STRUCTURAL BIOLOGY

ISSN 2059-7983

# Molecular basis of transcriptional repression of anti-CRISPR by anti-CRISPR-associated 2

So Yeon Lee,[a,b]‡ Gi Eob Kim[a,b]‡ and Hyun Ho Park[a,b]*

[a]College of Pharmacy, Chung-Ang University, Seoul 06974, Republic of Korea, and [b]Department of Global Innovative Drugs, Graduate School of Chung-Ang University, Seoul 06974, Republic of Korea. *Correspondence e-mail: xrayleox@cau.ac.kr

Received 7 July 2021
Accepted 4 November 2021

Edited by P. Langan, Oak Ridge National Laboratory, USA

‡ These authors contributed equally to this work.

**Keywords:** Aca2; adaptive immunity; anti-CRISPR-associated 2; anti-CRISPR; CRISPR–Cas; *Oceanimonas smirnovii*; crystal structure.

**PDB reference**: Aca2, 7ezy

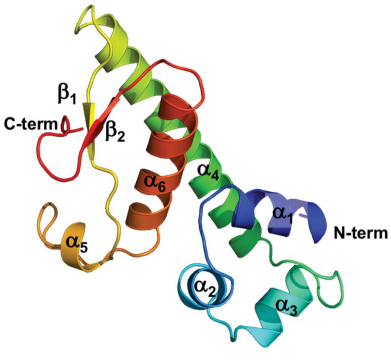
**Supporting information**: this article has supporting information at journals.iucr.org/d

CRISPR–Cas systems are well known host defense mechanisms that are conserved in bacteria and archaea. To counteract CRISPR–Cas systems, phages and viruses have evolved to possess multiple anti-CRISPR (Acr) proteins that can inhibit the host CRISPR–Cas system via different strategies. The expression of *acr* genes is controlled by anti-CRISPR-associated (Aca) proteins that bind to an upstream promoter and regulate the expression of *acr* genes during transcription. Although the role of Aca as a transcriptional repressor has been demonstrated, the mechanism of action of Aca has not been determined. Here, the molecular mechanism underlying the Aca2-mediated transcriptional control of *acr* genes was elucidated by determining the crystal structure of Aca2 from *Oceanimonas smirnovii* at a high resolution of 1.92 Å. Aca2 forms a dimer in solution, and dimerization of Aca2 is critical for specific promoter binding. The promoter-binding strategy of dimeric Aca2 was also revealed by performing mutagenesis studies. The atomic structure of the Aca family shown in this study provides insights into the fine regulation of host defense and immune-escape mechanisms and also demonstrates the conserved working mechanism of the Aca family.

## 1. Introduction

For protection against invasion by phages and other mobile genetic elements, bacteria and archaea have evolved to develop clustered regularly interspaced short palindromic repeats (CRISPR)–CRISPR-associated protein (Cas) systems. CRISPR are repetitive invader-originated DNA sequences found in the genomes of bacteria and archaea that can detect/destroy the genomic sequences of invaders during subsequent infections (Sorek *et al.*, 2008; Barrangou *et al.*, 2007; Mojica & Rodriguez-Valera, 2016). To execute CRISPR–Cas systems in the host, Cas proteins form a complex with a trans-activating CRISPR RNA (crRNA) which is present in the host CRISPR array (Brouns *et al.*, 2008; Marraffini, 2015). This Cas protein–crRNA complex detects and cleaves specific DNA or RNA sequences of invaders that are complementary to the crRNA (Marraffini, 2015; Brouns *et al.*, 2008). Due to the ability of CRISPR–Cas systems to cleave the desired DNA sequence, they can be applied to edit specific genes within organisms in various applications ranging from basic biological research to disease treatments (Hsu *et al.*, 2014; Wang *et al.*, 2013).

To counteract the CRISPR–Cas systems of bacteria and archaea that confer resistance to foreign genetic material from invaders, the evolution of phages has resulted in the development of multiple anti-CRISPR (Acr) proteins that can inhibit the host CRISPR–Cas machinery via different strategies (Borges *et al.*, 2017; Bondy-Denomy, 2018; Zhu *et al.*,



2018; Kim *et al.*, 2020; Lee *et al.*, 2020). Identifying *acr* genes is difficult as they are highly divergent in sequence. However, more than 60 *acr* genes have been identified based on genome examinations performed using advanced machine-learning tools (Borges *et al.*, 2017; Eitzinger *et al.*, 2020; Pawluk *et al.*, 2016). Interestingly, although the sequences and genomic locations of *acr* genes are diverse, they are consistently found upstream of a specific gene encoding a DNA-binding protein called anti-CRISPR-associated (Aca) protein, forming an *acr–aca* operon. Aca represses the transcription of *acr* genes (Birkholz *et al.*, 2019; Stanley *et al.*, 2019). Seven different Aca families have been identified in various phage and bacterial species (Stanley *et al.*, 2019; Borges *et al.*, 2017). One of the Aca families, Aca2, is associated with transcriptional control of the expression of various Acr proteins (Birkholz *et al.*, 2019). Genomic investigations have shown that Aca2 from *Pectobacterium carotovorum* phage ZF40 is associated with the *acrIF8* gene, while Aca2 from *Shewanella xiamenensis* is associated with the *acrIF10* gene (Birkholz *et al.*, 2019).

The main goal of this study was to elucidate the molecular mechanism underlying the Aca2-mediated transcriptional control of *acr* genes in order to better understand the regulatory role of Aca proteins in the anti-CRISPR system. A consensus has not been established with respect to the molecular architecture of the Aca family. This is the first study to demonstrate the structure of Aca2. In this study, we describe a strategy for the binding of Aca2 to the promotor of the *acr–aca* operon by providing the crystal structure of Aca2 from *Oceanimonas smirnovii* at a high resolution of 1.92 Å.

## 2. Materials and methods

### 2.1. Protein expression and purification

The expression plasmid for the full-length Aca2 sequence from *O. smirnovii* (corresponding to amino acids Met1–Glu125) was constructed by inserting the synthesized gene product, digested at the NdeI and XhoI restriction sites, into a pET-21a vector. The gene sequence was derived from the NCBI reference sequence WP_019933869, and gene synthesis was conducted by BIONICS (Seoul, Republic of Korea). The expression vector was delivered into *Escherichia coli* strain BL21 (DE3) using heat shock at 42°C. The transformed bacteria were spread on a lysogeny broth (LB) agar plate containing ampicillin and were incubated at 37°C for 16 h. A single recombinant colony was selected and cultured overnight at 37°C in 5 ml LB containing 50 µg ml$^{-1}$ ampicillin, after which the cells were transferred and cultured on a large scale in 1 l medium. When the optical density at 600 nm had reached approximately 0.7–0.8, 0.25 m*M* isopropyl *β*-D-1-thiogalactopyranoside (IPTG) was added to the medium to induce gene expression. The IPTG-treated cells were chilled on ice and further cultured at 20°C for 18 h using a shaking incubator. Subsequently, the bacterial cells were harvested and the pellet was resuspended in 16 ml lysis buffer (20 m*M* Tris–HCl pH 7.9, 500 m*M* NaCl). After adding phenylmethanesulfonyl fluoride, a serine protease inhibitor (Sigma–Aldrich, St Louis, USA), the cells were disrupted by sonication on ice with 12 bursts of 30 s each and a 90 s interval between each burst. The cell lysate was centrifuged at 10 000*g* and 4°C for 30 min to remove the cell debris. The supernatant was collected and mixed overnight with nickel–nitrilotriacetic acid (Ni–NTA) resin (Qiagen, Hilden, Germany) by gentle agitation at 4°C. The resulting mixture was loaded onto a gravity-flow column. The Ni–NTA resin in the column was washed with 100 ml lysis buffer to remove unbound proteins. For elution, 0.6 ml elution buffer (20 m*M* Tris–HCl pH 7.9, 500 m*M* NaCl, 250 m*M* imidazole) was added to the column five times. The resulting eluate was concentrated to 20 mg ml$^{-1}$ and subjected to size-exclusion chromatography (SEC). SEC purification was conducted using an ÄKTA-explorer system (GE Healthcare, Chicago, USA) equipped with a Superdex 200 Increase 10/300 GL 24 ml column (GE Healthcare) pre-equilibrated with SEC buffer (20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl). The SEC peak fractions were pooled, concentrated to 9 mg ml$^{-1}$, flash-frozen in liquid nitrogen and stored at −80°C until further use. Protein purity was assessed by SDS–PAGE.

### 2.2. Crystallization and data collection

The crystals were obtained by the typical hanging-drop vapor-diffusion method performed at 20°C. 1 µl 9 mg ml$^{-1}$ protein sample in buffer consisting of 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl was mixed with an equal volume of reservoir solution consisting of 0.1 *M* Tris–HCl pH 8.5, 20% PEG 1000 and the droplet was allowed to equilibrate against 500 µl reservoir solution. The crystallization conditions were further optimized and finally adjusted to obtain a buffer composition consisting of 0.1 *M* Tris–HCl pH 8.9, 25% PEG 1000. Diffraction-quality crystals appeared in five days and grew to maximum dimensions of 0.2 × 0.2 × 0.2 mm. For data collection, the crystals were soaked in mother liquor supplemented with 20%(*v*/*v*) glycerol as a cryoprotectant solution, mounted and flash-cooled in a nitrogen stream at −178°C. Diffraction data were collected on beamline 5C at the Pohang Accelerator Laboratory (PAL), Pohang, Republic of Korea at a wavelength of 1.0000 Å. The diffraction data were indexed, integrated and scaled using *HKL*-2000 (Otwinowski & Minor, 1997).

### 2.3. Structure determination and analysis

The structure was resolved by the molecular-replacement (MR) phasing method performed using *Phaser* in *Phenix* (Liebschner *et al.*, 2019). A modified version of a 3D structure-prediction model developed using *SWISS-MODEL*, in which 40 amino-acid residues were removed at the N-terminus, was used as the search model. Model building and refinement were performed using *Coot* (Emsley *et al.*, 2010) and *phenix.refine* in *Phenix*, respectively. The quality of the model was validated using *MolProbity* (Chen *et al.*, 2010). All structural figures were generated using *PyMOL* (DeLano & Lam, 2005).

### 2.4. Mutagenesis

Site-directed mutagenesis was performed using a QuikChange kit (Stratagene) according to the manufacturer's protocol. The presence of the mutation was then confirmed by sequencing. Mutant proteins were prepared using the same method as described above.

### 2.5. Sequence alignment

The amino-acid sequences of Aca2 from various species were analyzed using *Clustal Omega* (http://www.ebi.ac.uk/Tools/msa/clustalo/).

### 2.6. SEC–multi-angle light-scattering (MALS) analysis

The absolute molar mass of Aca2 in solution was determined using MALS. The target protein, purified by affinity chromatography using nickel–nitrilotriacetic acid resin, was filtered using a 0.2 µm syringe filter and loaded onto a Superdex 200 10/300 gel-filtration column (GE Healthcare) that had been pre-equilibrated with SEC buffer. The mobile phase buffer flowed at a rate of 0.4 ml min$^{-1}$ at 25°C. A DAWN TREOS MALS detector (Wyatt Technology, Santa Barbara, USA) was connected to the ÄKTAexplorer system (GE Healthcare). The molecular mass of bovine serum albumin was used as the reference value. The absolute molecular mass was assessed using *ASTRA* (Wyatt Technology).

### 2.7. Electrophoretic mobility shift assay (EMSA)

Various concentrations of purified proteins, both wild type and mutant, were pre-incubated with 8 µg synthesized DNA at 4°C for 60 min in a final volume of 20 µl of a buffer consisting of 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl. The reaction was then run on 6% agarose gel for 30 min at 100 V.

### 2.8. Accession code

Atomic coordinates and structure factors for the reported crystal structure have been deposited in the Protein Data Bank as entry 7ezy.

## 3. Results and discussion

### 3.1. Structure of Aca2 from *O. smirnovii*

For the initial structural study of Aca2, the *aca2* gene from *O. smirnovii* was codon-optimized for expression in the bacterial system, synthesized and cloned into the expression plasmid. The codon-optimized gene expressed well in the bacterial system. The gene product was purified using a quick two-step chromatography method comprising Ni–NTA affinity chromatography followed by SEC. Based on the SEC profile, with elution of the protein at a volume of approximately 13–18 ml, we speculated that Aca2 was present in a dimeric state in solution (Fig. 1*a*). However, the tailing asymmetric peak indicates the formation of other tentative oligomers in solution.

The purified Aca2 protein was crystallized and diffraction data were collected to a resolution of 1.92 Å. Owing to the absence of a structural homolog of Aca2, as would be required for the molecular-replacement (MR) phasing method, structure determination was initially attempted by the *ab initio* phasing method using *ARCIMBOLDO_BORGES* (Sammito *et al.*, 2013). This method utilizes small helix and sheet fragments available in the Protein Data Bank. However, the attempt was unsuccessful. The structure was finally resolved via the MR phasing method using a modified version of a 3D structure-prediction model generated using *SWISS-MODEL* as the search model. In this structure, 40 amino-acid residues were removed at the N-terminus. The MR search model was modeled using the structure of the putative cytoplasmic protein Ydil as a template (PDB entry 1s4k; Midwest Center for Structural Genomics, unpublished work), which exhibits only 19% sequence identity to Aca2. The final structure of Aca2 was refined to $R_{work}$ = 20.0% and $R_{free}$ = 23.1%. Diffraction data and refinement statistics are summarized in Table 1. Two molecules, *A* and *B*, were observed in the crystallographic asymmetric unit (Fig. 1*b*). The final model contained full-length Aca2 (residues 1–124) in both molecules. The last amino-acid residue (Glu125) and the extra LEHHHHHH residues at the C-terminus, which were derived from the expression construct and used in Ni–NTA affinity chromatography, were not included in the final model owing to

**Table 1**
Data-collection and refinement statistics.

Values in parentheses are for the outermost resolution shell. R.m.s.d., root-mean-square deviation.

| | |
|---|---|
| Data collection | |
| Space group | $P4_12_12$ |
| *a*, *b*, *c* (Å) | 64.13, 64.13, 135.52 |
| $\alpha$, $\beta$, $\gamma$ (°) | 90, 90, 90 |
| Resolution range (Å) | 28.98–1.92 |
| Total reflections | 575398 |
| Unique reflections | 22344 |
| Multiplicity | 25.8 (11.33) |
| Completeness (%) | 99.86 (99.81) |
| Mean $I/\sigma(I)$ | 18.1 (2.32) |
| $R_{merge}$† | 0.135 |
| Wilson *B* factor (Å$^2$) | 32.29 |
| Refinement | |
| Resolution range (Å) | 28.98–1.92 |
| Reflections | 22342 |
| $R_{work}$ | 0.2003 |
| $R_{free}$ | 0.2330 |
| No. of molecules in the asymmetric unit | 2 |
| No. of non-H atoms | |
| Macromolecules | 2209 |
| Solvent | 217 |
| Average *B*-factor values (Å$^2$) | |
| Overall | 36.99 |
| Macromolecules | 36.51 |
| Solvent | 41.42 |
| Ramachandran plot | |
| Favored (%) | 97.95 |
| Allowed (%) | 2.05 |
| Outliers (%) | 0.00 |
| Rotamer outliers (%) | 0 |
| Clashscore | 4.29 |
| R.m.s.d., bond lengths (Å) | 0.001 |
| R.m.s.d., angles (°) | 2.24 |

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the *i*th observed intensity of reflection *hkl* and $\langle I(hkl)\rangle$ is the average intensity obtained from multiple measurements.

unclear electron density (Fig. 1*b*). The overall structure of Aca2 consisted of an $\alpha$-helical bundle fold comprising six $\alpha$-helices ($\alpha_1$–$\alpha_6$) with two parallel $\beta$-sheets (Figs. 1*c* and 1*d*). Analysis of the surface electrostatic potential of Aca2 showed that the N-terminal region of Aca2 was composed of three $\alpha$-helices, $\alpha_1$–$\alpha_3$, and was negatively charged. The C-terminal region was positively charged (Fig. 1*e*). The structures of the two molecules in the asymmetric unit were almost identical, with a root-mean-square deviation (r.m.s.d.) of 0.34 Å (Fig. 1*f*). *B*-factor analysis showed higher *B* factors for the $\alpha_4$–$\beta_1$ connecting loop (average 62.34 Å$^2$). However, most of the structure was rigid, with low *B* factors (average of 36.58 Å$^2$; Fig. 1*g*).

### 3.2. Aca2 forms a dimer in solution

Determination of the working stoichiometry of the Aca family remains a major issue, and previous studies have proposed a monomeric or dimeric structure for the family (Birkholz *et al.*, 2019; Stanley *et al.*, 2019). We analyzed the exact stoichiometry of Aca2 in solution by calculating the absolute molecular mass using MALS. Although peak-tailing was detected in the SEC experiment, the molecular mass of the main peak was calculated to be 49.8 kDa (1.2% fitting error) by MALS (Fig. 2*a*). Considering that the theoretically calculated molecular mass of Aca2 is 18.22 kDa, the experimental molecular mass suggested a size slightly larger than that of a dimeric form but smaller than that of a trimeric form

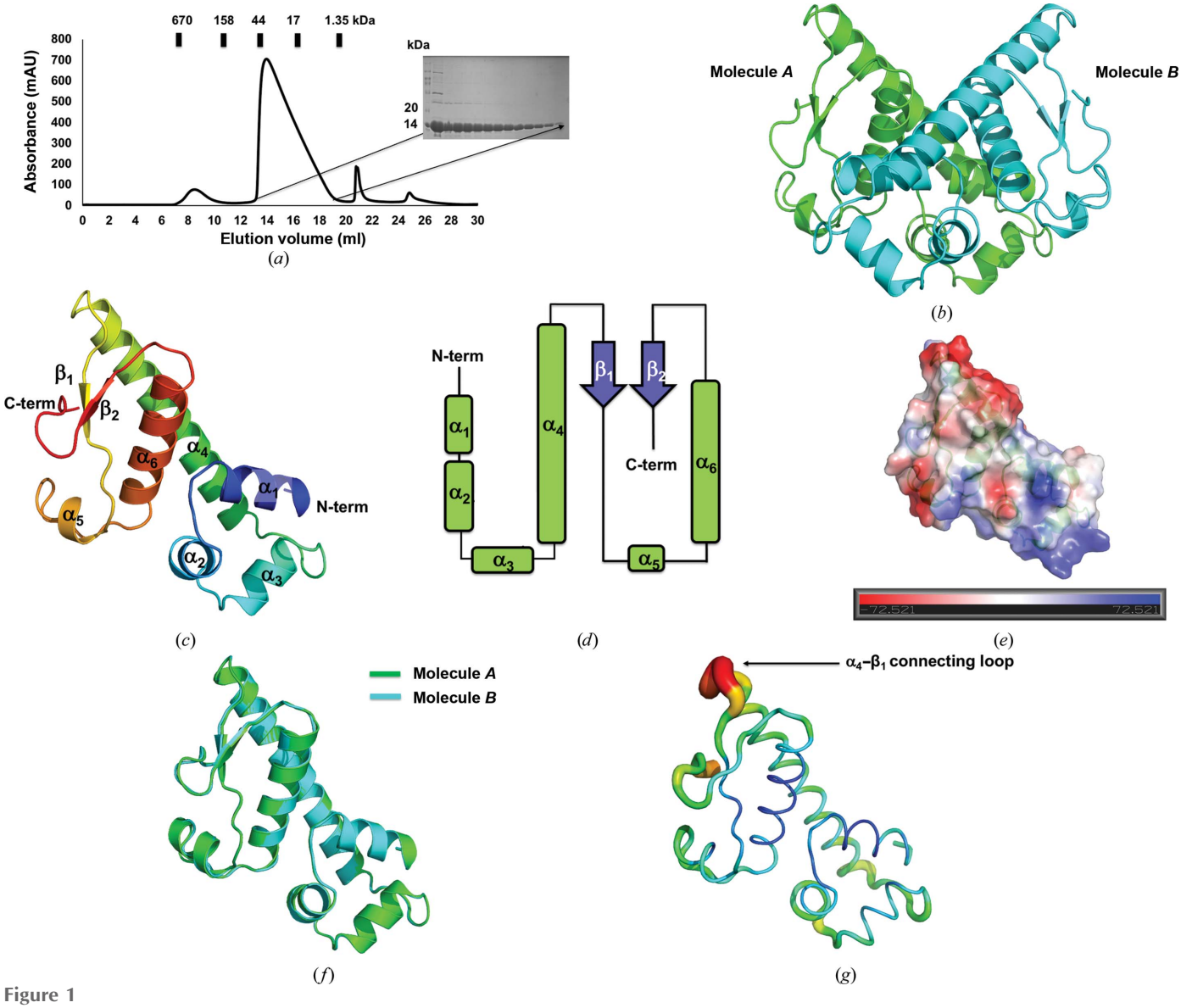
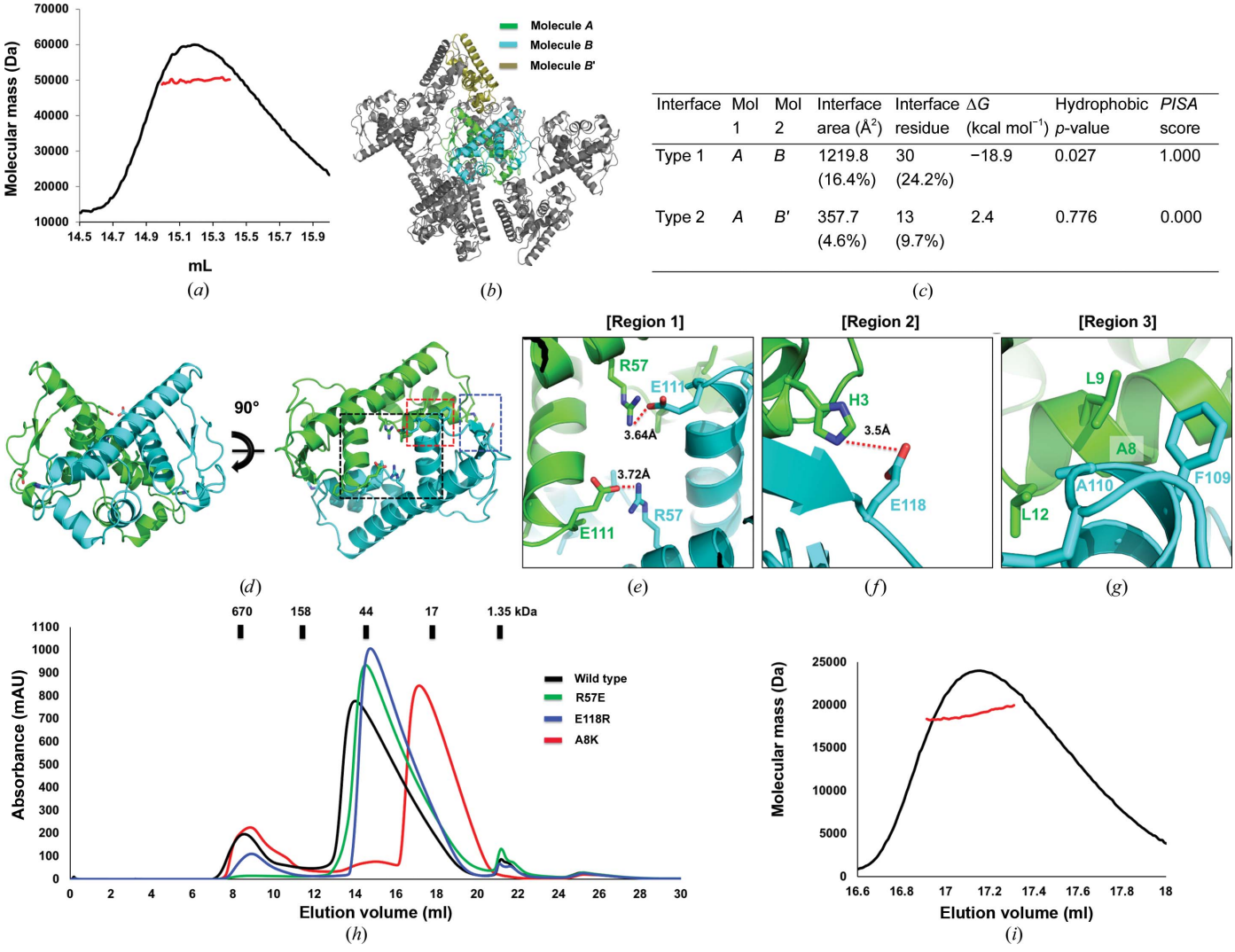



**Figure 1**
Crystal structure of Aca2 from *O. smirnovii* (osAca2). (*a*) Size-exclusion chromatography profile of osAca2. The SDS–PAGE gel shows the protein eluting at the peak position. (*b*) Cartoon representation of the structure of dimeric osAca2 presented as the crystallographic asymmetric unit. (*c*) Cartoon representation of the monomeric osAca2 structure. The color of the chain gradually moves through the spectrum from blue to red from the N-terminus to the C-terminus. (*d*) Topological representation of the osAca2 structure. (*e*) Surface electrostatic potential of osAca2. The scale bar ranges from −7.2 *kT*/e (red) to 7.2 *kT*/e (blue). (*f*) Superimposition of the structures of the molecules present in one asymmetric unit. (*g*) *B*-factor distribution in the structure of osAca2. The structure is presented in a putty representation, and rainbow colors from red to violet demonstrate the order of the *B*-factor values. The region with the highest *B* factor, corresponding to the $\alpha_4$–$\beta_1$ connecting loop, is indicated.

for Aca2 in solution. As we observed Aca2 in a dimeric or trimeric form in solution, we analyzed the crystallographic packing to search for symmetrical molecules and to further understand the tentative dimeric or trimeric structure of Aca2. The tentative trimeric form of Aca2 was not formed, although a putative alternative dimeric structure may be formed between molecule *A* and a symmetry molecule *B*′ (Fig. 2*b*). Due to the possibility of two types of interactions between molecules *A* and *B* and molecules *A* and *B*′, we analyzed the protein–protein interactions (PPIs) in both the *A*/*B* dimer and the *A*/*B*′ dimer using the *PDBePISA* PPI-calculating server (Fig. 2*c*). The interface of the *A*/*B* dimer had a complex-formation significance score of 1.000; this score ranges from 0 to 1 with increasing relevance of the interface to complex formation. The interface of the *A*/*B*′ dimer scored 0.00, indicating that the *A*/*B* dimer might be a real dimer formed in solution (Fig. 2*c*). In the *A*/*B* dimer, the total dimer surface buried an area of 1219.8 Å$^2$, which represents 16.4% of the total surface area. A total of 60 residues (30 from each molecule) were involved in PPI (Fig. 2*c*). Dimer formation was mediated by three distinct interface regions (Fig. 2*d*). The first region was formed in the central region of the symmetric dimer (Fig. 2*e*). Hydrogen bonds formed by Arg57 and Glu111 from each molecule were the primary forces maintaining the integrity of this dimeric interface. The second interface region was formed between the N-terminal helix of one molecule and




(*a*) (*b*)

| Interface | Mol 1 | Mol 2 | Interface area (Å$^2$) | Interface residue | Δ*G* (kcal mol$^{-1}$) | Hydrophobic *p*-value | *PISA* score |
|---|---|---|---|---|---|---|---|
| Type 1 | *A* | *B* | 1219.8 (16.4%) | 30 (24.2%) | −18.9 | 0.027 | 1.000 |
| Type 2 | *A* | *B*′ | 357.7 (4.6%) | 13 (9.7%) | 2.4 | 0.776 | 0.000 |

(*c*) (*d*) (*e*) (*f*) (*g*) (*h*) (*i*)

**Figure 2**
Dimeric structure of osAca2 and analysis of its interface. (*a*) Multi-angle light-scattering (MALS) profile corresponding to the main peak of SEC. The red line indicates the experimental molecular mass analyzed by MALS. (*b*) Crystallographic packing symmetry analysis. Thw two molecules found in the asymmetric unit are indicated by the green and cyan ribbon structures, while the other symmetric molecules are indicated by gray ribbon structures. The symmetrical molecule that might form a dimer with molecule *A* is indicated in yellow. (*c*) Table summarizing the interaction details of the two types of interfaces analyzed by the *PISA* server. (*d*) The tentative dimeric structure of osAca2 generated and analyzed by symmetry analysis and the *PISA* server. The regions of PPI magnified and presented in (*e*), (*f*) and (*g*) are indicated by a black dashed square for region 1, a blue dashed square for region 2 and a red dashed square for region 3. Close-up views of three different PPIs, region 1 (*e*), region 2 (*f*) and region 3 (*g*), in the dimeric structure of osAca2. The red dashed lines indicate hydrogen bonds. (*h*) Verification of the PPI via mutagenesis. SEC profiles comparing the position of eluted peaks for wild-type and mutant proteins. (*i*) MALS profile of the A8K mutant of osAca2. The red line indicates the experimental molecular mass analyzed by MALS.

the C-terminal loop of another molecule. His3 formed a hydrogen bond to Gly118 from the opposite molecule to stabilize this interface (Fig. 2*f*). The final interface (region 3) was generated by hydrophobic interactions between several hydrophobic residues, including Ala8, Leu9 and Leu12 from the $\alpha_1$ helix of one molecule and Phe109 and Ala110 from the $\alpha_6$ helix of its counterpart (Fig. 2*g*). In the case of the *A*/*B*′ dimer, the total dimer surface buried 357.7 Å$^2$, which represents 4.6% of the total surface area calculated by *PDBePISA* (Fig. 2*c*). This PPI analysis indicated that the *A*/*B* dimer is an appropriate dimer that might be formed in solution. To confirm the PPI analysis data, we performed a mutagenesis study. Because Arg57, Glu118 and Ala8 were primary interface residues involved in the formation of three different PPI regions in the *A*/*B* dimer, they were mutated to glutamic acid, arginine and lysine, respectively, producing R57E, E118R and A8K mutants. These mutants were then subjected to SEC to analyze the effect of the mutations on dimer formation. Although the R57E and E118R mutations did not affect dimer formation, the A8K mutant (a tentative region 3-disrupted mutant) showed a definite disruptive effect by generating a new peak at a different position corresponding to the size of a monomer in the SEC profile (Fig. 2*h*). The absolute molecular mass of the newly generated tentative monomer peak of the A8K mutant was further calculated by MALS to confirm the change in stoichiometry on mutagenesis. The molecular mass of the moved peak corresponding to the size of a monomer produced by the A8K mutant was 18.9 kDa (5.0% fitting error), indicating that this peak indeed represented a monomeric form of Aca2 that was produced by disruption of the dimer interface in region 3 (Fig. 2*i*). Based on the determined dimeric structure in the asymmetric unit, analysis of crystal packing and mutagenesis studies, we concluded that Aca2 exists as a dimer in solution and that the dimer of Aca2 is mainly held together by hydrophobic interactions.

### 3.3. Aca2 directly binds to the promoter of the *acr6* gene and the dimeric structure is critical for promoter recognition

The presence of Aca proteins was initially studied using Aca1 from bacteriophage and Aca2 from *P. carotovorum* phage ZF40 (Birkholz *et al.*, 2019; Stanley *et al.*, 2019; Bondy-Denomy *et al.*, 2013). Because previous biochemical studies showed a direct interaction of Aca proteins with the promoter region of *acr* genes, we tested the direct interaction of *O. smirnovii* Aca2 (osAca2) with specific DNA. Electrostatic surface analysis showed that the upper side of the dimer formed a highly negatively charged pocket. The bottom side of the dimer formed a highly positively charged pocket that might correspond to the position of DNA interaction (Fig. 3*a*). In addition, the bottom side of the osAca2 dimer generated a round hole that could accommodate the rod-shaped DNA, indicating that osAca2 binds to DNA via the bottom side of the deep groove formed by dimeric osAca2 (Fig. 3*b*).

To elucidate the direct interactions of the promoter and the molecular mechanism underlying promoter recognition by osAca2 to regulate the expression of *acr* genes, we analyzed the DNA sequences of the *acrIF6–aca2* promoter of *O. smirnovii* and examined the tentative Aca2 binding sites. We found two inverted repeats (P1 and P2) upstream of the *acrIF6* gene which might represent the binding site for osAca2 (Fig. 3*c*). To analyze the interaction of osAca2 with the promoter sequences, we synthesized three DNA fragments, P1–P2 (GTTCGCAATTGCGAACTAAGATGGAACCAG ATTCGAGATTGGCTCGAATC), P1 (GTTCGCAATTGC GAAC) and P2 (GATTCGAGATTGGCTCGAATC), and performed EMSAs. Upon the incubation of purified osAca2 protein with the P1–P2 DNA fragment, which contains two putative osAca2 binding sites, a concentration-dependent gel shift was clearly observed, indicating that osAca2 directly binds to this region of the *acrIF6* promoter of *O. smirnovii* (Fig. 3*d*). To confirm the specific interaction of osAca2 with the P1–P2 DNA fragment, EMSA was performed with three DNA fragments (N1, N2 and N3) which did not contain any inverted sequences. All three nonspecific DNA fragments did not demonstrate a shift upon incubation with osAca2, indicating that osAca2 specifically interacts with the P1–P2 DNA fragment (Fig. 3*e*). When the P1 and P2 DNA fragments were individually incubated with osAca2, a gel shift similar to that observed with the P1–P2 fragment was noticed, indicating that both P1 and P2 might be osAca2 binding sites in the promoter of osAca2 (Figs. 3*f* and 3*g*). However, a larger band shift with a lower concentration of the P1 DNA fragment indicated stronger binding of osAca2 to the P1 site.

To determine whether dimerization of osAca2 was critical for binding of the promoter, we performed EMSA again using a dimerization-defective mutant form of osAca2 (the A8K mutant). The A8K mutant did not demonstrate a gel shift with any DNA fragment even at the highest concentration, indicating that the dimerization-defective mutant has lost its ability to bind to specific DNA sequences (Fig. 3*h*). In conclusion, osAca2 directly binds to specific inverted sequences in the promoter of the *acrIF6* gene and dimerization of osAca2 is critical for promoter binding.

### 3.4. Strategy of promoter recognition by osAca2

To identify the DNA-binding region of osAca2, which is a crucial functional region, a phylogenetic analysis of homologous Aca2 sequences from various species was performed using the *ConSurf* server (Ashkenazy *et al.*, 2016). This analysis, in conjunction with sequence comparison, showed that half of the residues from the N-terminus, consisting of $\alpha_1$–$\alpha_3$, were more conserved, while the last region comprising the residues at the C-terminus was less conserved, indicating that the N-terminal region of osAca2 might be crucial for its function (Figs. 4*a* and 4*b*). This sequence analysis also indicated that $\alpha_2$ and $\alpha_3$ might form the typical helix–turn–helix (HTH) motif, which is a well known DNA-binding motif, supporting the hypothesis that the N-terminal region of osAca2 is critical for binding the promoter and for regulation of *acr* gene expression. The conservation of Ala8, which was critical for the dimerization of Aca2, indicates that dimerization is critical for the proper functioning of Aca proteins.

To examine the tentative HTH motif-mediated promoter-binding activity of osAca2, we performed structural modeling and docking using the *HDOCK* server (Yan *et al.*, 2017). In the top docking model, osAca2 recognized the promoter sequence of DNA by inserting $\alpha_3$ of the HTH motif into the minor groove of the double-helix DNA (Fig. 4*c*). Arg30, Gln33, Tyr34 and Arg39 were involved in reading the palindromic sequence of GTT and showed specific interactions in the promoter region of the *acr6* gene in the genome of *O. smirnovii* (Figs. 4*c* and 4*d*). Among the interacting residues, Arg30 was the major DNA-readout contributor and was localized in the center of the deep minor groove. Arg34 and Arg39 were also localized in the upper side of the center of the minor groove and might be marginally involved in the readout of specific DNA sequences. Gln33 was localized on the outer side of the minor groove and was involved in mediating a direct interaction with the DNA backbone. These observations indicated that Arg30, Tyr34 and Arg39 of osAca2 might be critical for generating the readout corresponding to a specific sequence, while Gln33 might be important for gripping the DNA sequence. The distance between the Arg30 residues in each molecule of dimeric osAca2 was approximately 32 Å

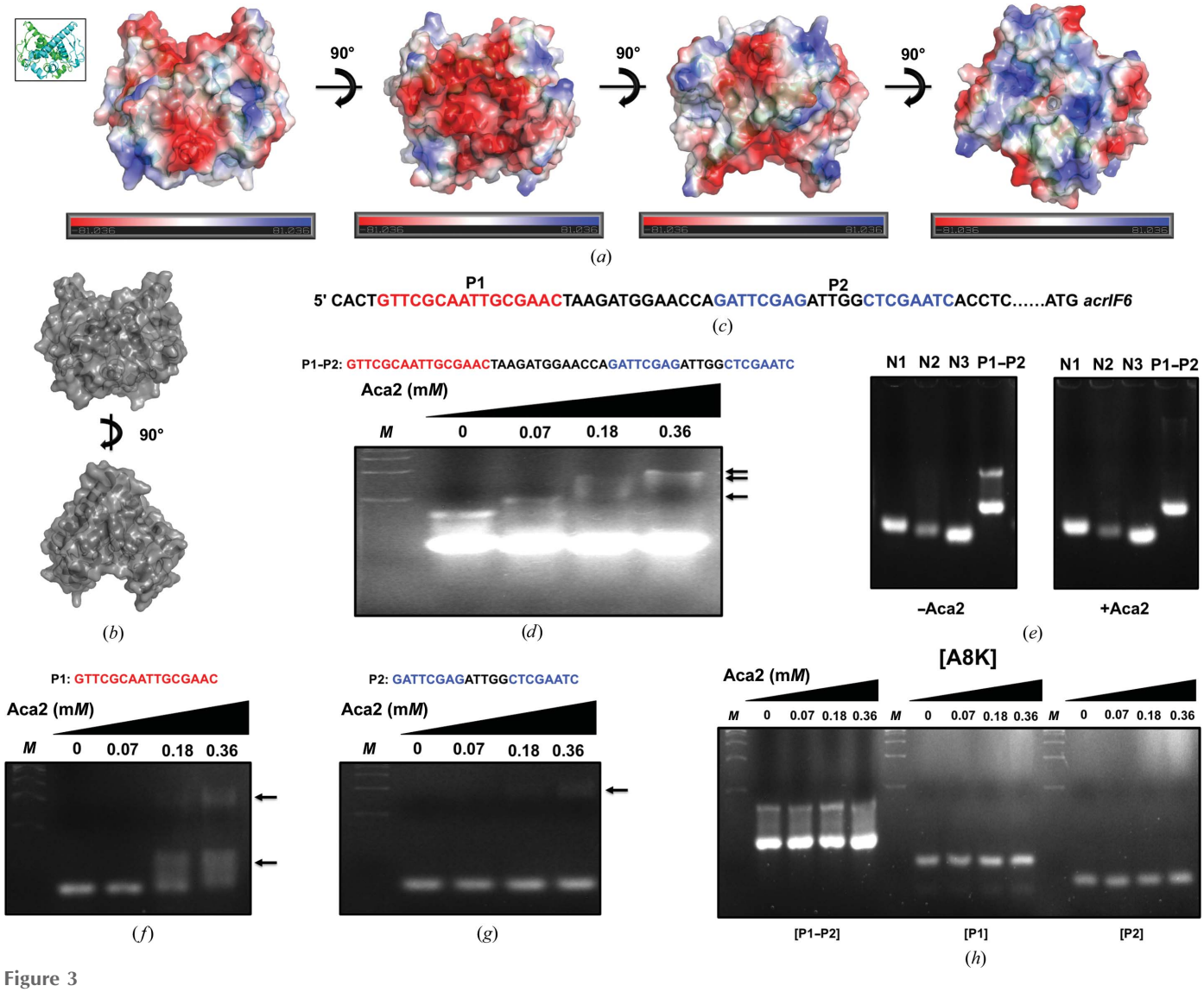



**Figure 3**
Promoter-binding activity of osAca2. (*a*) Surface electrostatic features of dimeric osAca2. The scale bar ranges from −81.0 *kT*/e (red) to 81.0 *kT*/e (blue). The direction of the dimeric structure of osAca2 observed at the start of the rotation is shown by the cartoon figure on the left. (*b*) The overall shape of the surface of osAca2. (*c*) Putative osAca2 binding sites in the promoter region of the *acr6* gene of *O. smirnovii*. Two inverted-repeat sequences (P1 and P2) upstream of the *acrIF6* gene were identified by sequence analysis. (*d*) Mobility shift of the synthesized P1–P2 DNA probe containing two inverted sequences (P1 and P2), with increasing concentrations of osAca2, on an agarose gel. The synthesized nucleotide sequences are shown on the abovementioned gel. Concentration dependence is indicated by the black triangle. Lane *M* contains molecular size markers. The shifted DNA bands are indicated by a black arrow. (*e*) Sequence-specific interaction of osAca2. N1, N2 and N3 indicate negative controls representing DNA fragments containing no inverted sequences. (*f*, *g*) Mobility shift of the synthesized DNA probes P1 (*f*) and P2 (*g*) observed on an agarose gel. (*h*) Mobility shift of synthesized DNA probes observed on an agarose gel with increasing concentrations of A8K mutant osAca2. The DNA probes used in the experiment are indicated below the gel. Concentration dependence is indicated by the black triangle. Lane *M* contains molecular size markers. The concentrations of osAca2 protein are indicated for each lane in the gel.

(Fig. 4*e*). Because the length of a complete turn of double-helical DNA is approximately 34 Å, this distance of 32 Å is a reasonable distance that can explain the recognition of two deep minor grooves by dimeric molecules of osAca2.

To experimentally confirm our docking model, we performed a mutagenesis study. Since Arg30, Gln33, Tyr34 and Arg39 were analyzed as being the main DNA-interacting residues, they were mutated to aspartic acid, arginine, tryptophan and tryptophan, respectively, producing R30D, Q33R, Y34W and R39W mutants. Each mutant was purified, and the effects of the mutation on the disruption of DNA binding were analyzed using EMSA. Upon incubation of the four purified mutants with the P1–P2 DNA fragment, the concentration-dependent gel shift noted on incubation with wild-type osAca2 was not observed, indicating that all four residues were important for P1–P2 promoter recognition (Fig. 4*f*).

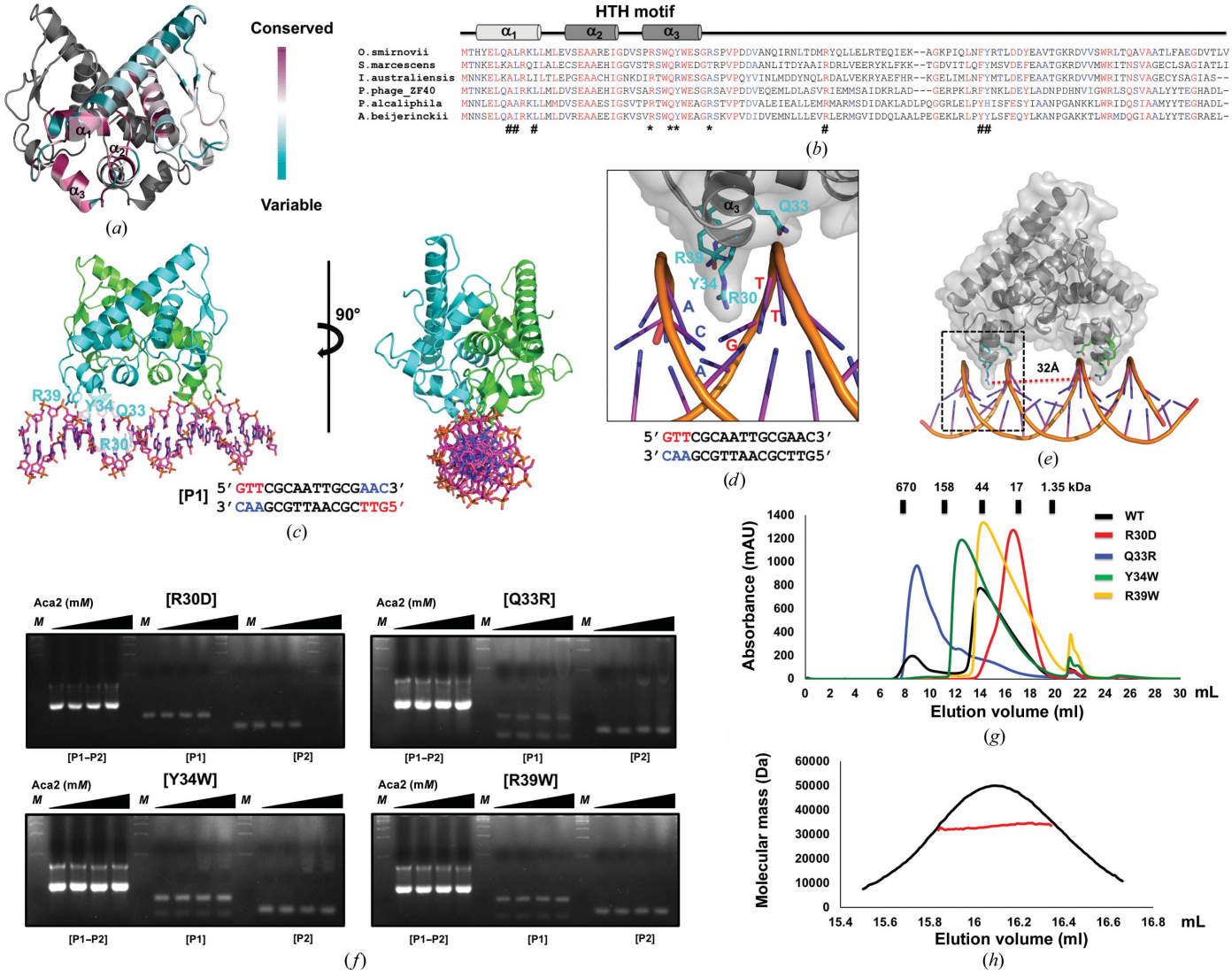



**Figure 4**
Putative strategy of promoter recognition by dimeric osAca2. (*a*) Cartoon representation of dimeric osAca2 colored according to the degree of amino-acid sequence conservation as analyzed by the *ConSurf* server. (*b*) Sequence alignment of Aca2 from different species. Mostly conserved and partially conserved residues are colored red and blue, respectively. The location of the three helices $\alpha_1$–$\alpha_3$ at the N-terminus containing the HTH motif are shown above the sequence. # indicates conserved residues involved in formation of the dimeric interface. * indicates conserved residues that might be involved in DNA recognition. (*c*) Structural model of the osAca2–P1 promoter complex generated by the *HDOCK* server. The sequence of the P1 putative osAca2 binding site used for modeling and docking is provided below the structural model. Arg30, Gln33, Tyr34 and Arg39 involved in the readout of the palindromic GTT sequence are labeled in the structure. (*d*) Close-up view of the putative binding region between osAca2 and promoter DNA. Residues involved in the readout of the palindromic GTT sequence are labeled. (*e*) Cartoon representation showing the distance between Arg30 in each molecule of dimeric osAca2. The black dashed box indicates the position magnified in (*d*). (*f*) Mobility shift of synthesized DNA probes observed on an agarose gel with increasing concentrations of DNA binding-disturbed mutants, including R30D, Q33R, Y34W and R39W. The DNA probes used in the experiment are indicated below the gel. Concentration-dependence is indicated by the black triangle. A concentration gradient, 0, 0.07, 0.18 and 0.36 m*M*, was applied for each mutant protein. Lane *M* contains molecular size markers. (*g*) Verification of the effect of mutagenesis on the stoichiometric alteration of osAca2. The SEC profiles compare the positions of the eluted peaks for the wild type and each mutant protein. (*h*) MALS profile of the R30D mutant of osAca2. The red line indicates the experimental molecular mass analyzed by MALS.

**Table 2**
Structural similarity search using *DALI*.

| Protein (PDB code) | *Z*-score | R.m.s.d. (Å) | Identity (%) | Reference |
|---|---|---|---|---|
| MqsA (3o9x) | 8.9 | 2.4 (61 residues) | 18 | Brown *et al.* (2011) |
| HigA2 (5j9i) | 8.7 | 2.6 (61 residues) | 20 | Hadži *et al.*, 2017) |
| ClgR (3f51) | 7.7 | 2.3 (54 residues) | 20 | Russo *et al.* (2009) |
| SaPI (6h49) | 7.3 | 6.2 (71 residues) | 11 | Ciges-Tomas *et al.* (2019) |

However, when the P1 and P2 DNA fragments were individually incubated with each mutant, the R30D and R39D mutants did not demonstrate any gel shift with the DNA, whereas the Q33R and Y34W mutants showed a gel shift at high concentration (greater than 0.18 m*M* mutant protein). This indicated that Arg30 and Arg39 are the most critical residues for the promoter interaction mediating the repressor function. However, all four residues, Arg30, Gln33, Tyr34 and Arg39, were important for interaction with the combined P1–P2 promoter.

Due to the critical effect of the dimeric structure of osAca2 on binding to the specific promoter DNA, the effect of mutagenesis on the stoichiometry of osAca2 was examined by SEC. For accurate measurements of the peak shift, the same protein concentration (~2 mg ml$^{-1}$) was used for SEC. As shown in Fig. 4(*g*), the R39W mutant produced a peak profile similar to the wild type, whereas the three mutants demonstrated varied SEC profiles. The Q33R mutant generated a void peak, indicating that Q33R affected the solubility of the osAca2 protein (Fig. 4*g*). However, the Q33R mutant also produced a dimeric peak. We used this dimeric form of osAca2 in the EMSA. Interestingly, the R30D mutant produced a smaller-sized peak. However, the elution point was similar to that of the wild type, eluting from the SEC column at approximately 14 ml (Fig. 4*g*). To eliminate the uncertainty in the effect of the R30D mutant on the stoichiometry, the molecular size of the R30D mutant was confirmed by MALS. The molecular mass of the moved peak produced by the R30D mutant was 33.3 kDa (1.3% fitting error), indicating that this peak indeed represents the dimeric form and that mutagenesis did not affect the stoichiometry of osAca2 (Fig. 4*h*). Notably, the Y34W mutant produced a peak corresponding to a larger size compared with the wild type, implying that the Y34W mutation produced new oligomeric forms. Further analysis of the effect of the Y34W mutation should be performed.

To obtain further insight into the function of osAca2 based on its structure, we compared the structure of osAca2 with its structural homologs using the *DALI* server (Holm & Sander, 1995). Because the structure of osAca2 was the first structure to be elucidated in the Aca family, the structural homologs of osAca2 reported by the *DALI* server were not functionally related to Aca2. The top four matches are as follows: MqsA (PDB entry 3o9x; Brown *et al.*, 2011), HigA2 (PDB entry 5j9i; Hadži *et al.*, 2017), ClgR (PDB entry 3f51; Russo *et al.*, 2009) and SaPI (PDB entry 6h69; Ciges-Tomas *et al.*, 2019) (Table 2). Although they were not functionally related to Aca2, all of the proteins selected by the *DALI* server contained an HTH motif for DNA/RNA binding. Although the sequence identity between osAca2 and MqsA was only approximately 18%, the structures of the three helices at the N-terminal region of osAca2 containing the HTH motif superimposed well with the HTH region composed of $\alpha$3–$\alpha$4 in MqsA, with an r.m.s.d. of 2.4 Å (Fig. 5*a*, Table 2). Because the MqsA structure was presented in a DNA-complexed form, we superimposed the osAca2 structure with the DNA-bound form of the dimeric MqsA structure (Fig. 5*b*). The position of the HTH motif of osAca2 corresponded to that of the HTH motif of MqsA, confirming that the $\alpha_2$–$\alpha_3$ region of osAca2 contains the HTH motif, which might recognize the target DNA sequence. To summarize, Aca2, which is a repressor of the expression of *acr* genes, forms a dimer in solution. The dimeric form is critical for the binding of specific DNA. The N-terminal region of osAca2 comprising the $\alpha_2$–$\alpha_3$ helices contains the HTH motif that is specific for the palindromic sequence of the promoter DNA. Arg30 in the $\alpha$3 helix is the most critical residue involved in DNA binding. The structural and biochemical study of Aca2 provides insights into the function of the Aca2 family and thus provides a prototype for the mechanism of action of the Aca family containing an HTH motif (Fig. 5*c*).

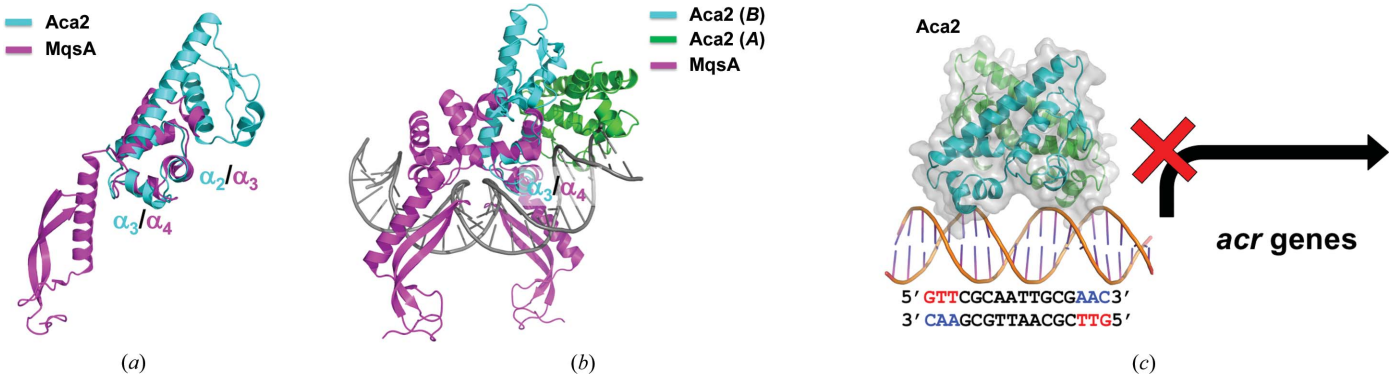



**Figure 5**
Putative model of *acr* gene repression by Aca2. (*a*) Structural comparison of monomeric osAca2 with monomeric MqsA by superimposition. The superimposed position showing the HTH motif of MqsA is labeled. (*b*) Structural comparison of dimeric osAca2 with the dimeric MqsA–DNA complex by superimposition. (*c*) Putative model of *acr* gene repression by Aca2. The tentative nucleotide sequence in the P1 promoter corresponding to osAca2 is highlighted in red and blue.

## Acknowledgements

We thank the staff of the 5C beamline at PAL, Pohang, Republic of Korea for their assistance in data collection.

## Funding information

This study was supported by the Basic Science Research Program of the National Research Foundation of Korea (NRF) of the Ministry of Education, Science and Technology (NRF-2017M3A9D8062960, NRF-2018R1A4A1023822 and NRF-2021R1A2C3003331).

## References

Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. (2016). *Nucleic Acids Res.* **44**, W344–W350.
Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D. A. & Horvath, P. (2007). *Science*, **315**, 1709–1712.
Birkholz, N., Fagerlund, R. D., Smith, L. M., Jackson, S. A. & Fineran, P. C. (2019). *Nucleic Acids Res.* **47**, 9658–9665.
Bondy-Denomy, J. (2018). *ACS Chem. Biol.* **13**, 417–423.
Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. (2013). *Nature*, **493**, 429–432.
Borges, A. L., Davidson, A. R. & Bondy-Denomy, J. (2017). *Annu. Rev. Virol.* **4**, 37–59.
Brouns, S. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J., Snijders, A. P., Dickman, M. J., Makarova, K. S., Koonin, E. V. & van der Oost, J. (2008). *Science*, **321**, 960–964.
Brown, B. L., Wood, T. K., Peti, W. & Page, R. (2011). *J. Biol. Chem.* **286**, 2285–2296.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Ciges-Tomas, J. R., Alite, C., Humphrey, S., Donderis, J., Bowring, J., Salvatella, X., Penadés, J. R. & Marina, A. (2019). *Nat. Commun.* **10**, 3676.
DeLano, W. L. & Lam, J. W. (2005). *Abstr. Pap. Am. Chem. Soc.* **230**, U1371–U1372.
Eitzinger, S., Asif, A., Watters, K. E., Iavarone, A. T., Knott, G. J., Doudna, J. A. & Minhas, F. (2020). *Nucleic Acids Res.* **48**, 4698–4708.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Hadži, S., Garcia-Pino, A., Haesaerts, S., Jurėnas, D., Gerdes, K., Lah, J. & Loris, R. (2017). *Nucleic Acids Res.* **45**, 4972–4983.
Holm, L. & Sander, C. (1995). *Trends Biochem. Sci.* **20**, 478–480.
Hsu, P. D., Lander, E. S. & Zhang, F. (2014). *Cell*, **157**, 1262–1278.
Kim, G. E., Lee, S. Y. & Park, H. H. (2020). *FEBS Open Bio*, **10**, 2532–2540.
Lee, S. Y., Kim, G. E., Kim, Y. G. & Park, H. H. (2020). *Biochem. Biophys. Res. Commun.* **533**, 751–757.
Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* D**75**, 861–877.
Marraffini, L. A. (2015). *Nature*, **526**, 55–61.
Mojica, F. J. & Rodriguez-Valera, F. (2016). *FEBS J.* **283**, 3162–3169.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pawluk, A., Staals, R. H., Taylor, C., Watson, B. N., Saha, S., Fineran, P. C., Maxwell, K. L. & Davidson, A. R. (2016). *Nat. Microbiol.* **1**, 16085.
Russo, S., Schweitzer, J. E., Polen, T., Bott, M. & Pohl, E. (2009). *J. Biol. Chem.* **284**, 5208–5216.
Sammito, M., Millán, C., Rodríguez, D. D., de Ilarduya, I. M., Meindl, K., De Marino, I., Petrillo, G., Buey, R. M., de Pereda, J. M., Zeth, K., Sheldrick, G. M. & Usón, I. (2013). *Nat. Methods*, **10**, 1099–1101.
Sorek, R., Kunin, V. & Hugenholtz, P. (2008). *Nat. Rev. Microbiol.* **6**, 181–186.
Stanley, S. Y., Borges, A. L., Chen, K. H., Swaney, D. L., Krogan, N. J., Bondy-Denomy, J. & Davidson, A. R. (2019). *Cell*, **178**, 1452–1464.
Wang, H., Yang, H., Shivalila, C. S., Dawlaty, M. M., Cheng, A. W., Zhang, F. & Jaenisch, R. (2013). *Cell*, **153**, 910–918.
Yan, Y., Zhang, D., Zhou, P., Li, B. & Huang, S. Y. (2017). *Nucleic Acids Res.* **45**, W365–W373.
Zhu, Y., Zhang, F. & Huang, Z. (2018). *BMC Biol.* **16**, 32.