



OPEN Structural basis of dimerization and cascade formation by Cas5

Yong Jun Kang^{1,2,3}, Hyun Ji Ha^{1,3}, Hyo Been Jin^{1,2}, SoYeon Lee^{1,2} & Hyun Ho Park^{1,2}✉

CRISPR-Cas systems are essential for prokaryotic adaptive immune mechanisms; however, the structural details of many subtype-specific components remain unclear. Herein, we report the crystal structure and biophysical characterization of Cas5 from *Moraxella bovoculi* (MboCas5), a component of the type I-C CRISPR-Cas system. We found that *M. bovoculi* encodes both type I-C and type III-B systems, and that MboCas5 forms a dimer that is stabilized by key interactions, including a salt bridge between R72 and D167. Structural comparisons with other Cas5 homologs and AlphaFold 3 predictions further validated the unique dimer configuration, suggesting that it is conserved across species. Additionally, structural comparison revealed a highly flexible loop region, which likely undergoes conformational changes upon Cascade assembly and might mediate interactions with Cas8 and crRNA. Overall, the findings provided structural and mechanistic insights into Cas5 function and could potentially contribute to our understanding of the assembly of type I-C Cascade complexes.

Clustered regularly interspaced short palindromic repeats (CRISPRs) provide an RNA-based adaptive immune system that protects archaea and bacteria from the invasion of foreign genetic elements such as phages and plasmids^{1–3}. CRISPR loci comprise conserved repeat sequences of approximately 20–50 base pairs, interspersed with variable spacer sequences, which were derived during prior such invasion^{1,4,5}. Together with CRISPR-associated (*cas*) genes encoding Cas proteins, these components enable the recognition and destruction of foreign genetic elements^{6,7}. The CRISPR-Cas system functions similar to the adaptive immune system of higher organisms, acquiring genetic information from past infections to mount targeted immune response and protect the host from subsequent invasion.

CRISPR-Cas systems are categorized into two classes based on the organization of the CRISPR locus and the composition of Cas proteins. Class 1 systems use multi-subunit Cas protein complexes to execute their functions, whereas class 2 systems rely on a single, multi-domain Cas protein for various activities^{6,8}. These classes are further divided into six types (I to VI) according to their mechanisms of action; types I, III, and IV belong to Class 1, whereas types II, V, and VI belong to Class 2^{5,6}.

Among the diverse CRISPR-Cas systems, type I is the most widespread and is characterized by a multiprotein complex known as the CRISPR-associated complex for antiviral defense (Cascade)^{9,10}. Cascade recognizes the invasion of a foreign DNA through a CRISPR-derived RNA (crRNA) guide, resulting in target DNA degradation via nuclease Cas3^{6,7}. The type I system is further classified into seven subtypes (I-A to I-G) based on the composition of Cas proteins. In particular, the type I-C system forms a minimal Cascade complex, which differs from other type I subtypes in terms of structural and functional organization^{10,11}. Unlike many type I subtypes, which include Cas6 for crRNA processing, the type I-C Cascade lacks Cas6 in its operon and comprises only three core proteins: Cas5, Cas7, and Cas8^{12,13}.

Cas5 plays a central role in Cascade complex stabilization and is essential for crRNA binding and target DNA recognition^{14,15}. The primary role of Cas5 is to cleave pre-crRNA and generate mature crRNA of unit length¹⁴. Despite its functional importance, the structural details of Cas5 remain largely unknown. In this study, we aimed to determine the structure of Cas5 from *Moraxella bovoculi* (hereafter called MboCas5) to gain insights into its role in Cascade assembly and function in a type I-C CRISPR-Cas system. By elucidating its structural features and comparing it with other Cas5 proteins from different types, we provide in-depth understanding of how the type I-C Cascade assembles. In addition, we elucidated the actual dimer structure of Cas5 through structural analysis, mutagenesis, and AI-based predictions. Based on these findings, we proposed a model for Cas5 dimer assembly and suggested that structural rearrangements of Cas proteins contribute to Cascade complex formation.

¹College of Pharmacy, Chung-Ang University, Seoul 06974, Republic of Korea. ²Department of Global Innovative Drugs, Graduate School of Chung-Ang University, Seoul 06974, Republic of Korea. ³Yong Jun Kang and Hyun Ji Ha contributed equally to this work. ✉email: xrayleox@cau.ac.kr

Results

Structure analysis of Cas5 from *M. bovoculi*

To understand the CRISPR-Cas system of *M. bovoculi*, we conducted a genetic analysis of the bacterium and found that it possessed two types of CRISPR-Cas systems: type III-B and type I-C (Fig. 1A). Typically, the type I-C CRISPR-Cas system forms a Cascade complex that lacks Cas6, which is responsible for RNA processing. The results showed that the Cas6 gene was absent, thereby indicating that the *M. bovoculi* CRISPR-Cas system follows the typical characteristics of a type I-C CRISPR-Cas system. Among the two CRISPR-Cas systems, we aimed to understand the assembly and mechanism of the type I-C system using structural biology approaches. To achieve this, we purified the proteins constituting the type I-C CRISPR-Cas system, including Cas1, Cas2, Cas3, Cas4, Cas5, Cas7, and Cas8. Among the proteins, MboCas5 was well expressed in *E. coli*, and we successfully obtained a large quantity of the protein for use in our structural studies with His-tag affinity chromatography and size-exclusion chromatography (SEC). The SEC results showed a nearly single peak containing the target protein, MboCas5, which eluted at approximately 14 mL, appearing between γ -globulin (158 kDa) and ovalbumin (44 kDa) (Fig. 1C and D). Based on the elution volume, we inferred that MboCas5 possibly existed as a dimer in solution.

Following purification, the MboCas5 protein was successfully crystallized and the crystal structure was resolved using the molecular replacement (MR) phasing method, with the predicted structure from AlphaFold3 (AF3) employed as the search model¹⁶. Table 1 presents an overview of the crystallographic and refinement statistics.

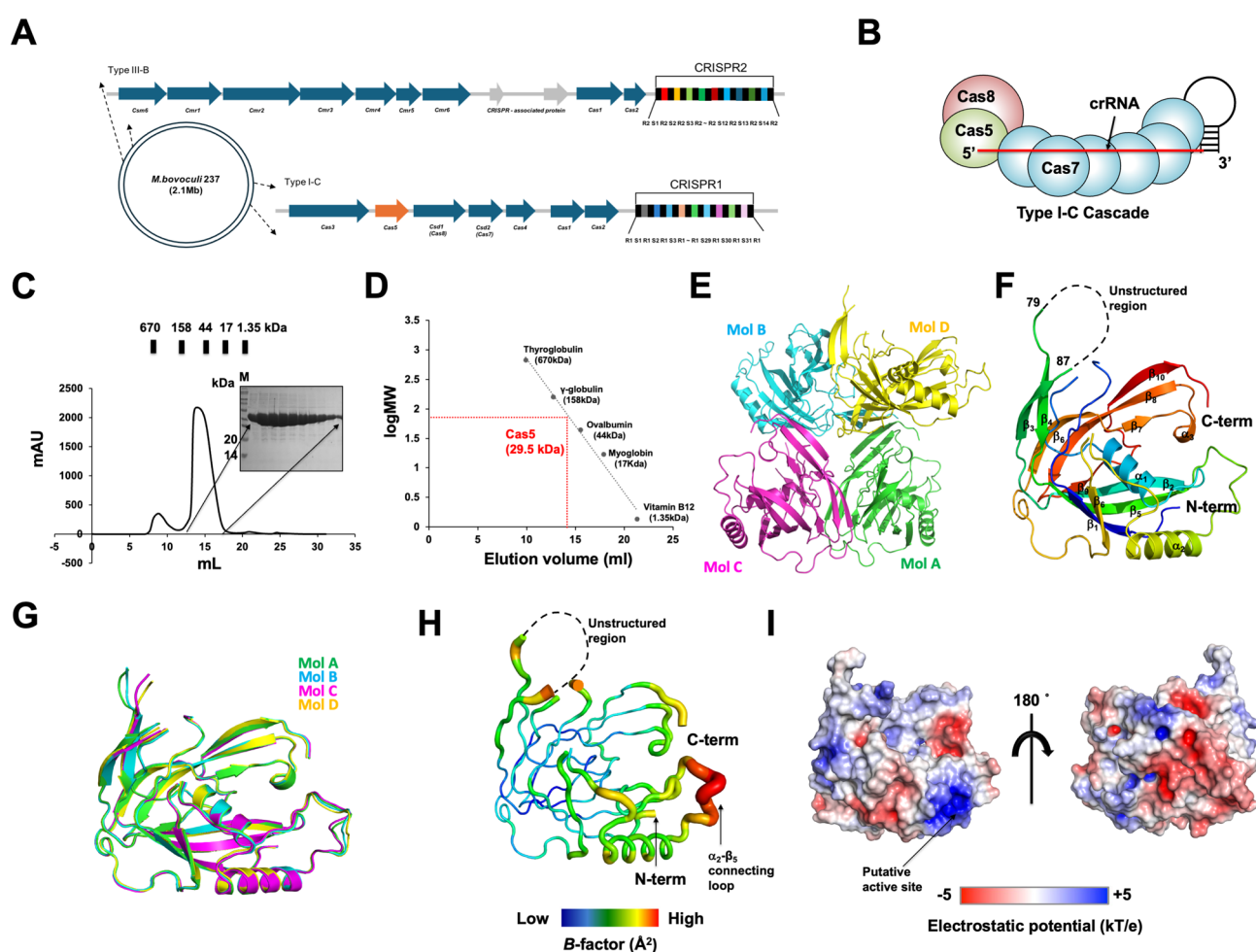


Fig. 1. Overall structure of MboCas5. (A) Schematic overview of the two different types of CRISPR-Cas systems present in *M. bovoculi*. (B) Overview of the type I-C Cascade. (C) SEC profile of MboCas5, along with an SDS-PAGE gel image, showing the eluted protein at the peak position. (D) Elution volume line-fitting in SEC plotted against the size marker and the logarithm of the molecular weight of MboCas5. The red point on the fitting line signifies the elution volume of MboCas5. (E) Four molecules of MboCas5 in the same crystallographic asymmetry unit (ASU). (F) Cartoon representation of MboCas5 structure. The color of the chain from the N- to the C-terminus gradually moves through the spectrum from blue to red. (G) Superposition of four MboCas5 structures detected in the same ASU. (H) B-factor distribution across the structure of MboCas5. The structure is shown in a putty representation, and rainbow colors from red to violet represent the B-factor values. (I) Electrostatic surface representation of MboCas5 from two angles at 180°. The scale ranges from -5.0 kT/e (red) to $+5.0$ kT/e (blue).

Data collection	
Space group	P2 ₁
Unit cell parameter <i>a</i> , <i>b</i> , <i>c</i> (Å)	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	73.03, 92.98, 83.82
α , β , γ (°)	90, 96, 90
Resolution range (Å) ¹	29.49–3.33 (3.449–3.33)
Total reflections ¹	102,988 (10,623)
Unique reflections ¹	16,138 (1,600)
Multiplicity ¹	6.4 (6.6)
Completeness (%) ¹	98.16 (99.56)
Mean <i>I</i> / σ (<i>I</i>) ¹	8.83 (2.38)
<i>R</i> _{merge} (%) ²	23.03 (84.9)
Wilson <i>B</i> -factor (Å ²)	76.34
Refinement	
Resolution range (Å)	29.49–3.33
Reflections	16,133
<i>R</i> _{work} (%)	22.21
<i>R</i> _{free} (%)	27.48
Number of molecules in the ASU	4
Number of non-hydrogen atoms	6,852
Macromolecules	6,852
Solvent	0
Average <i>B</i> -factor values (Å ²)	75.41
Macromolecules	75.41
Solvent	0
Ramachandran plot:	
favoured/allowed/outliers (%)	95.56/4.44/0.00
Rotamer outliers (%)	0
Clashscore	3.78
RMSD bonds (Å)/angles (°)	0.003/0.6

Table 1. Data collection and refinement statistics. ¹Values for the outermost resolution shell in parentheses. ² $R_{\text{merge}} = \sum_h \sum_i |I(h)_i - \langle I(h) \rangle| / \sum_h \sum_i I(h)_i$, where *I*(*h*) is the observed intensity of reflection *h*, and $\langle I(h) \rangle$ is the average intensity obtained from multiple measurements. ASU indicates asymmetric unit.

Four molecules, MolA to MolD, were found in the asymmetric unit (ASU), and the final model comprised the MboCas5 protein spanning from F8 to P223 (Fig. 1E). Seven N-terminal residues and 20 C-terminal residues were not constructed due to their untraceable electron density. Notably, in all four molecules, seven residues (positions 80–86) could not be traced and were, therefore, not included in the final model (Fig. 1F). The overall structure of MboCas5 was composed of 2 α -helices ($\alpha 1$ and $\alpha 2$) and 10 β -sheets ($\beta 1$ – $\beta 10$) (Fig. 1F). The four molecules in a single ASU exhibited similar structures. When superimposed, the root mean square deviation (RMSD) values for molA, molB, molC, and molD were 0.6, 0.5, 0.6, and 0.8, respectively (Fig. 1G). *B*-factor analysis revealed that the structure of MboCas5 is generally rigid, with low *B*-factors (average 72.2 Å²), except for the $\alpha 2$ – $\beta 5$ connecting loop region (average 122 Å²) and the unstructured region, indicating the presence of a few flexible loops despite the overall rigidity of the protein in solution (Fig. 1f). Electrostatic surface analysis revealed intricate patterns of positive, negative, and neutral charges (Fig. 1I). Based on the observation that the region identified as the active site in previous Cas5 studies also forms a basic patch in MboCas5 (Fig. 1I), MboCas5 was predicted to function similarly by binding to RNA substrates through this basic surface (Table 2).

Dimeric structure of MboCas5

The actual stoichiometry of Cas5 remains controversial due to conflicting reports of it existing as a dimer in some bacterium species, and as a monomer in others^{17,18}. As Cas5 stoichiometry likely influences its functional mechanism, we determined the exact stoichiometry of MboCas5 in solution by calculating the absolute molecular mass using multi-angle light scattering (MALS); the molecular mass of the main peak was calculated as 60.2 kDa (3.1% fitting error) (Fig. 2A). Considering that the theoretically calculated molecular weight of MboCas5 was 29.5 kDa, the experimental molecular mass of MboCas5, as analyzed by MALS, suggested that it formed dimers in solution.

To understand the dimer formation of MboCas5 in detail, we analyzed the protein-protein interaction (PPI) interface using the PDBePISA PPI-calculating server. The PISA score, which ranges from 0 to 1 as the relevance of the interface to complex formation increases, was 0.5 for the BD and AC dimers, indicating them to be the possible biologically relevant dimeric forms of MboCas5 (Fig. 2B). Since both BD and AC dimers essentially

Proteins (accession numbers)	Z-score	RMSD (Å)	Identity (%)
Cas5d from <i>A. halodurans</i> (4F3M)	26.7	1.5 (193/213)	49
Cas5d from <i>S. pyogenes</i> (3VZH)	26.5	1.4 (186/194)	37
Cas5d from <i>X. oryzae</i> (3VZI)	22.1	2.1 (181/214)	31
Cas5 from <i>D. vulgaris</i> (7KHA)	20	2.2 (181/220)	26

Table 2. Structural similarity search using DALI²⁵.

represent the same dimeric form, we analyzed the BD dimer in detail, as a representative. For the BD dimer, the total dimer surface buried an area of 716.1 Å², representing 6.2% of the total surface area. Twenty residues, representing 9.8% of the total MboCas5 residues, were involved in PPI formation (Fig. 2B). PPI was formed by 12 hydrogen bonds (H-bonds) and four salt bridges (Fig. 2C and D). The salt bridges formed by residues R72 and D167 from each molecule are the primary forces that maintain the integrity of the dimeric interface. Additionally, residues Q38, T73, Q168, and D170 from each monomer were found to be involved in the formation of H-bonds.

To validate the predicted dimeric form, we generated two mutants that were expected to disrupt PPI. In one mutant, namely Q38W, Q38 (a residue involved in hydrogen bond formation at the center of the PPI interface) was replaced with tryptophan. In the other mutant, R72W, R72 (a residue that forms a salt bridge and serves as a major force driving PPI formation) was replaced by tryptophan. The two mutants and wild-type MboCas5 were subjected to SEC to analyze the dimer formation. SEC analysis showed that dimer formation was disrupted in the R72W variant, as evidenced by delayed elution from SEC, corresponding to the size of the monomer (Fig. 2E). The absolute molecular weight of the R72W mutant was determined using MALS. The calculated weight of 34.4 kDa (3.5% fitting error) confirmed that the R72W mutation resulted in an MboCas5 monomer (Fig. 2F and G). Based on the structure analysis, crystal packing, and mutagenesis studies, followed by a dimer disruption assay, we concluded that MboCas5 exists as a BD dimer in solution. The Q38W mutant remained a dimer, suggesting that the disruption of the H-bond mediated by Q38 alone was not sufficient to destabilize the overall dimeric structure (Fig. 2E and G). To further validate our dimer structure, structural information was used to predict its dimeric form using AF3. The predicted model closely resembled the experimentally determined dimer structure (Fig. 2H and I). Taken together, we confirmed that the dimer indeed represents the authentic dimeric form of Cas5.

Thus far, the only Cas5 dimer structure proposed based on structural data is SmuCas5 (Fig. 2J)¹⁸. When we compared the structure of SmuCas5 dimer with that of our MboCas5 dimer, we observed substantial differences, particularly in the orientation of the dimerization interface (Fig. 2C and J). To explore a more plausible dimeric form of SmuCas5, we used AF3 to predict the dimer structure. The predicted SmuCas5 dimer clearly differed from the previously published SmuCas5 dimer, since it formed through an alternative set of protein interfaces (Fig. 2K). Superimposition of the predicted SmuCas5 dimer onto our MboCas5 dimer revealed that both utilized similar interfaces for dimerization (Fig. 2L). Therefore, together with our structural, mutational, and modeling studies, we concluded that the identified Cas5 dimer structure likely represents a physiologically relevant dimer form. Furthermore, we proposed that the dimerization mode may be applicable to other Cas5 proteins.

Structural comparison of MboCas5 with Cas5 homologs from other species

Till date, Cas5 structure has been determined and reported in five different bacterial species^{14,19,20}. To perform a comparative structural analysis of the Cas5 proteins and MboCas5, we used the DALI server to search for structurally similar proteins (Fig. 3A). Cas5 from *A. halodurans* (AhaCas5)¹⁴, which shares the highest sequence identity (49%) with MboCas5, was the most structurally similar homolog. This was followed by *S. pyogenes* Cas5 (SpyCas5)²⁰, *X. oryzae* Cas5 (XorCas5)²⁰, and *D. vulgaris* Cas5 (DvuCas5)¹⁹ in order. The Z-scores ranged from approximately 20 to 27 and the RMSD values were between 1.5 and 2.2 Å. When we superimposed all five Cas5 structures, including MboCas5, we found that the overall fold of the main domain was highly conserved (Fig. 3B). However, notable structural differences were observed in the region corresponding to the unstructured loop of MboCas5. Specifically, the region is composed of two long helices in XorCas5, whereas it adopts an extended loop conformation in DvuCas5, clearly distinguishing it from MboCas5 (Fig. 3B). Pairwise structural alignments revealed that the unstructured loop region of MboCas5 remained unmodelled in both AhaCas5 and SpyCas5, indicating this region to likely be intrinsically disordered (Fig. 3C). In contrast, the corresponding regions in XorCas5 and DvuCas5 formed well-defined structural elements, such as extended helices in XorCas5 and a long loop in DvuCas5 (Fig. 3C).

Among the bacterial Cas5 proteins, SmuCas5 is one of the most extensively studied in terms of its catalytic activity and active site characterization¹⁸. In SmuCas5, the active site is composed of a catalytic triad formed by Y50, K120, and H121, and an RNA substrate-binding site formed by K52 and R134. Sequence alignment revealed these residues to be conserved in MboCas5 (Fig. 3D), suggesting that MboCas5 likely forms a similar active site and performs a comparable function (Fig. 3E). However, in XorCas5 and DvuCas5, some of the

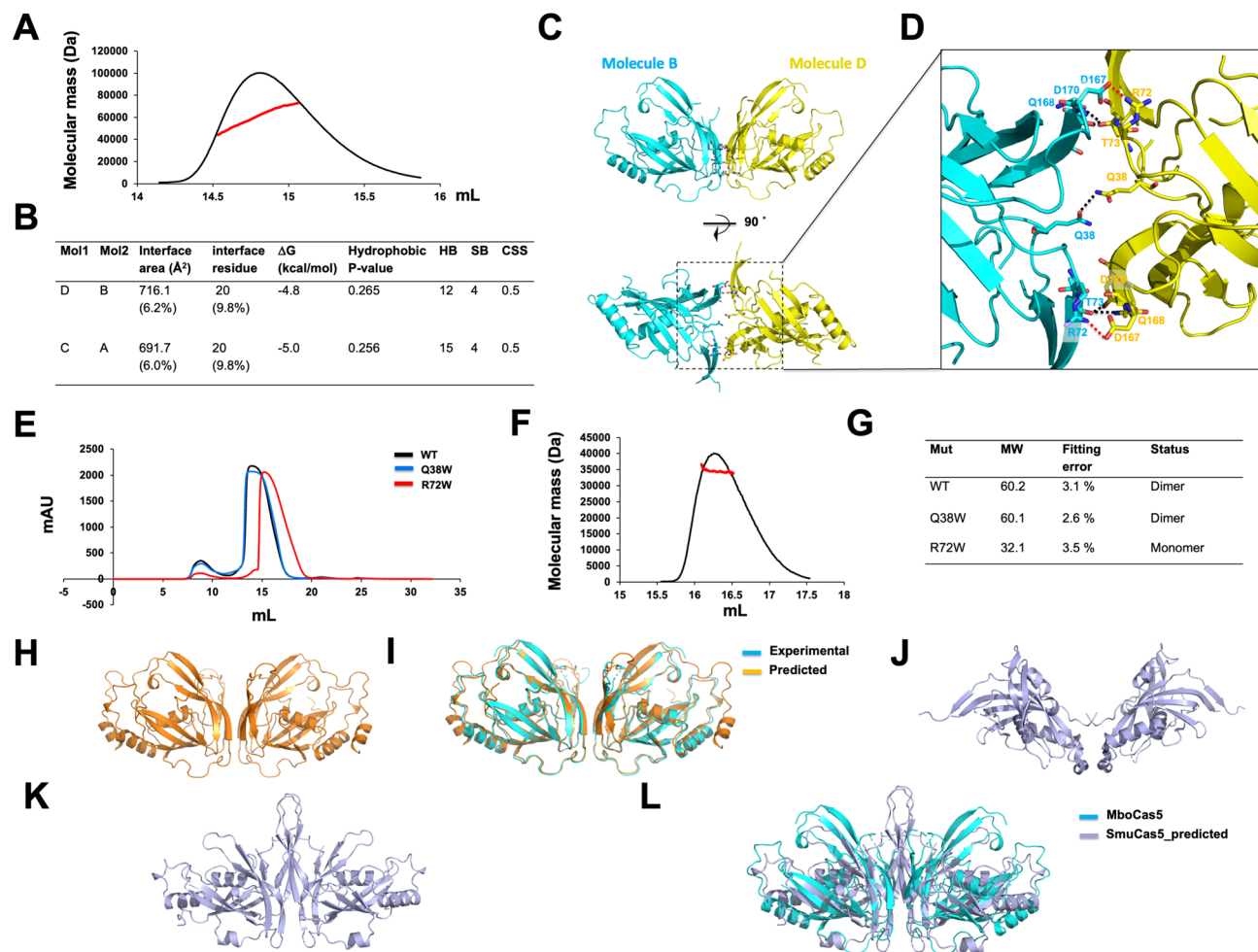


Fig. 2. Unique dimeric structure of MboCas5 and its interface analysis. (A) Profile of MALS analysis. The experimental MALS data (red line) plotted on the SEC chromatogram (black) at 280 nm, with the x-axis representing the SEC elution volume and the y-axis representing the absolute molecular mass. (B) Table summarizing the result of a PISA analysis. (C) Predicted dimeric structure of MboCas5. The model shows a dimer formed by MolB and MolD. (D) Close-up view of the MolB/MolD dimer PPI interface. Key residues contributing to the PPI are labeled. Salt bridges and hydrogen bonds are indicated by red dotted lines and black dotted lines, respectively. (E) SEC profiles of the R72W mutant, which is predicted to be important for dimer PPI formation, and the Q38W mutant, which is expected to be unrelated to PPI formation. (F) MALS profile of the R72W mutant. (G) Table summarizing the MALS results of WT, R72W, and Q38W. (H) Cartoon representation of the MboCas5 dimer structure predicted by AF3. (I) Cartoon representation showing the superposition of the MboCas5 dimer structure predicted by AlphaFold 3 (orange) and the experimentally determined dimer structure revealed in this study (cyan). (J) Previously reported dimeric structure of Cas5. Among all the Cas5 proteins characterized till date, a dimeric structure has been reported for SmuCas5 (PDB ID: 4R0J). (K) Cartoon representation of the SmuCas5 dimer structure predicted by AF3. (L) Cartoon representation showing the superposition of the SmuCas5 dimer structure predicted by AF3 (light blue) and the experimentally determined dimer structure revealed in this study (cyan).

residues critical for catalytic activity were only partially conserved, suggesting that there may be differences in their enzymatic activities and substrate specificities.

Structural adjustment of Cas proteins in cascade for secure complex formation

Among the known Cas5 structures, DvuCas5 is unique in that it has only been observed as part of the Cascade complex, rather than as a standalone protein¹⁹. In the structure of the type I-C Cascade, Cas5 interacts with Cas8 and crRNA, and a long loop—referred to as the “thumb motif” extends outward to engage with Cas7 (Fig. 4A and B). Interestingly, the long-loop region corresponds to an unstructured or unresolved region in Cas5 from other species, including MboCas5. In XorCas5, although the loop is present, it extends in a completely different direction (Fig. 4C). The observations suggested that the region is highly flexible when Cas5 exists in isolation, allowing it to adopt multiple conformations. Upon Cascade assembly, the region likely undergoes a conformational rearrangement to tightly interact with Cas8, thereby stabilizing the complex.

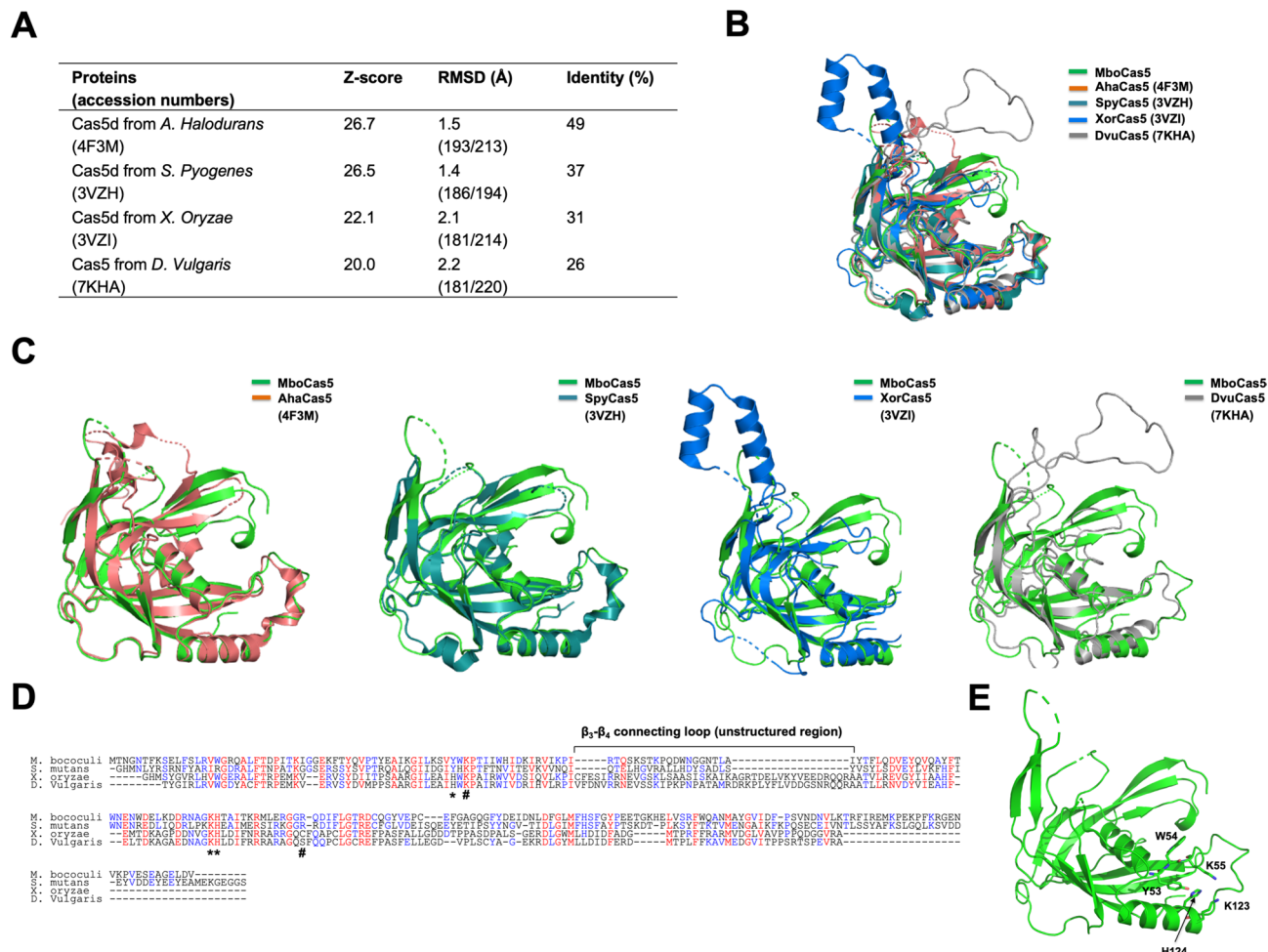


Fig. 3. Structure comparison with Cas5 homologs from other species. (A) Table summarizing the results of the DALI search. (B) Cartoon representation showing the structural superposition of the top four hits identified by the DALI search. The comparison highlights the overall structural similarity among the predicted homologs. (C) Cartoon representations of pairwise structural superpositions between MboCas5 and each of the top four hits identified by the DALI search, illustrating the degree of structural similarity for each comparison. (D) Multiple sequence alignment of MboCas5 with Cas5 proteins from three other species in which functional studies have been performed. The residues highlighted by black boxes represent those previously reported to be important for Cas5 activity. (E) Cartoon representation of MboCas5 showing the locations of residues that correspond to those previously identified as important for Cas5 activity in other species.

Such a conformational adaptation for secure assembly is not unique to Cas5. Recent studies have reported similar structural changes in other Cascade components, including Cas3 and Cas8¹¹. Cas3, for example, undergoes a significant conformational change upon recruitment to the Cascade complex, allowing for a deeper and tighter interaction with Cas8 (Fig. 4D, E, and F). Similarly, Cas8 itself undergoes structural rearrangement to accommodate Cas3 (Fig. 4E). Our findings supported the idea that Cascade complex formation involves evolutionarily optimized conformational shifts across multiple proteins to ensure stable and functionally precise assembly.

Comparison of CrRNA processing activity between monomeric and dimeric forms of MboCas5

Within the Cascade complex, Cas5 typically exists as a monomer; however, several Cas5 proteins from different species have been reported to form dimers when expressed alone, and these dimers are known to retain crRNA processing activity even in the absence of other Cascade components. In our study, MboCas5 was found to form a dimer when expressed independently, and this dimerization was disrupted by the R72W mutation. Does dimerization play a critical role in Cas5 function? To address this question, we compared the crRNA processing activities of the wild-type (dimeric) and R72W mutant (monomeric) forms of MboCas5. In this activity assay, we used only the repeat region as the substrate, an approach that is well-established and widely used by other researchers, to ensure higher RNA stability and achieve more accurate cleavage analysis¹⁴. The experimental

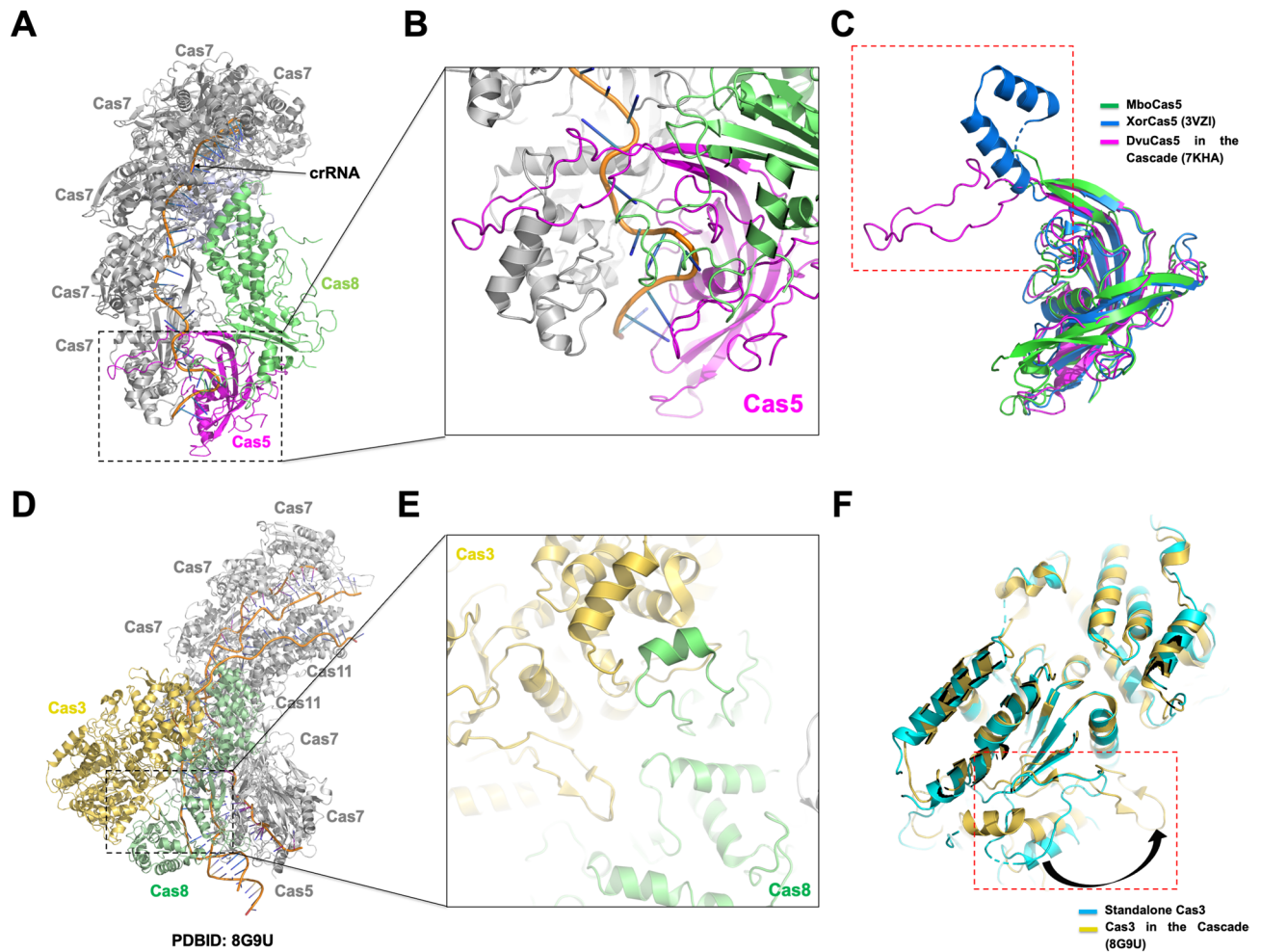


Fig. 4. Structural adjustment of Cas proteins in Cascade for secure complex formation. (A) Structure of type I-C Cascade from *Desulfovibrio vulgaris* (PDB ID: 7KHA). (B) Magnified view of the dashed box in panel (a), showing the interaction between Cas5 and Cas7 in the Cascade complex. (C) Cartoon representation of the superposition of three Cas5 structures. Regions with significant structural variations are highlighted with red dashed boxes. (D) Complex structure of type I-C Cascade and Cas3 from *Neisseria lactamica* (PDB ID: 8G9U). (E) Magnified view of the dashed box in panel (d), showing the interaction between Cas3 and Cas8 in the Cascade complex. (F) Cartoon representation of the superposition of two Cas3 structures. Regions with significant structural variations are highlighted with red dashed boxes.

results demonstrated that both forms exhibited similar levels of crRNA processing activity (Fig. 5A and B). These findings indicate that Cas5 possesses intrinsic crRNA processing capability in both its monomeric and dimeric states.

The next question was whether the dimeric form of Cas5 can exist within the Cascade complex. To address this, we performed a structural analysis by superimposing one molecule of our dimeric MboCas5 onto the monomeric Cas5 in the reported type I-C Cascade structure. We then examined whether the second molecule of the dimer could fit within the Cascade architecture without steric clashes. The analysis revealed that the second molecule of the MboCas5 dimer causes a clear steric collision with Cas8, indicating that the dimeric form of Cas5 cannot be accommodated within the type I-C Cascade complex (Fig. 5C and D).

Discussion

In this study, we present the high-resolution crystal structure of Cas5 from *Moraxella bovoculi* (MboCas5), a core component of the type I-C CRISPR-Cas system, and provide biochemical and mutational evidence supporting its unique dimeric assembly. Our findings reveal that MboCas5 adopts a conserved Cas5 fold yet displays distinctive structural elements, particularly a flexible loop region and a novel dimerization interface, which likely underpin its functional role in Cascade assembly and crRNA handling.

One of the most striking observations is the stable unique dimer formation of MboCas5 in solution, confirmed by SEC-MALS and mutagenesis. Although Cas5 proteins from certain bacterial species, such as *Streptococcus mutans*, have been reported to form dimers¹⁸, previous structural interpretations of their dimerization interfaces differ substantially from our findings. By combining crystallographic analysis, PISA-based interface scoring,

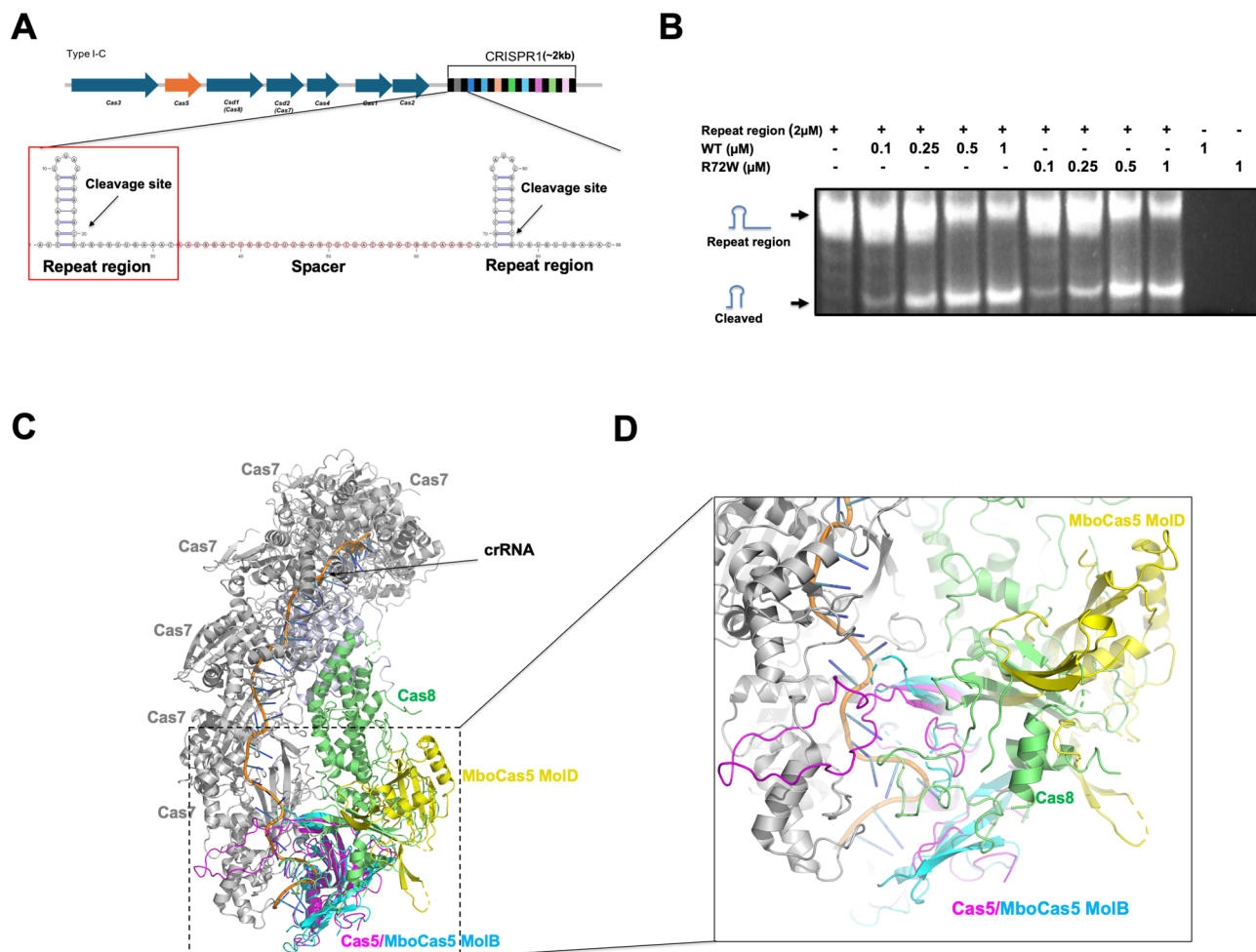


Fig. 5. Comparison of crRNA processing activity between monomeric and dimeric forms of MboCas5, and structural analysis of Cas5 forms within the Cascade complex. (A) The sequence and predicted secondary structure of the designed substrate for MboCas5. The region used as the substrate is highlighted with a red box. The cleavage site predicted to be recognized by Cas5 is indicated with an arrow. (B) Comparison of the crRNA processing activity between the wild-type (WT; dimeric) and the R72W mutant (monomeric) forms of MboCas5. The positions of the uncleaved pre-crRNA and the cleaved crRNA products are indicated. (C) Structural superimposition of Cas5 in the type I-C Cascade complex with the dimeric MboCas5 to examine the possible spatial position of the Cas5 dimer within the Cascade architecture. (D) Enlarged view of the boxed region indicated by the dashed lines in panel (C).

targeted mutagenesis, and AF 3 modeling, we demonstrate that the MboCas5 dimer interface is stabilized primarily by salt bridges between R72 and D167, supported by a network of hydrogen bonds. Disruption of the R72-mediated salt bridge abolished dimerization, underscoring its critical role. This mechanism appears to be conserved *in silico* across other Cas5 orthologs, suggesting a potential universal dimerization mode among type I-C Cas5 proteins. Given that stoichiometry and oligomerization state can significantly influence Cascade assembly, our structural insights may help explain subtype-specific differences in complex stability and crRNA loading.

Structural comparisons with Cas5 homologs from diverse CRISPR-Cas subtypes highlight both conserved and variable features. The catalytic triad residues (Y50, K120, H121) and RNA-binding residues (K52, R134) are preserved in MboCas5, implying conservation of crRNA binding and processing potential. In contrast, the unstructured loop region in MboCas5 corresponds to well-formed helices or extended loops in some homologs (e.g., *Xanthomonas oryzae* Cas5), indicating structural plasticity in this region. Such flexibility likely facilitates the substantial conformational rearrangements required during Cascade assembly, as seen in *Desulfovibrio vulgaris* Cas5 when bound to Cas8 and crRNA. This adaptability may enable MboCas5 to accommodate the minimalistic protein composition of type I-C Cascade while maintaining complex stability.

Our observations also align with a broader paradigm in CRISPR-Cas biology: conformational adaptability of multiple Cascade components to achieve precise, high-affinity assembly. Similar to the loop rearrangement in Cas5, structural changes in Cas3 and Cas8 upon complex formation have been reported to optimize protein-

protein interactions¹¹. These dynamic adjustments likely represent an evolutionary strategy to ensure robust target recognition and degradation in the face of genomic diversity among invading elements.

Beyond structural characterization, our results have potential implications for the engineering of CRISPR-Cas systems. The identification of a structurally and functionally critical dimerization interface opens opportunities for rational modulation of Cascade assembly by targeted mutagenesis or small-molecule binding. Such approaches could be used to alter complex stability, crRNA binding dynamics, or interference activity, potentially broadening the biotechnological applicability of type I-C systems. Nevertheless, our structural model represents MboCas5 in isolation, without bound crRNA or partner proteins. It remains to be determined whether the unique dimer revealed in this study persists within the assembled Cascade or undergoes rearrangement upon complex formation. Furthermore, while AF3 predictions suggest conservation of the identified interface in other Cas5 proteins, experimental validation across species is needed to confirm its generality. High-resolution structural studies of complete type I-C Cascade complexes containing MboCas5, ideally in multiple functional states, will be essential to fully elucidate the mechanistic relevance of our observations.

The Cas5c family can be divided into two subgroups, A and B, based on their amino acid sequence composition^{18,20}. Members of subgroup A, such as BhaCas5 and SmuCas5, possess an additional ~ 30 amino acids at the C-terminal region. In contrast, subgroup B members, including XorCas5, lack this C-terminal extension but contain an insertion of approximately 20 residues between the $\beta 3$ and $\beta 4$ strands. MboCas5 contains an insertion of approximately 30 residues at the C-terminus and lacks about 23 residues between $\beta 3$ and $\beta 4$. These characteristics indicate that MboCas5 belongs to the Cas5c subgroup within subgroup A.

In conclusion, this work provides structural and mechanistic insights into MboCas5, defining its unique dimeric assembly and conformational flexibility as key features likely contributing to type I-C Cascade function. This work also revealed that Cas5 can exist as either a monomer or a dimer in isolation, and both forms possess crRNA-processing activity. However, within the Cascade complex, Cas5 is incorporated exclusively as a monomer, consistent with its structural compatibility and functional role in the complex assembly. By integrating crystallography, biophysics, mutagenesis, and comparative analysis, we establish a framework for understanding Cas5-mediated assembly principles that may extend to other CRISPR-Cas systems. These findings advance our molecular understanding of Cascade architecture and may inform future efforts to harness type I-C systems for genome engineering applications.

Methods

Cloning, overexpression, and purification

The expression plasmid for full-length Cas5 from *M. bovoculi*, corresponding to amino acids 1–245 (MboCas5), was constructed by inserting the synthesized gene product into a pET21a vector. The gene sequence was obtained from GenBank (Accession ID: WP_036362185), and gene synthesis was conducted using BIONICS (Seoul, Republic of Korea). The synthesized gene was digested using the restriction enzymes *NdeI* and *XhoI* and inserted into a pET21a plasmid vector (Novagen, Madison, WI, USA).

The expression plasmid encoding the gene was transformed into *Escherichia coli* BL21 (DE3) competent cells using heat shock at 42 °C. Subsequently, the cells were spread onto plates containing lysogeny broth (LB) agar with 75 μ g/mL ampicillin and incubated overnight at 37 °C. A single colony was selected and cultured in 5–10 mL of LB containing 100 μ g/mL ampicillin overnight at 37 °C, after which the cells were inoculated and cultured on 1 L of LB containing 100 μ g/mL ampicillin. When the optical density at 600 nm reached approximately 0.8, the cultured cells were cooled in ice, and 0.25 mM isopropyl β -D-1-thiogalactopyranoside (IPTG) was added to induce the expression of the MboCas5 gene. The cells were further cultured for 18 h at 20 °C.

The cultured cells expressing MboCas5 were collected by centrifugation at 2000 $\times$ g for 15 min at 20 °C, resuspended in 20 mL of lysis buffer (20 mM Tris-HCl, pH 8.0, and 500 mM NaCl). The cells were then lysed by ultrasonication on ice. The cell lysates were centrifuged at 10,000 $\times$ g for 30 min at 4 °C to remove the cell debris, and the supernatant was collected and incubated with nickel-nitrilotriacetic acid (Ni-NTA) affinity resins (Qiagen, Hilden, Germany) for 2 h.

After incubation, the supernatant was applied to a gravity-flow column (Bio-Rad, Hercules, CA, USA), and the resin was washed with 40 mL of lysis buffer to remove all unbound proteins. Next, 0.65 mL of 5 $\times$ elution buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, and 250 mM imidazole) was added to the column to elute the target protein. The eluted protein from the affinity resin was loaded onto a Superdex 200 Increase 10/300 GL column (GE Healthcare, Waukesha, WI, USA) connected to an ÄKTA Pure system (GE Healthcare), which was pre-equilibrated with the SEC buffer (20 mM Tris-HCl, pH 8.0, and 150 mM NaCl). The eluted peak fractions from SEC were collected, pooled, and concentrated to 6.8 mg/mL, flash-frozen in liquid nitrogen, and stored at –80 °C until use. Sodium dodecyl sulphate-polyacrylamide gel electrophoresis was used to assess the purity of the protein.

Crystallization and X-ray diffraction data collection

Crystallization of MboCas5 was performed by the hanging drop vapor diffusion method at 20 °C. The optimized crystals were acquired by equilibrating a mixed drop containing 1 μ L of protein solution (6.8 mg/mL of protein in SEC buffer) and 1 μ L of reservoir solution containing 10% (w/v) polyethylene glycol (PEG) 8000, 0.1 M HEPES pH 7.5, and 8% (v/v) ethylene glycol against 500 μ L of reservoir solution. Crystals were formed within 25 days. Diffraction data were collected using the BL-5 C beamline of the Pohang Accelerator Laboratory (Pohang, Republic of Korea). Data were processed using HKL2000²¹.

Structure determination and analysis

The structure of MboCas5 was determined using the MR method with the Phaser program²² of the PHENIX Suite²³. The MR search model used was the AlphaFold3-predicted MboCas5 model¹⁶. Model building and refinement were performed using Coot²⁴ and phenix.refine, respectively, in the PHENIX suite²³.

SEC-MALS analysis

The molecular weight of MboCas5 in solution was determined using SEC coupled with multi-angle light scattering (SEC-MALS). The purified protein was loaded onto a Superdex 200 Increase 10/300 GL column (GE Healthcare) connected to an ÅKTA FPLC system (GE Healthcare). The system was coupled with a mini-DAWN TREOS MALS detector (Wyatt Technology, Santa Barbara, CA, USA). Data were collected at a flow rate of 0.5 mL/min with measurements recorded every 0.5 s and processed using the ASTRA software (Wyatt Technology). The molecular weight of bovine serum albumin was used as the reference.

Mutagenesis

Site-directed mutagenesis was performed using a QuickChange kit (Stratagene, San Diego, CA, USA) according to the manufacturer's instructions. The introduced mutations were verified using sequencing. Successfully validated mutant proteins were expressed and purified following the same protocol used for the wild-type protein.

Structural homology search using the DALI server

Structurally related proteins were identified using the DALI server²⁵. A PDB file containing only molecule A of MboCas5 was submitted to the PDB search panel with default search parameters.

Pre-crRNA cleavage assay

CRISPR repeat RNAs, corresponding to the CRISPR array identified in the *M. bovoculi* genome, were synthesized using the HiScribe[®] T7 Quick High Yield RNA Synthesis Kit (New England Biolabs, USA) from double-stranded DNA templates. The synthesized RNAs were purified with the Monarch[®] RNA Cleanup Kit (New England Biolabs, USA). DNA templates containing a 5' T7 promoter sequence for RNA synthesis were synthesized by Bionics (Seoul, Republic of Korea).

CRISPR repeat RNAs (2.0 µM) were incubated with wild-type or mutant MboCas5 proteins (0.1, 0.25, 0.5, and 1.0 µM) in reaction buffer containing 20 mM HEPES (pH 7.5) and 150 mM KCl at room temperature for 25 min. Reaction mixtures were resolved on a 14% TBE polyacrylamide gel (acrylamide/bis-acrylamide, 37.5:1) and visualized by SYBR Gold staining (Invitrogen, USA).

Data availability

All data generated or analysed during this study are included in this published article. Uncropped original gels are provided at the Supplementary Figure 1.

Received: 27 August 2025; Accepted: 12 December 2025

Published online: 15 December 2025

References

- Sorek, R., Kunin, V. & Hugenholtz, P. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nat. Rev. Microbiol.* **6**, 181–186 (2008).
- Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712 (2007).
- Mojica, F. J. & Rodríguez-Valera, F. The discovery of CRISPR in archaea and bacteria. *FEBS J.* **283**, 3162–3169 (2016).
- Barrangou, R. The roles of CRISPR-Cas systems in adaptive immunity and beyond. *Curr. Opin. Immunol.* **32**, 36–41 (2015).
- Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
- Marraffini, L. A. CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61 (2015).
- Brouns, S. J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
- Wiegand, T., Karambelkar, S., Bondy-Denomy, J. & Wiedenheft, B. Structures and strategies of Anti-CRISPR-Mediated immune suppression. *Annu. Rev. Microbiol.* **74**, 21–37 (2020).
- Liu, T. Y. & Doudna, J. A. Chemistry of class 1 CRISPR-Cas effectors: Binding, editing, and regulation. *J. Biol. Chem.* **295**, 14473–14487 (2020).
- Hidalgo-Cantabrana, C. & Barrangou, R. Characterization and applications of type I CRISPR-Cas systems. *Biochem. Soc. Trans.* **48**, 15–23 (2020).
- Kim, D. Y., Lee, S. Y., Ha, H. J. & Park, H. H. Structural basis of Cas3 activation in type I-C CRISPR-Cas system. *Nucleic Acids Res.* **52**, 10563–10574 (2024).
- Csorgo, B. et al. A compact Cascade-Cas3 system for targeted genome engineering. *Nat. Methods.* **17**, 1183– (2020).
- Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by cascade and Cas3. *Science* **361**, eaat839 (2018).
- Nam, K. H. et al. Cas5d protein processes pre-crRNA and assembles into a cascade-like interference complex in subtype I-C/Dvulg CRISPR-Cas system. *Structure* **20**, 1574–1584 (2012).
- Zhao, H. et al. Crystal structure of the RNA-guided immune surveillance cascade complex in *Escherichia coli*. *Nature* **515**, 147–150 (2014).
- Jumper, J. et al. Highly accurate protein structure prediction with alphafold. *Nature* **596**, 583–589 (2021).
- Mulepati, S. & Bailey, S. In vitro reconstitution of an *Escherichia coli* RNA-guided immune system reveals unidirectional, ATP-dependent degradation of DNA target. *J. Biol. Chem.* **288**, 22184–22192 (2013).
- Lemak, S. et al. Structural and biochemical insights into CRISPR RNA processing by the Cas5c ribonuclease SMU1763 from *Streptococcus mutans*. *J. Biol. Chem.* **297**, 101251 (2021).
- O'Brien, R. E. et al. Structural basis for assembly of non-canonical small subunits into type I-C cascade. *Nat. Commun.* **11**, 5931 (2020).

20. Koo, Y., Ka, D., Kim, E. J., Suh, N. & Bae, E. Conservation and variability in the structure and function of the Cas5d endoribonuclease in the CRISPR-mediated microbial immune system. *J. Mol. Biol.* **425**, 3799–3810 (2013).
21. Otwinowski, Z. *DENZO Data Processing Package* (Yale University, C.T., 1990).
22. McCoy, A. J. Solving structures of protein complexes by molecular replacement with phaser. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 32–41 (2007).
23. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
24. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
25. Holm, L. & Sander, C. Dali: a network tool for protein structure comparison. *Trends Biochem. Sci.* **20**, 478–480 (1995).

Acknowledgements

We express our gratitude to the beamline staff at PLS-5 C of the Pohang Light Source (Pohang, Korea) for their assistance with X-ray diffraction data collection.

Author contributions

H.H.P. designed and supervised the project. Y.J.K. and H.J.H. performed cloning, expression, and protein purification. Y.J.K. and S.Y.L. crystallized the proteins, collected X-ray data, and solved the protein structure. H.B.J. performed SEC-MALS experiments. Y.J.K. and H.J.H. wrote the initial manuscript, and H.H.P. edited the manuscript. All authors discussed the results, commented on the manuscript, and approved the final version.

Funding

This study was supported by the National Research Foundation of Korea (NRF), funded by the Korean government (MSIT) [grant number RS-2025-02316334], and by the Chung-Ang University Graduate Research Scholarship in 2025.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-32766-5>.

Correspondence and requests for materials should be addressed to H.H.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025