX monomer might be critical for in vivo recognition of PA-CoA and its effect on DNA binding (see below).

Sequence and structural analysis from PaaX were performed by combining HHpred (a server for protein remote homology detection and 3D structure prediction [41]), DALI (a network service for comparing 3D protein structures [42]) and Foldseek (a searching algorithm to align either the experimentally elucidated or predicted 3D structure of a query protein against a database based on tertiary amino acid interactions [43]). Results provided by HHpred and confirmed by Foldseek revealed that the highest overall sequence identity encompassing the three domains (18–26 %) occurred with two deposited but non-published structures putatively functioning as transcriptional regulators (PDB code: 3L09 from *Jannaschia* sp.; PDB code: 3KFW from *Mycobacterium tuberculosis*) (Fig. S2) [3,4]. A deeper analysis excluding these two proteins showed PaaX sequence homologs grouped either by the N-terminal wHTH domain or by the βαβαβ-fold domain, while not finding any homolog exclusively through the C-terminal domain. Thus, sequence homologs to the wHTH domain of PaaX were found in many transcription regulator proteins, such as RosR from *Halobacterium salinarum* (PDB code 6QFD), the closest one with a rmsd of 2.3 Å for 77 Cα atoms superimposition. Remarkably, as stated above, the sequence of the PaaX βαβαβ domain was detected as being homologous to the Cas2 endoribonuclease from *E. faecalis* (PDB code 5XVN, rmsd of 2.5 Å for 65 Cα atoms superimposition), which is an essential protein for the bacterial defence system mediated by CRISPR [53–55]. These results were confirmed by the DALI analysis, which yielded very similar results (data not shown).

3.2.2. Dimerization of PaaX

The consensus sequence of the *Pa*, *Px* and *Pz* promoters that are recognised by PaaX involves two palindromic 6-bp sequences (5′-TGATTC[…]GAATCA-3′) separated by a 26–28-bp stretch [6,8], leading to a mean theoretical distance between them of 118 Å (assuming a standard B-DNA conformation) and suggesting a dimeric configuration for the protein. To ascertain the oligomeric state of PaaX in solution, the purified protein was subjected to sedimentation velocity at different protein concentrations. The analysis of data as a distribution of sedimentation coefficients showed a single major species (>95 %) with an *s* value of 3.6 S and a standard *s* value ($s_{20,w}$) of 4.0 S at around 10 μM protein, both for PaaX and PaaX-C4 (Fig. 3A). On the other hand, a sedimentation equilibrium assay (Fig. 3B) was carried out to evaluate the average size of the protein. The sedimentation equilibrium gradient of PaaX at 8.7 μM fitted best with a single species of about 75,000 ± 2000 Da. This is compatible with PaaX being a dimer in solution, taking into account that the sequence-derived molecular mass of the histidine-tagged PaaX monomer is 36,636 Da.

As we observed an unequivocal tendency of PaaX to dimerize in solution either in the presence or in the absence of its PA-CoA ligand, we analyzed the possible dimeric arrangements observed within the crystal unit cell. In this vein, two different dimer possibilities were envisaged, either by interaction through the β-hairpins contained in their wHTH domains or through interaction between their Cas2-like domains. To shed light on the likeliness of each configuration, we performed ab initio complex formation prediction by using AI algorithms as implemented in AlphaFold2 [28]. All solutions given by AlphaFold2 converged to a single dimer arrangement mode through the Cas2-like domains, which matched one of the mentioned dimers observed in the crystal packing (Fig. 2C–D). In this regard, it should be mentioned that Cas2 dimerization is needed to produce the functional Cas1-Cas2 CRISPR integrase complex [53] and, in fact, the superimposition of *E. faecalis* Cas2 (PDB code, 5XVN) and the PaaX Cas2-like domain dimeric structures reveals a similar interaction mode involving 4 β-sheets from each monomer (Fig. 2E) reinforcing the plausibility of the proposed dimer. Nevertheless, the dimerization interface in PaaX is built mainly by hydrophobic forces, in which 10 residues from each monomer contribute to create a strong hydrophobic core that is flanked by an external ionic triad conformed by residues Arg-185 and Glu-188 from one chain and Glu-116 from the other one (Fig. 2D), whereas while Cas2 displayed both hydrophobic and polar contacts (Fig. 2E).

3.2.3. Prediction of DNA binding mode

As the structural recognition of DNA by *E. faecalis* Cas2 has already been described [54,55]), we explored DNA binding to the PaaX dimer by using the 3D superimposition of PaaX Cas2-like domains onto the Cas2:

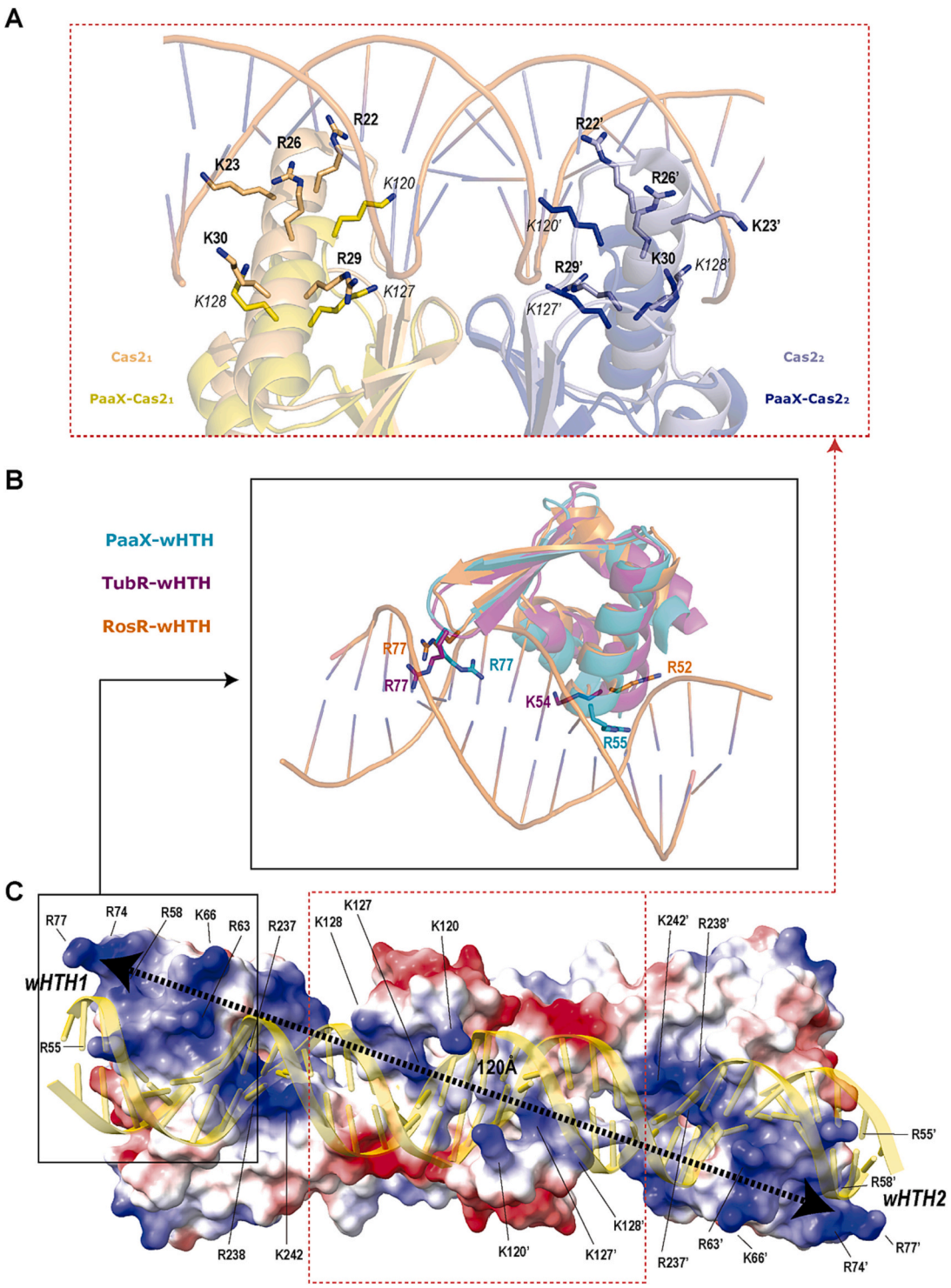

Fig. 4. Putative dsDNA binding mode of PaaX. (A) PaaX Cas2-like dimer (yellow-dark blue) superimposed onto the Cas2 dimer in complex with dsDNA (PDB code, 5XVN, bright orange-slate blue) displaying the key residues involved in DNA binding. (B) PaaX wHTH domain (in blue) superimposed onto the wHTH binding domain from RosR (PDB code, 6QFD, in orange) and TubR (PDB code, 4ASO, in purple) proteins. Conserved residues involved in DNA binding in this region are represented as capped sticks and labelled. (C) Electrostatic potential surface of crystal PaaX dimer and its dsDNA binding mode. Blue and red regions represent positive and negative potentials, respectively, while white regions represent non-polar surfaces.

DNA complex (PDB code 5XVN) (Fig. 4). Interestingly, the Cas2 α helix involved in DNA binding by insertion into its major groove is also conserved in Paax (residues 119–133). This interaction is amplified in the dimer where the α helix from each monomer contributes to the DNA binding (Fig. 4). The distribution of positively charged residues along this helix was also partially conserved in PaaX with Lys-127 and Lys-128 oriented to DNA similarly to residues Arg-29 and Arg-30 in Cas2 (Fig. 4A). Furthermore, the wHTH domain provides a second DNA-binding region, as seen for residues Arg-55 and Arg-77 which can be aligned to DNA interacting residues Arg-52 and Arg-77 of RosR protein (PDB code 6QFR) or Lys-54 and Arg-77 of TubR protein (PDB code 4ASO) (Fig. 4B). Collectively, the analysis of the electrostatic potential

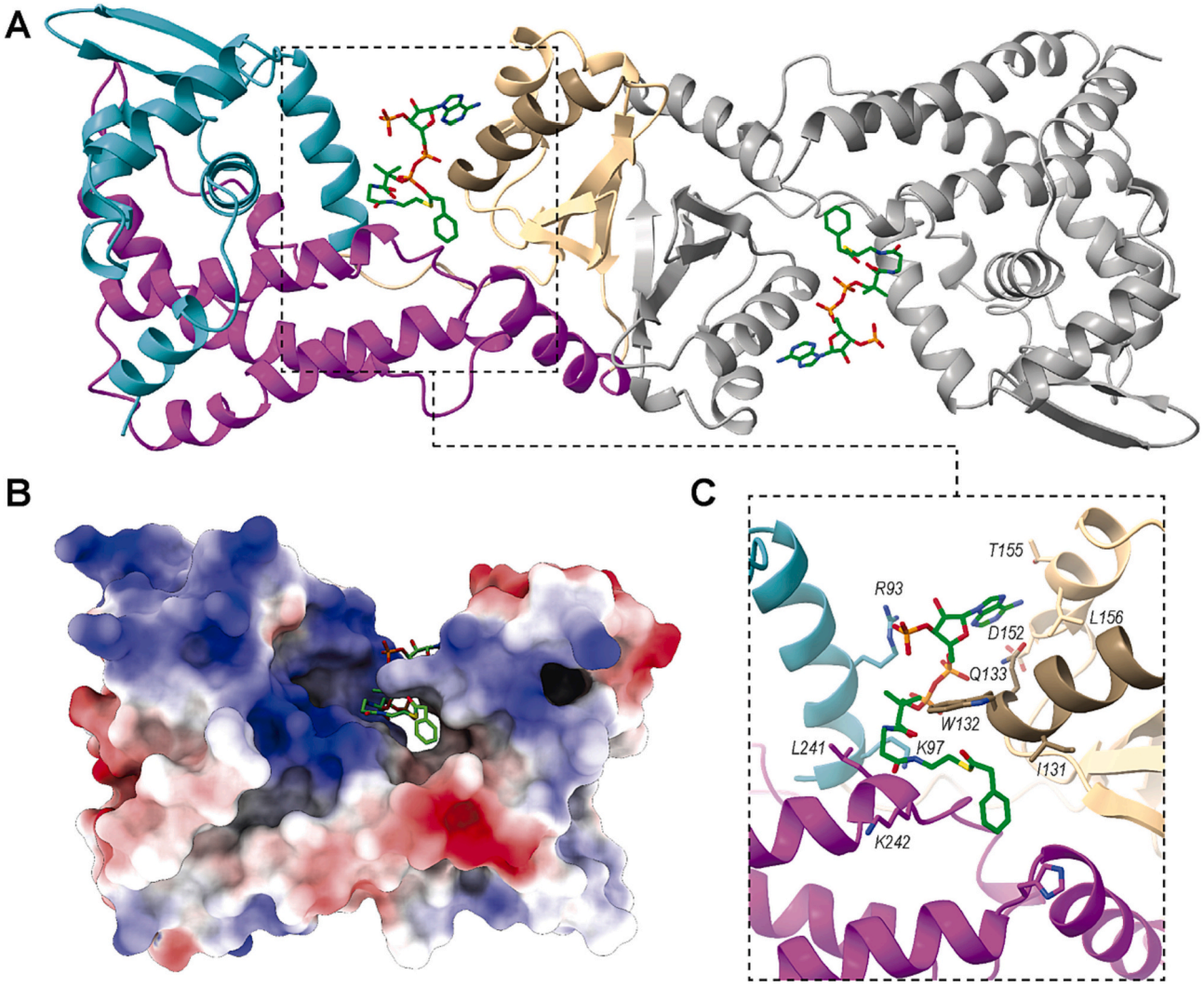

Fig. 5. In silico docking of PA-CoA on PaaX-C4 crystal structure. (A) Docking results on PaaX dimer. Dark cyan, light brown and purple ribbons indicate the wHTH, Cas 2-like and C-terminal domains, respectively. The putative DNA-binding helix comprising residues 119–133 and belonging to the Cas 2-like domain is depicted in dark brown. Docked PA-CoA molecules are shown in stick representation. (B) Electrostatic potential representation of the PA-CoA:PaaX docking complex model. Blue and red regions represent positive and negative potentials, respectively, while white regions represent non-polar surfaces. (C) detailed view of the PA-CoA binding site and the residues contacting the ligand.

surface in the dimer revealed a 120 Å long basic patch spanning from wHTH1 to wHTH2, in which the DNA molecule can be accommodated (Fig. 4C).

Inspection of the PaaX dimer structure also allowed us to identify the positions of the mutated cysteines, with the exception of Cys-312 as this C-terminal region could not be solved as mentioned above (Fig. S3). Those mutations that rendered variants less able to be released from their promoter sequences upon induction by PA-CoA (i.e. Cys-168 and Cys-189, Table 1) are located in the Cas 2-like domain near the dimerization interface and far from the DNA-binding sequences. Besides any possible differences in gene expression levels, since both wild-type and C4 mutant bound to DNA with similar efficiency in the absence of PA-CoA (Fig. 1), it is likely that these mutations did not affect the competent form to bind DNA but rather somehow hamper the conformational changes caused by the binding of the inductor (see below). Of note, none of the homologous sequences identified by HHPred, DALI or Foldseek contained any cysteines in the same positions.

3.2.4. *Model of phenylacetyl-coenzyme A binding*

PA-CoA has been previously shown to prevent PaaX binding to DNA in EMSA assays, therefore demonstrating that it is the true inducer of the PAA degradation pathway [6,11]. Taking advantage of the elucidated 3D structure of the protein we attempted the modeling of PA-CoA binding by in silico docking procedures (Section 2.11). The lowest energy conformations generated by SwissDock [46] displayed the PA-CoA molecule bound, in a zig-zag configuration, in the positively charged deep crevice between the three domains of the protein (Fig. 5A–B). The ligand would be stabilized by several residues as detailed in Fig. 5C: *i*) the adenosine moiety could establish hydrogen bonds with the side chains of Arg-93, Gln-133, Asp-152 and Thr-155, as well as a van der Waals contact with Leu-156; *ii*) the charged phosphate groups can be stabilized through ionic interactions with Arg-93 and Lys-97; *iii*) the pantoic acid moiety is hydrogen-bound to the side-chain of Lys-97 and the peptidic backbone of Leu-241; *iv*) the β-Ala moiety contacts Trp-132 and Lys-242 through van der Waals forces, and, finally, *v*) the phenylacetyl moiety is involved in two simultaneous T-shaped π-π interactions with Trp-132 and His-191 at the bottom of the crevice and after an abrupt kink in the ligand molecule produced by the protruding side-chain of Trp-132 (Fig. 5B–C). Two remarkable features arise from this model. First, the predicted binding site of PA-CoA overlaps with the 119–133 DNA-binding α-helix (Fig. 5A, dark brown-coloured helix), so that the entry of the ligand into the crevice would very likely interfere

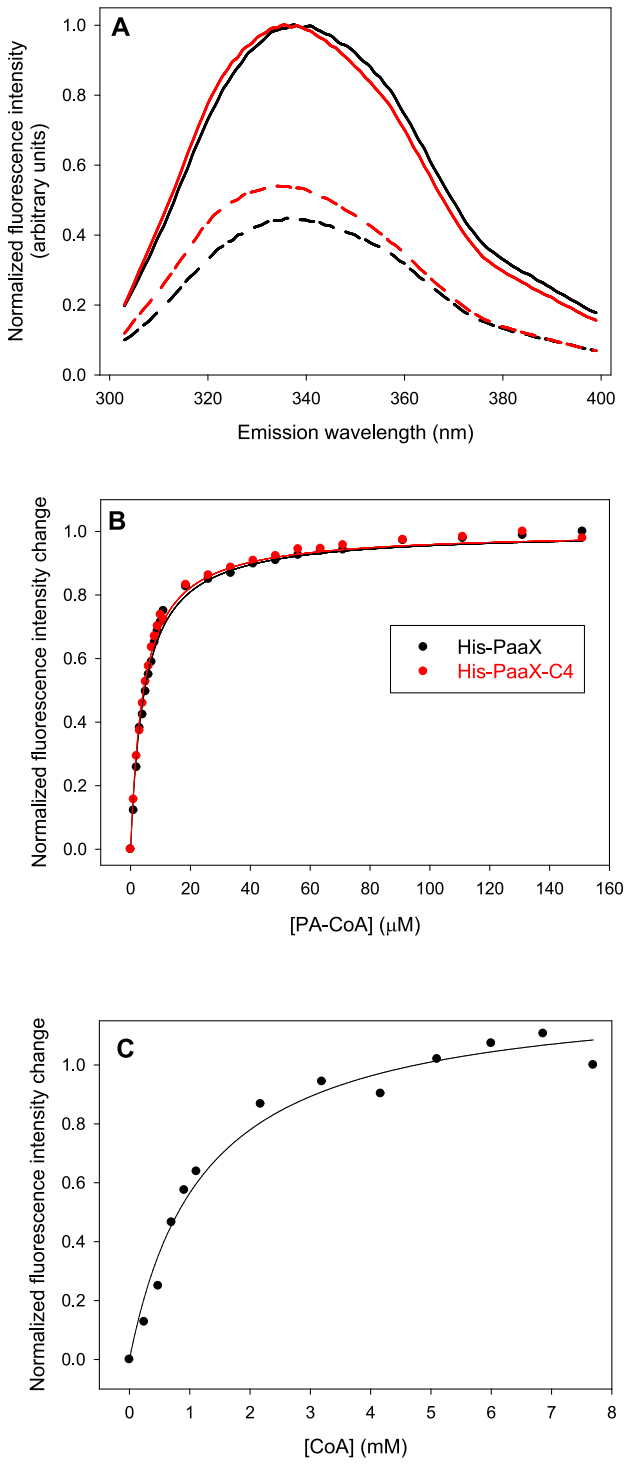

Fig. 6. Binding of PA-CoA to PaaX. (A) Tryptophan fluorescence emission spectra of PaaX (black) or PaaX-C4 (red) in the absence (solid line) or in the presence (dashed line) of 19.6 μM PA-CoA. Data are normalized to the maximum intensity of each protein without ligand. (B) Titrations of PaaX (black circles) and PaaX-C4 (red circles) with PA-CoA monitored by decrease in intrinsic fluorescence emission at 340 nm. Changes in fluorescence intensity were normalized for clarity of presentation. Lines represent the best fit for each titration of a model of one type of binding sites with ligand concentrations equivalent to the target (Eq. (5)). (C) Titration of PaaX CoA. Lines represent the best fit for each titration of a model of one type of binding sites with ligand concentrations in excess of the target (Eq. (6)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Biophysical data of PaaX-ligand interaction.

Protein	K_d (μM)[a] (PA-CoA)	K_d (μM)[b] (CoA)	Δt_m (°C) (20 μM PA-CoA)	Δt_m (°C) (200 μM PA-CoA)	Δt_m (°C) (1.2 mM CoA)
PaaX	4.7 ± 0.2	1200 ± 200	+1.2 ± 0.1	+3.3 ± 0.1	+1.2 ± 0.2
PaaX-C4	4.3 ± 0.1	ND[c]	+0.9 ± 0.1	+3.9 ± 0.1	ND[c]

[a] Dissociation constant calculated according to Eq. (5).
[b] Dissociation constant calculated according to Eq. (6).
[c] ND: Not determined.

with the recognition of the DNA by the protein. On the other hand, while the CoA moiety itself establishes several interactions with the protein, the specificity of binding is provided by a highly stabilizing T-shaped π-π stacking network that includes the phenylacetyl group of the ligand and the Trp-132 and His-191 aromatic residues. It should be mentioned that T-shaped π-π contacts have been described as being the most prevalent among all possible aromatic interactions established between proteins and their ligands [56]. It is worth mentioning, that the structural homologs of *E. coli* Paax (the putative transcriptional regulator from *Jannaschia* sp. CCS1 - PDB code: 3L09 - and the uncharacterized protein Rv0674 from *Mycobacterium tuberculosis*, - PDB code: 3KFW) also present a related cavity with a basic electrostatic potential at that region. Blind docking of PA-CoA ligand onto these structures also places the ligand in the same crevice as in Paax (Fig. S4).

3.3. Experimental validation of the PaaX structural model

The availability of the 3D structure of the PaaX repressor constitutes a solid basis to provide a molecular explanation of several experimental features of the protein. It should be remarked here that the N-terminal histidine tag employed for the purification of the protein, although too flexible to be solved in the crystal structure, should be in any case unequivocally placed in a solvent-exposed region and completely separated from the dimerization interface, the DNA-binding motifs or the presumed PA-CoA binding sites (Fig. 2C), so any interference of this sequence in the biophysical experiments described below is most likely to be minimal, if any at all.

3.3.1. Binding of PA-CoA

We first studied the biophysical characteristics of the binding of PA-CoA to PaaX using intrinsic fluorescence spectroscopy. The protein was excited at 295 nm, a wavelength at which PA-CoA does not absorb, to monitor any conformational changes exclusively in the protein. The emission spectrum of PaaX has a maximum at 337 nm (Fig. 6A), indicating the partial burial of the emitting tryptophan residues. Addition of increasing amounts of PA-CoA induced a remarkable decrease in fluorescence intensity together with a slight shift to shorter wavelengths (336 nm) (Fig. 6A). A similar result was observed with PaaX-C4 (Fig. 6A). This suggests that aromatic residues are located in or near the PA-CoA binding site and, in fact, according to the docking model (Fig. 5C), the indole group of Trp-132 would be placed in close proximity to the terminal phosphate substituent of the tetrahydrofuran group in the ligand, which in turn might act as an efficient quencher of its fluorescence [57]. Highly similar titration curves of PaaX or PaaX-C4 with PA-CoA were obtained by monitoring the decrease in tryptophan fluorescence emission (Fig. 6B), both showing a saturation profile that was analyzed using a single-site binding model (Eq. (5)) which resulted in dissociation constants (K_d) of 4.7 ± 0.2 and 4.3 ± 0.1 μM, respectively (Table 3), and confirming that the cysteine substitutions in PaaX-C4 do not strongly affect PA-CoA binding.

To assess the degree of specificity and strength provided by the phenylacetyl moiety of PA-CoA, binding of free coenzyme A (CoA) was

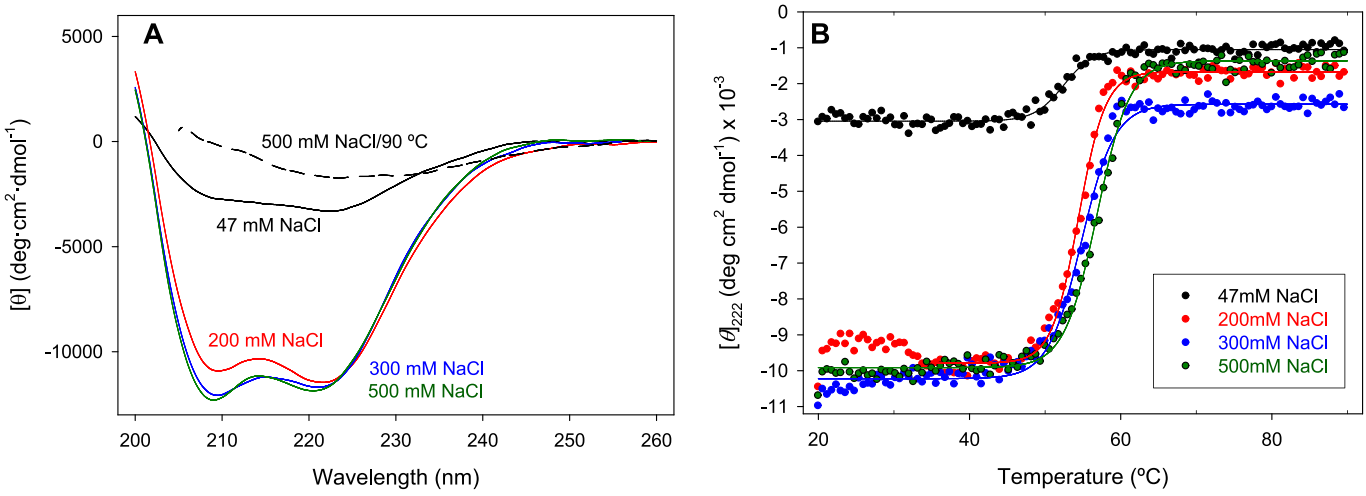

Fig. 7. Dependence of PaaX thermal stability on ionic strength. (A) Far-UV CD spectra of PaaX at 20 °C and at different NaCl concentrations and at 90 °C at 500 mM NaCl. All the samples were prepared in 50 mM phosphate at pH 8 with 20 % glycerol and the stated NaCl concentrations. (B) Temperature scans monitored by CD signal at 222 nm at different NaCl concentrations. Sigmoidal functions were fitted to obtain the midpoints of the transitions shown in Table 3.

also monitored by intrinsic fluorescence spectroscopy (Fig. 6C). Similarly to PA-CoA, titration of PaaX with CoA also produced a decrease in the tryptophan fluorescence intensity, albeit with a ~250-fold lower affinity (K_d = 1.2 ± 0.2 mM). This is in accordance with the previous report where the addition of high concentrations (800 μM) of CoA was unable to completely release PaaX from DNA in the EMSA assays [18]. Moreover, the titration of PaaX with CoA was repeated in the presence of 10 mM phenylacetic acid (PAA), although no changes in the binding curve were observed (data not shown), indicating that both moieties must be linked in a single molecule for efficient binding.

Sedimentation velocity experiments were performed to test the effect of PA-CoA in PaaX oligomerization. The addition of 50 μM PA-CoA, a sufficient amount to saturate the protein (Fig. 6B) did not alter the sedimentation coefficient distribution of the protein (Fig. 3A), suggesting that the ligand does not have an effect on the oligomerization state of the protein even though its predicted binding site contains elements of the Cas2-like dimerization domain (Fig. 5A).

Finally, the interaction of PaaX with PA-CoA was also assessed by far-UV circular dichroism (CD). Fig. S5 shows a wavelength spectrum typical of a protein with high content in α-helix, with marked minima at 208 and 222 nm [58]. Spectral deconvolution by the CDNN utilities [59] yielded a 41 % α-helical content, in reasonable accordance with the actual content derived from 3D structure of the protein (51 %) (Fig. 2A), considering the accuracy of such deconvolution procedures. The ellipticity signal decreased in the presence of PA-CoA (Fig. S5), with an estimated decrease in the amount of α-helix content (41 % to 34 %) concomitant with an increase in the amount of random coil (44 % to 49 %) was estimated, suggesting that a part of the protein becomes more flexible upon PA-CoA binding, as also speculated above (Fig. S1).

3.3.2. *Thermal stability and effects of ionic strength*

The thermal stability of PaaX was also assessed by CD (Fig. 7). The wavelength spectrum of PaaX at 20 °C registered at moderate to high ionic strength (300–500 mM NaCl) displays the above-mentioned minima at 208 and 222 nm ascribed to α-helical structure (Fig. 7A). A CD-monitored temperature scan performed in these conditions by following the ellipticity signal at 222 nm displayed a single co-operative transition (Fig. 7B) leading, at high temperatures, to a wavelength spectrum indicative of mixed α and β structures [58] (Fig. 7A). Moreover, the signal was severely reduced, as the protein visibly aggregated in the cuvette. Besides, the midpoint of the thermal unfolding curve (t_m) depended on the temperature scanning rate (data not shown). These observations indicate that PaaX unfolds irreversibly at high temperatures, probably forming intermolecular, aggregating folding intermediates that accumulate rapidly upon denaturation and, consequently, a thermodynamic analysis of the curves was not possible. Instead, midpoints of the transitions (t_m) were obtained by adjusting the curves to a simple sigmoidal function with no physical meaning and were used as rough indicators of protein stability (Table 4).

Table 4
Effect of ionic strength on the thermal stability of PaaX. Experiments were carried out in 20 mM sodium phosphate, pH 8 buffer.

[NaCl] (mM)	t_m (°C)
47	51.6 ± 0.2
200	54.9 ± 0.1
300	55.4 ± 0.1
500	57.2 ± 0.1

The secondary structure and thermal stability of the protein where highly dependent on NaCl concentration. Thermal denaturation midpoints increased with increasing ionic strength (Table 4), and the wavelength spectrum showed some hints of secondary structure loss at 200 mM NaCl, while low ionic strength conditions (47 mM NaCl) caused the CD signal to be dramatically lost concomitantly with a decrease of the absorbance of the sample, which is again indicative of a loss of soluble protein due to aggregation (Fig. 7A). This also translated to a significantly decreased midpoint temperature, almost 6 degrees lower than the 500 mM NaCl case (Fig. 7B, Table 4). A possible explanation for this behaviour might take into account the existence of extensive clusters of positive charges both in the DNA-binding motifs of the protein and in one side of the presumed PA-CoA-binding crevice, together with smaller clusters of negatively charged acidic residues in the Cas2-like domain (Fig. 4C, Fig. S4A). Electrostatic repulsions within these regions could prevent the protein from properly folding in low salt conditions, leading to exposed hydrophobic patches prone to intermolecular association, whereas addition of a high ionic strength (or its polyanionic DNA substrate) that screens such repulsions, might lead to overall protein stabilization.

We also compared the thermal stability of wild-type PaaX with the C168A/C264A/C312A, C189A/C264A/C312A and C4 variants (Table 1). Thermal transitions were similar to the wild-type protein, with a single cooperative transition and aggregation at high temperatures (data not shown), but with only slightly lower transition midpoints, confirming that the mutations did not largely affect the overall

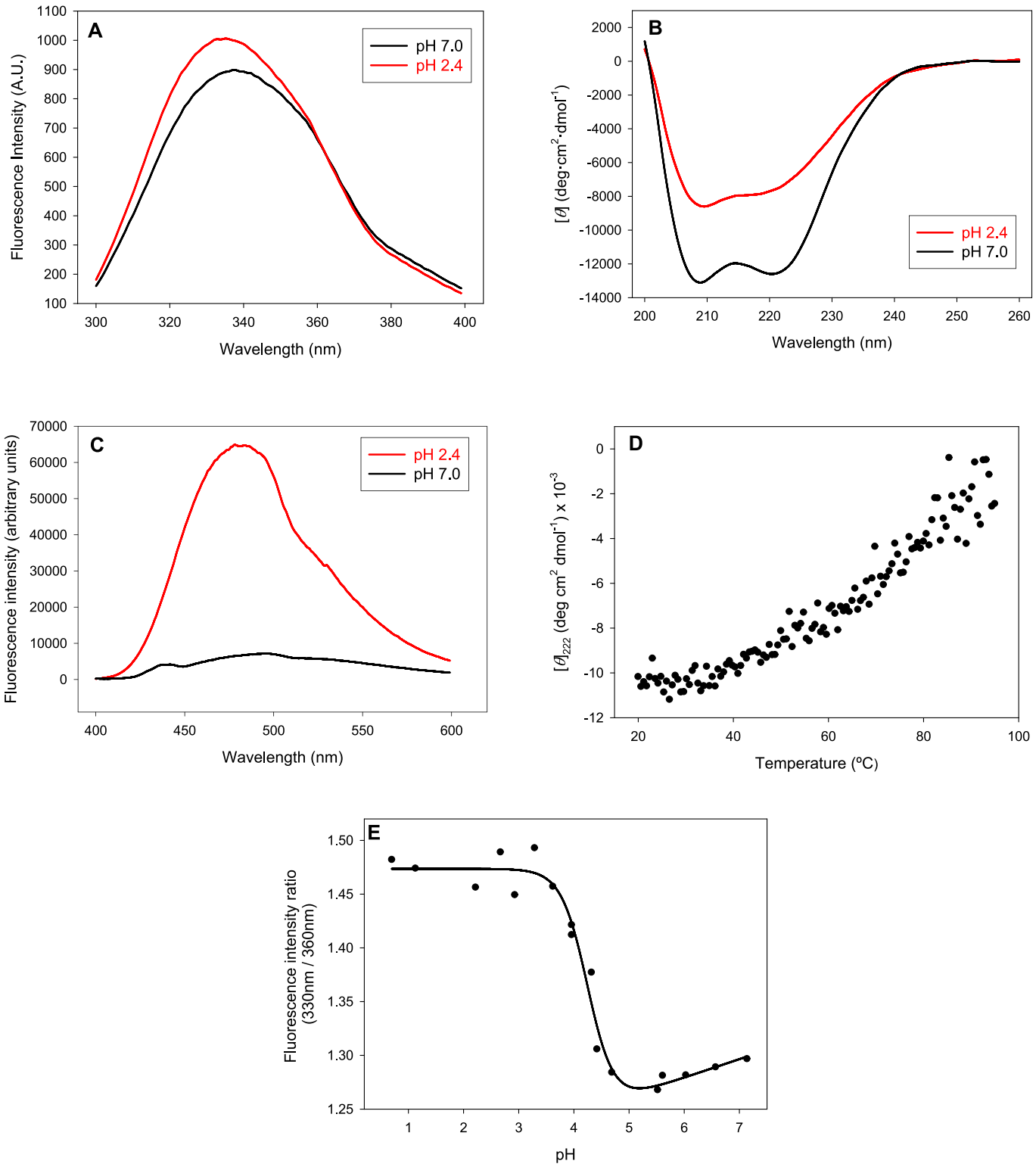

Fig. 8. Stability of PaaX at acid pH. PaaX was titrated with different buffers from pH 7.0 to pH 0.5 as described in Materials an Methods. (A) and (B), intrinsic fluorescence emission and far-UV CD spectra, respectively, of PaaX at pH 7.0 and pH 2.4. (C) ANS emission spectra of PaaX at pH 7 and pH 2.4 (dotted line). (D), temperature scan of PaaX at pH 2.4. (E) Titration of PaaX with different pHs. The solid line represents the best fit to a model to determine pKa and the number of protons involved in the transition (Eq. (7)).

protein structure.

On the other hand, addition of micromolar concentrations of PA-CoA enhanced thermostability for both to the wild-type and C4 PaaX variant to a similar degree, whereas free CoA produced only a modest change in t_m and in the presence of a much higher concentration (Table 3). This reinforces our previous suggestion that the complete PA-CoA structure is necessary for an efficient binding (Fig. 6).

3.3.3. *Stability at acid pH*

Upon lowering the pH of the solution, the spectroscopic

characteristics of PaaX changed substantially. At pH 2.0 the tryptophan fluorescence emission spectrum of PaaX displayed a moderate increase in intensity and a blue-shift of the emission maximum to 335 nm (Fig. 8A) suggesting a deeper burial of tryptophan residues. However, changes in the CD wavelength spectrum were more remarkable, showing a loss of about 50 % of the ellipticity and a change in shape (Fig. 8B). Moreover, the 8-anilinonaphthalene-1-sulfonic acid (ANS) probe, when added at pH 2.4, induced a remarkable increase in its fluorescence intensity that was otherwise negligible at pH 7.0 (Fig. 8C), suggesting that at low pH the probe is bound to solvent-exposed hydrophobic patches of the protein [60]. Furthermore, a thermal scan of PaaX at low pH monitored by far-UV CD lacked any clear cooperative transition (Fig. 8D) adding evidence that the specific packing of the hydrophobic core is disrupted and that the burial of tryptophan residues takes place within non-specific, residual hydrophobic regions. All these results suggest that PaaX is transformed at acidic pH into a partially folded state of the so-called "molten-globule" type, with a moderate content of remaining secondary structure, a loose tertiary packing and extensive solvent-exposed hydrophobic patches [61]. A pH titration of the protein was monitored by tryptophan fluorescence emission (Fig. 8E). A single cooperative transition was obtained that was fitted to a two-state model of pH-dependent unfolding (Eq. (7)), yielding a titration pKa of 4.3 $\pm$ 0.1 and the uptake of approximately 2 protons in the transition (1.8 $\pm$ 0.4). This suggests that the protonation state of two Asp/Glu residues governs the transition between the native structure of PaaX at neutral pH and the intermediate accumulated at pH 2.4. It can be speculated that the above-mentioned ionic triad involving Arg-185 and Glu-188 (monomer A) and Glu-116 (monomer B), that "clamp" the hydrophobic dimerization core in the Cas2-like domain (Fig. 2D) are candidates to be essential in maintaining the dimer conformation, so protonation in these two places at low pH could involve dimer dissociation and exposure of the mentioned hydrophobic regions.

4. Conclusions

The crystal structure of a variant of the PaaX repressor from *E. coli* W lacking its four cysteines reveals a novel fold in the family of transcription regulators. PaaX is revealed as a dimer where each monomer contains three domains, namely an N-terminal wHTH domain, a middle domain with homology to *E. faecalis* Cas2 protein and a C-terminal domain with a previously unreported structure. The protein dimerizes through the Cas2-like domain, creating a positively charged, continuous surface amenable to interact with its cognate promoter sequences [8]. There is a deep, cavity between the wHTH and Cas2-like domains which is amenable to bind a molecule of its activating ligand, PA-CoA per monomer. Presumably, the hinge present in between the wHTH and Cas2-like domains allows some flexibility to reposition these domains, as suggested by the different packing of the two domains in the three molecules of the crystal asymmetric unit, allowing them to accommodate the ligand. It is therefore evident that any spatial rearrangement of these two domains upon PA-CoA binding would dramatically affect its DNA binding capacity, finally causing the release of the protein. On the other hand, we have carried out the first biophysical characterization of this type of fold, and found that the stability of the protein is strongly dependent on the pH and ionic strength, and that is prone to undertake conformational changes leading to partially folded intermediates of the "molten globule" kind. Whether this marginal stability is specific of PaaX or can be extended to other transcriptional regulators with this similar fold (e.g. the *Jannaschia* sp. and *M. tuberculosis* proteins described above) deserves further investigation.

In summary, the structure and biophysical assays described here may provide novel and revealing clues on the mechanism by which the important regulator PaaX represses expression of the phenylacetic acid and other aromatic compound catabolic pathways in *Escherichia coli* and many other bacteria.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijbiomac.2023.127935.

CRediT authorship contribution statement

Víctor M. Hernández-Rocamora: Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **Rafael Molina:** Investigation, Formal analysis, Writing – review & editing. **Alejandra Alba:** Investigation. **César Carrasco-López:** Investigation. **Alzoray Rojas-Altuve:** Investigation. **Santosh Panjikar:** Investigation. **Ana Medina:** Investigation. **Isabel Usón:** Investigation. **Carlos Alfonso:** Investigation, Formal analysis. **Beatriz Galán:** Formal analysis. **Germán Rivas:** Formal analysis. **Juan A. Hermoso:** Formal analysis, Funding acquisition, Project administration, Writing – review & editing. **Jesús M. Sanz:** Supervision, Conceptualization, Formal analysis, Funding acquisition, Project administration, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to thank Dr. José Luis García for kindly providing the pX2BS plasmid and the plasmids and reporter strain needed to study the function of PaaX in vivo, to Cristina Fernández for her support and counseling on performing the Miller experiments and to Dr. Javier Gómez for assistance in the analysis of the PA-CoA binding data. We are grateful to the staff of the Protein Engineering Group at Instituto de Biología Molecular y Celular, for their support. This research was funded by the following sources: Grants PID2019-105126RB-I00, PID2022-139209OB-C21 (MCIN/AEI/10.13039/501100011033/and ERDF A way of making Europe), TED2021-129747B-C22 (AEI/10.13039/501100011033/NextGenerationEU/PRTR) and CIBER-Consorcio Centro de Investigación Biomédica en Red (CIBERES, Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación, Spain) to JMS; grants PID2020-115331GB-100 funded by MCIN/AEI/10.13039/501100011033 and CRSII5_198737/1 (Swiss National Science Foundation) to JAH; grant PID2021-128751NB-I00 (MICINN/AEI/FEDER/UE) to IU, and grant RYC2021-030916-I by the Spanish Agencia Estatal de Investigación to RM. VMH-R was supported by a FPU PhD fellowship from Spanish *Ministerio de Educación y Ciencia*.

References

[1] T.D.H. Bugg, M. Ahmad, E.M. Hardiman, R. Rahmanpour, Pathways for degradation of lignin in bacteria and fungi, Nat. Prod. Rep. 28 (2011) 1883–1896, https://doi.org/10.1039/c1np00042j.
[2] G. Fuchs, M. Boll, J. Heider, Microbial degradation of aromatic compounds - from one strategy to four, Nat. Rev. Microbiol. 9 (2011) 803–816, https://doi.org/10.1038/nrmicro2652.
[3] M. Jiao, W. He, Z. Ouyang, Q. Shi, Y. Wen, Progress in structural and functional study of the bacterial phenylacetic acid catabolic pathway, its role in pathogenicity and antibiotic resistance, Front. Microbiol. 13 (2022), 964019, https://doi.org/10.3389/fmicb.2022.964019.
[4] A.M. Grishin, M. Cygler, Structural organization of enzymes of the phenylacetate catabolic hybrid pathway, Biology (Basel) 4 (2015) 424–442, https://doi.org/10.3390/biology4020424.
[5] J.M. Luengo, J.L. García, E.R. Olivera, The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications, Mol. Microbiol. 39 (2001) 1434–1442, https://doi.org/10.1046/j.1365-2958.2001.02344.x.
[6] A. Ferrández, B. Miñambres, B. García, E.R. Olivera, J.M. Luengo, J.L. García, E. Díaz, Catabolism of phenylacetic acid in *Escherichia coli*: characterization of a new aerobic hybrid pathway, J. Biol. Chem. 273 (1998) 25974–25986, https://doi.org/10.1074/jbc.273.40.25974.
[7] R. Teufel, V. Mascaraque, W. Ismail, M. Voss, J. Perera, W. Eisenreich, W. Haehnel, G. Fuchs, Bacterial phenylalanine and phenylacetate catabolic pathway revealed, Proc. Natl. Acad. Sci. U. S. A. 107 (2010) 14390–14395, https://doi.org/10.1073/pnas.1005399107.
[8] H.S. Kim, T.S. Kang, J.S. Hyun, H.S. Kang, Regulation of penicillin G acylase gene expression in *Escherichia coli* by repressor PaaX and the cAMP-cAMP receptor protein complex, J. Biol. Chem. 279 (2004) 33253–33262, https://doi.org/10.1074/jbc.M404348200.

[9] B. Galán, J.L. García, M.A. Prieto, The PaaX repressor, a link between penicillin G acylase and the phenylacetyl-coenzyme A catabolon of *Escherichia coli* W, J. Bacteriol. 186 (2004) 2215–2220, https://doi.org/10.1128/JB.186.7.2215-2220.2004.
[10] T. Del Peso-Santos, D. Bartolomé-Martín, C. Fernández, S. Alonso, J.L. García, E. Díaz, V. Shingler, J. Perera, Coregulation by phenylacetyl-coenzyme A-responsive PaaX integrates control of the upper and lower pathways for catabolism of styrene by *Pseudomonas* sp. strain Y2, J. Bacteriol. 188 (2006) 4812–4821, https://doi.org/10.1128/JB.00176-06.
[11] C. Fernández, E. Díaz, J.L. García, Insights on the regulation of the phenylacetate degradation pathway from *Escherichia coli*, Environ. Microbiol. Rep. 6 (2014) 239–250, https://doi.org/10.1111/1758-2229.12117.
[12] A. Ferrández, J.L. García, E. Díaz, Transcriptional regulation of the divergent paa catabolic operons for phenylacetic acid degradation in *Escherichia coli*, J. Biol. Chem. 275 (2000) 12214–12222, https://doi.org/10.1074/jbc.275.16.12214.
[13] F. Ripoll, S. Pasek, C. Schenowitz, C. Dossat, V. Barbe, M. Rottman, E. Macheras, B. Heym, J.L. Hermann, M. Daffé, R. Brosch, J.L. Risler, J.L. Gaillard, Non mycobacterial virulence genes in the genome of the emerging pathogen *Mycobacterium abscessus*, PloS One 4 (2009) 5660, https://doi.org/10.1371/journal.pone.0005660.
[14] R. Alkasir, Y. Ma, F. Liu, J. Li, N. Lv, Y. Xue, Y. Hu, B. Zhu, Characterization and transcriptome analysis of *Acinetobacter baumannii* persister cells, Microb. Drug Resist. 24 (2018) 1466–1474, https://doi.org/10.1089/mdr.2017.0341.
[15] S. Kashyap, P. Sharma, N. Capalash, Potential genes associated with survival of *Acinetobacter baumannii* under ciprofloxacin stress, Microbes Infect. 23 (2021), https://doi.org/10.1016/j.micinf.2021.104844.
[16] T.A. McDonald, N.T. Holland, C. Skibola, P. Duramad, M.T. Smith, Hypothesis: phenol and hydroquinone derived mainly from diet and gastrointestinal flora activity are causal factors in leukemia, Leukemia 15 (2001) 10–20, https://doi.org/10.1038/sj.leu.2401981.
[17] J. Sambrook, D.W. Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 2001.
[18] C. Fernandez, Advances in the Structural Characterization of the Phenylacetic Acid Catabolism in *Escherichia coli*, Universidad Complutense de Madrid, 2005 (Ph.D. Thesis).
[19] S. Fernández, V. de Lorenzo, J. Pérez-Martin, Activation of the transcriptional regulator XylR of Pseudomonas putida by release of repression between functional domains, Mol. Microbiol. 16 (1995) 205–213, https://doi.org/10.1111/j.1365-2958.1995.tb02293.x.
[20] J.H. Miller, Assay of β-galactosidase, in: Exp. Mol. Genet, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA, 1972.
[21] U.K. Laemmli, Cleavage of structural proteins during the assembly of the head of bacteriophage T4, Nature 227 (1970) 680–685.
[22] C.N. Pace, F. Vajdos, L. Fee, G. Grimsley, T. Gray, How to measure and predict the molar absorption coefficient of a protein, Protein Sci. 4 (1995) 2411–2423, https://doi.org/10.1002/pro.5560041120.
[23] J. Sedlak, R.H. Lindsay, Estimation of total, protein-bound, and nonprotein sulfhydryl groups in tissue with Ellman's reagent, Anal. Biochem. 25 (1968) 192–205, https://doi.org/10.1016/0003-2697(68)90092-4.
[24] A. Rojas-Altuve, C. Carrasco-López, V.M. Hernández-Rocamora, J.M. Sanz, J. A. Hermoso, Crystallization and preliminary X-ray diffraction studies of the transcriptional repressor PaaX, the main regulator of the phenylacetic acid degradation pathway in *Escherichia coli* W, Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67 (2011) 1278–1280, https://doi.org/10.1107/S1744309111029873.
[25] A. Vagin, A. Lebedev, MoRDa, an automatic molecular replacement pipeline, Acta Crystallogr. Sect. A Found. Adv. 71 (2015) s19, https://doi.org/10.1107/s2053273315099672.
[26] S. Panjikar, V. Parthasarathy, V.S. Lamzin, M.S. Weiss, P.A. Tucker, Auto-Rickshaw: an automated crystal structure determination platform as an efficient tool for the validation of an X-ray diffraction experiment, Acta Crystallogr. Sect. D Biol. Crystallogr. 61 (2005) 449–457, https://doi.org/10.1107/S0907444905001307.
[27] A.J. McCoy, R.W. Grosse-Kunstleve, P.D. Adams, M.D. Winn, L.C. Storoni, R. J. Read, Phaser crystallographic software, J. Appl. Cryst. 40 (2007) 658–674, https://doi.org/10.1107/S0021889807021206.
[28] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[29] C. Millán, M.D. Sammito, A.J. McCoy, A.F.Z. Nascimento, G. Petrillo, R.D. Oeffner, T. Domínguez-Gil, J.A. Hermoso, R.J. Read, I. Usón, Exploiting distant homologues for phasing through the generation of compact fragments, local fold refinement and partial solution combination, Acta Crystallogr. Sect. D Struct. Biol. 74 (2018) 290–304, https://doi.org/10.1107/S2059798318001365.
[30] A. Medina, E. Jiménez, I. Caballero, A. Castellví, J. Triviño Valls, M. Alcorlo, R. Molina, J.A. Hermoso, M.D. Sammito, R. Borges, I. Usón, Verification: model-free phasing with enhanced predicted models in ARCIMBOLDO_SHREDDER, Acta Crystallogr. Sect. D, Struct. Biol. 78 (2022) 1283–1293, https://doi.org/10.1107/S2059798322009706.
[31] M. Sammito, K. Meindl, I.M. de Ilarduya, C. Millán, C. Artola-Recolons, J. A. Hermoso, I. Usón, Structure solution with ARCIMBOLDO using fragments derived from distant homology models, FEBS J. 281 (2014) 4029–4045, https://doi.org/10.1111/febs.12897.
[32] R.J. Read, A.J. McCoy, A log-likelihood-gain intensity target for crystallographic phasing that accounts for experimental error, Acta Crystallogr. Sect. D Struct. Biol. 72 (2016) 375–387, https://doi.org/10.1107/S2059798315013236.
[33] I. Usón, G.M. Sheldrick, An introduction to experimental phasing of macromolecules illustrated by SHELX; new autotracing features, Acta Crystallogr. Sect. D Struct. Biol. 74 (2018) 106–116, https://doi.org/10.1107/S2059798317015121.
[34] I. Usón, G.M. Sheldrick, Model Building in SHELXE, BioRxiv, 2022, 2022.04.28.489939, https://doi.org/10.1101/2022.04.28.489939.
[35] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. Sect. D Biol. Crystallogr. 60 (2004) 2126–2132, https://doi.org/10.1107/S0907444904019158.
[36] M.D. Winn, G.N. Murshudov, M.Z. Papiz, Macromolecular TLS refinement in REFMAC at moderate resolutions, Methods Enzymol. 374 (2003) 300–321, https://doi.org/10.1016/S0076-6879(03)74014-2.
[37] P.D. Adams, P.V. Afonine, G. Bunkóczi, V.B. Chen, I.W. Davis, N. Echols, J. J. Headd, L.W. Hung, G.J. Kapral, R.W. Grosse-Kunstleve, A.J. McCoy, N. W. Moriarty, R. Oeffner, R.J. Read, D.C. Richardson, J.S. Richardson, T. C. Terwilliger, P.H. Zwart, PHENIX: a comprehensive Python-based system for macromolecular structure solution, Acta Crystallogr. Sect. D Biol. Crystallogr. 66 (2010) 213–221, https://doi.org/10.1107/S0907444909052925.
[38] L. Schrödinger, W. DeLano. http://www.pymol.org/pymol, 2020.
[39] E.F. Pettersen, T.D. Goddard, C.C. Huang, E.C. Meng, G.S. Couch, T.I. Croll, J. H. Morris, T.E. Ferrin, U.C.S.F. ChimeraX, Structure visualization for researchers, educators, and developers, Protein Sci. 30 (2021) 70–82, https://doi.org/10.1002/pro.3943.
[40] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, T.E. Ferrin, UCSF Chimera - a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612, https://doi.org/10.1002/jcc.20084.
[41] L. Zimmermann, A. Stephens, S.Z. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A.N. Lupas, V. Alva, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core, J. Mol. Biol. 430 (2018) 2237–2243, https://doi.org/10.1016/j.jmb.2017.12.007.
[42] L. Holm, DALI and the persistence of protein shape, Protein Sci. 29 (2020) 128–140, https://doi.org/10.1002/pro.3749.
[43] M. van Kempen, S.S. Kim, C. Tumescheit, M. Mirdita, C.L.M. Gilchrist, J. Söding, M. Steinegger, Foldseek: fast and accurate protein structure search, Nat. Biotechnol. (2023), https://doi.org/10.1038/s41587-023-01773-0.
[44] P. Schuck, Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling, Biophys. J. 78 (2000) 1606–1619, https://doi.org/10.1016/S0006-3495(00)76713-0.
[45] T.M. Laue, B.D. Shah, T.M. Ridgeway, S.L. Pelletier, Computer-aided interpretation of analytical sedimentation data for proteins, in: S. Harding, A. Rowe, J. Horton (Eds.), Anal. Ultracentrifugation Biochem. Polym. Sci, Royal Society of Chemistry, Cambridge, UK, 1992, pp. 90–125.
[46] A. Grosdidier, V. Zoete, O. Michielin, SwissDock, a protein-small molecule docking web service based on EADock DSS, Nucleic Acids Res. 39 (2011) W270–W277, https://doi.org/10.1093/nar/gkr366.
[47] M. Oliveberg, S. Vuilleumier, A.R. Fersht, Thermodynamic study of the acid denaturation of barnase and its dependence on ionic strength: evidence for residual electrostatic interactions in the acid/thermally denatured state, Biochemistry 33 (1994) 8826–8832, https://doi.org/10.1021/bi00195a026.
[48] N. Nagano, M. Ota, K. Nishikawa, Strong hydrophobic nature of cysteine residues in proteins, FEBS Lett. 458 (1999) 69–71, https://doi.org/10.1016/S0014-5793(99)01122-9.
[49] R. O'Dwyer, R. Razzaque, X. Hu, S.K. Hollingshead, J.G. Wall, Engineering of cysteine residues leads to improved production of a human dipeptidase enzyme in *E. coli*, Appl. Biochem. Biotechnol. 159 (2009) 178–190, https://doi.org/10.1007/s12010-008-8379-9.
[50] C. Stover, M.P. Mayhew, M.J. Holden, A. Howard, D.T. Gallagher, Crystal- lization and 1.1-Å diffraction of chorismate lyase from Escherichia coli, J. Struct. Biol. 129 (2000) 96–99, https://doi.org/10.1006/jsbi.1999.4205.
[51] S.R. Price, K. Nagai, Protein engineering as a tool for crystallography, Curr. Opin. Biotechnol. 6 (1995) 425–430, https://doi.org/10.1016/0958-1669(95)80072-7.
[52] S. Panjikar, V. Parthasarathy, V.S. Lamzin, M.S. Weiss, P.A. Tucker, On the combination of molecular replacement and single-wavelength anomalous diffraction phasing for automated structure determination, Acta Crystallogr. Sect. D Biol. Crystallogr. 65 (2009) 1089–1097, https://doi.org/10.1107/S0907444909029643.
[53] J.K. Nuñez, P.J. Kranzusch, J. Noeske, A.V. Wright, C.W. Davies, J.A. Doudna, Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity, Nat. Struct. Mol. Biol. 21 (2014) 528–534, https://doi.org/10.1038/ns mb.2820.
[54] Y. Xiao, S. Ng, K. Hyun Nam, A. Ke, How type II CRISPR-Cas establish immunity through Cas1-Cas2-mediated spacer integration, Nature 550 (2017) 137–141, https://doi.org/10.1038/nature24020.
[55] C. Hu, C. Almendros, K.H. Nam, A.R. Costa, J.N.A. Vink, A.C. Haagsma, S.R. Bagde, S.J.J. Brouns, A. Ke, Mechanism for Cas4-assisted directional spacer acquisition in CRISPR–Cas, Nature 598 (2021) 515–520, https://doi.org/10.1038/s41586-021-03951-z.
[56] Y. Zhao, J. Li, H. Gu, D. Wei, Y. Chang Xu, W. Fu, Z. Yu, Conformational preferences of π–π stacking between ligand and protein, analysis derived from crystal structure data geometric preference of π–π interaction, Interdiscip. Sci. – Comput. Life Sci. 7 (2015) 211–220, https://doi.org/10.1007/s12539-015-0263-z.

[57] T. Alev-Behmoaras, J.J. Toulmé, C. Hélène, Effect of phosphate ions on the fluorescence of tryptophan derivatives. Implications in fluorescence investigation of protein-nucleic acid complexes, Biochimie 61 (1979) 957–960, https://doi.org/10.1016/S 0300-9084(79)80246-1.
[58] N. Greenfield, G.D. Fasman, Computed circular dichroism spectra for the evaluation of protein conformation, Biochemistry 8 (1969) 4108–4116, https://doi.org/10.1021/bi00838a031.
[59] G. Böhm, R. Muhr, R. Jaenicke, Quantitative analysis of protein far UV circular dichroism spectra by neural networks, Protein Eng. Des. Sel. 5 (1992) 191–195, https://doi.org/10.1093/protein/5.3.191.
[60] G.V. Semisotnov, N.A. Rodionova, O.I. Razgulyaev, V.N. Uversky, A.F. Gripas', R. I. Gilmanshin, Study of the "molten globule" intermediate state in protein folding by a hydrophobic fluorescent probe, Biopolymers 31 (1991) 119–128, https://doi.org/10.1002/bip.360310111.
[61] O.B. Ptitsyn, Protein folding: hypotheses and experiments, J. Protein Chem. 6 (1987) 273–293, https://doi.org/10.1007/BF00248050.