A Dissertation for the Degree of Master of Science

Understanding the nuclease activity of *Xanthomonas albilineans* Cas2 via its solution structure and dynamics

Xanthomonas albilineans Cas2 단백질의
수용액 구조 및 다이나믹스 분석을 통한
핵산분해효소 기능 이해

February, 2018

Seoul National University

Biomodulation Major

Department of Agricultural Biotechnology

Migyeong Jeong

Understanding the nuclease activity of *Xanthomonas albilineans* Cas2 via its solution structure and dynamics

Xanthomonas albilineans Cas2 단백질의
수용액 구조 및 다이나믹스 분석을 통한
핵분해효소 기능 이해

지도교수 서 정 용

이 논문을 이학 석사 학위논문으로 제출함

2018년 2월

서울대학교 대학원

농생명공학부 바이오모듈레이션 전공

정 미 경

정 미 경의 석사학위논문을 인준함

2018년 1월

위 원 장 석 차 옥 (인)

부 위 원 장 서 정 용 (인)

위 원 배 의 영 (인)

Abstract

Understanding the nuclease activity of *Xanthomonas albilineans* Cas2 via its solution structure and dynamics

Migyeong Jeong
Biomodulation Major
Department of Agricultural Biotechnology
The Graduate School
Seoul National University

The CRISPR-Cas system confers prokaryotes an adaptive immunity. The system integrates a small genetic fragment from invader into the host CRISR-locus as a memory of infection. Distinctively, each types of the system carries different organizations of CRISPR-associated proteins (Cas), which are featured by their roles in certain steps of the immunity - acquisition, expression, and interference. Among various Cas proteins, Cas2 is the most conserved protein that functions as a scaffold cooperating with Cas1 during the acquisition step. Although studies have demonstrated the intrinsic metal-dependent nuclease activity of Cas2, the unnecessary role of its nuclease activity for proper acquisition process arises questions to the possession and mechanisms of its activity. Moreover, especially in type I-C, all of known Cas2 crystal structures are catalytically inactive conformational states, which makes even more ambiguous to understand the procedures for its nuclease activity to take place. Thus, in this study, we have investigated the solution state of *Xanthomonas albilineans* Cas2 (*Xa*Cas2) to

provide direct mechanistic evidences for its nuclease activity. Based on NMR experiments, residual dipolar couplings (RDC) analysis have resulted that *Xa*Cas2 adopts an inactive conformation in solution; and, the NMR relaxation data indicates that the active site and hinge regions possess highly dynamical motion as for the driving force of conformational change. In conclusion, the results suggest that *Xa*Cas2 is thermodynamically stable as inactive conformation in solution, and the forces from highly dynamic regions allow catalytically active conformation when encounter genetic substrates and preferred metal ions. Taken together, the intrinsic function of Cas2 can be explained by a dynamic equilibrium of conformational states that serves as a scaffold and as a nuclease on demand.

Keywords : Cas2 | nuclease activity | structure and dynamics | NMR

Student Number : 2015-22696

Table of Contents

List of Figures

List of Tables

List of Abbreviations

2D or 3D
Two or three dimensional

CRISPR
Clustered regularly interspaced palindromic repeat

DNA
Deoxyribonucleic acid

LB
Luria Bertini

M9
Minimal 9

NMR
Nuclear magnetic resonance

NOE
Nuclear Overhauser effect

OD
Optical density

pET
Plasmid for expression using T7 promoter

RNA
Ribonucleic acid

TEV
Tobacco Etch virus

w/v
Weight per volume

1. Introduction

1.1 The prokaryotic adaptive immune system, CRISPR-Cas system

Most of prokaryotes carry one or more of CRISPR-Cas (Clustered regularly interspaced short palindromic repeat-CRISPR-associated proteins) system in their genome which confers memories of previously encountered genetic materials [1–5]. The system is defined by presence of two major components, a CRISPR-locus and CRISPR-associated proteins (Cas) genes in the genome. The CRISPR-locus is featured by a number of palindromic repeat sequences that are regularly separated by variable sequences, called spacers. The transcription and translation of Cas genes generates functional Cas proteins that cooperate together with different roles in certain steps of the system as they were assigned to. The CRISPR-Cas system undergoes three distinctive steps, acquisition, expression, and interference, to establish the adaptive immunity. During the acquisition step, integration of exogeneous genetic elements facilitates the system to process the integrated genes into short segments called protospacers. Then, during the expression step, the incorporated spacers are transcribed into CRISPR-RNA (crRNA)s and interact with expressed Cas proteins to form a ribonucleoprotein complex as an effector complex. By the guidance of crRNA, the effector complex recognizes the same or similar genes by complementary base-pairs leading to degradation during the interference steps in the system [6–10] (Fig. 1A). Likewise, the CRISPR-Cas systems establish memories of invaders for instant facilitation of the defense system upon subsequent invasions.

1.2 The functional and structural characteristics of Cas2

The CRISPR-Cas system is categorized into two classes, six types and thirty-three subtypes according to different organizations of signature *cas* genes [11]. Cas1 and Cas2 are common in most CRISPR-Cas systems, and play an important role in the initiation of the adaptive immune system [12]. In *Escherichia coli*, Cas1 and Cas2 are essential elements for capturing foreign DNA segments, where Cas2 functions as a scaffold protein to control the length of the spacers [13,14]. Studies have been demonstrated that Cas2 exhibits nuclease activities against double-stranded DNA (dsDNA) and single-stranded RNA (ssRNA) in a divalent metal ion- and pH-dependent manner. Cas2 from *Sulfolobus solfataricus* exhibited a nuclease activity against ssRNA substrates [15], whereas Cas2 from *Bacillus halodurans* [16] and *Streptococcus pyogenes* [17] showed a nuclease activity against dsDNA substrates. To note, Cas2 from *Desulfovibrio vulgaris* [18] and *Thermococcus onnurineus* [19] did not show apparent nuclease activities, implying not all of Cas2 appear to have the nuclease functions which might be replaced by other components. Although Cas2 possesses nuclease activity, its nuclease activity was not essential for proper function of acquisition stage where Cas1-Cas2 complex integrates a new spacer segment into CRISPR-locus to establish immune memory [13]. Thus, the mechanism of Cas2 changing its state from functionally active to inactive reversibly is questioned.

Up-to-date, all defined crystal structures of Cas2s from several bacterial strains present a common ferredoxin fold of a homodimer with invariably catalytically inactive conformations that do not explain the nuclease activity [15–19]. It has been reported that a Cas2 dimer binds to a single divalent metal ion for the nuclease activity, and that highly conserved Asp in β1 strand is critical for the metal binding [16] (Fig. 8). The negatively charged residues, such as the Asp residues at the dimer interface are separated by 6–15 Å between subunits,

which most of them do not allow the Cas2 dimer to accommodate a single metal ion.

A recent study on the type I-C Cas2 from *Xanthomonas albilinean* has been postulated that Cas2 undergoes conformational changes to catalyze the nuclease reaction via C-terminal dynamical motion by demonstrating changes in the UV-fluorescence intensity of this area depend on different pH-environment, which are correlated to its functionally active conditions. Comparative analysis of crystal structures has predicted hinge motions in the β4–β5 loop region of Cas2 for the nuclease activity, but experimental evidences have been limited [16,17,20].

Here, we have investigated the solution structure and dynamics of Cas2 from gram-negative bacteria *Xanthomonas albilineans*, a main plant pathogen of leaf scald diseases [21,22]. *Xanthomonas albilineans* Cas2 (*Xa*Cas2) belongs to the subtype I-C CRISPR-Cas system (Fig. 1B), and the previous crystal structure presented a catalytically inactive state [20]. We demonstrate that *Xa*Cas2 adopts an inactive conformation in solution, but exhibits significant conformational dynamics at the hinge region and the metal binding site. The dynamic nature at the active site supports the conformational switch mechanism of Cas2 to function as a scaffold protein as well as a nuclease.

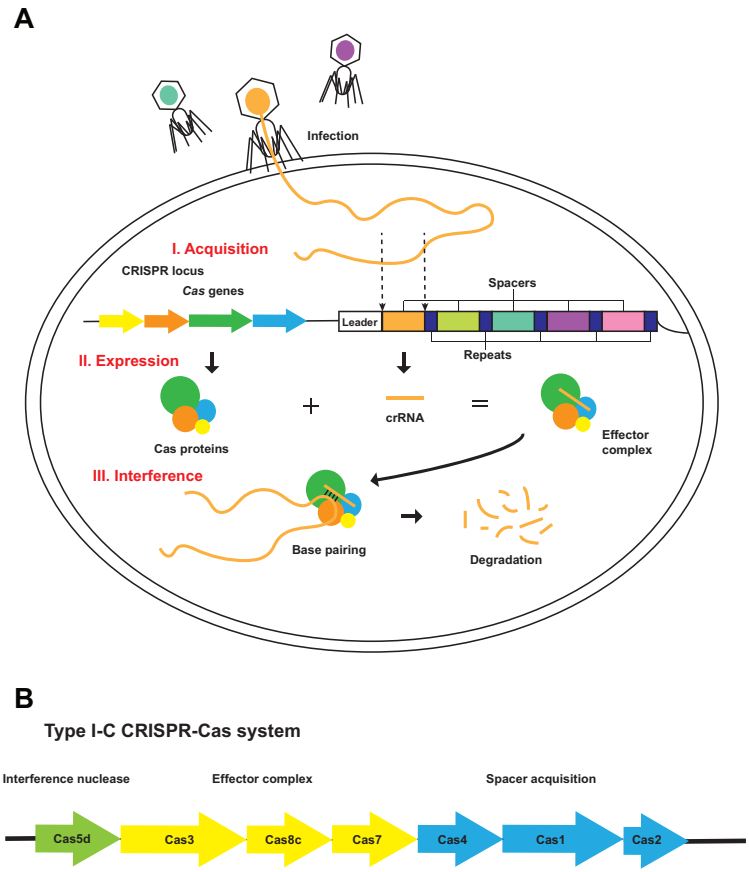

Figure 1. Overview of CRISPR-Cas system and gene organization of type I-C CRISPR-Cas system. (A) The three steps of CRISPR-Cas system : I. Acquisition, II. Expression, and III. Interference are described. Upon infection, the inserted foreign gene (bacteriophage) is incorporated into the CRISPR-locus of host (bacteria) genome during acquisition step. Then, during expression step, the inserted gene, a spacer is further transcribed into crRNA to form an effector complex with Cas proteins. Finally, during interference step, the active effector complex, base pairs with the invasive gene as a result of the degradation of the invasive gene. (B) Cas gene organization of type I-C CRISPR-Cas system. The Cas gene contents of type I-C system is described as arrows. (From right to left) The Cas proteins that are involved in spacer acquisition in the system are colored as blue, in expression stage to form an effector complex are colored as yellow, and in interference state as a nuclease is colored as green.

2. Materials and methods

2.1. Cloning, protein expression, and purification

The synthetic *Xanthomonas albilineans cas2* gene was cloned into a modified pET32a vector (Fig. 2) and verified by DNA sequencing. The cloned plasmid was introduced into *Escherichia coli* strain BL21star (DE3) (Invitrogen) cells, and were grown in LB or minimal media with $^{15}NH_4Cl$ and/or $^{13}C_6$-glucose as sole nitrogen or carbon sources, respectively (Table 1). Cells were grown at 37°C until OD_{600} reached 0.8, induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside, and harvested by centrifugation after 16 hours of induction at 17°C. The cells were resuspended in 20 mM sodium phosphate, pH 6.5, 200 mM NaCl, 20 mM Imidazole, 2 mM β-mercaptoethanol, and 1 mM phenylmethylsulfonylfluoride. Cell lysis was performed using Emulsiflex C3 (Avestin, Canada) and the cell lysate was centrifuged at 20,000 x g for 30 minutes. The clear supernatant was loaded onto a His-trap column equilibrated with 20 mM sodium phosphate, pH 6.5, 200 mM NaCl, 20 mM imidazole, and 2 mM β-mercaptoethanol, and eluted using 500 mM imidazole. To remove the His_6-tag by the TEV protease, a long flexible linker was required between the His_6-tag and *Xa*Cas2, resulting in an N-terminal GSGSGSGSG-tail after the TEV protease digestion reaction. The digestion reaction was loaded again onto the His-trap column, and the flow-through fraction was loaded onto a Superdex 75 column equilibrated with 20 mM sodium phosphate, pH 6.5, 200 mM NaCl, 2 mM β-mercaptoethanol. *Xa*Cas2 was incubated in 10 mM HEPES, pH 7.4, 100 mM NaCl, and 10 mM EDTA for an hour to remove metal ions and finally prepared in 10 mM HEPES, pH 7.4, and 100 mM NaCl.

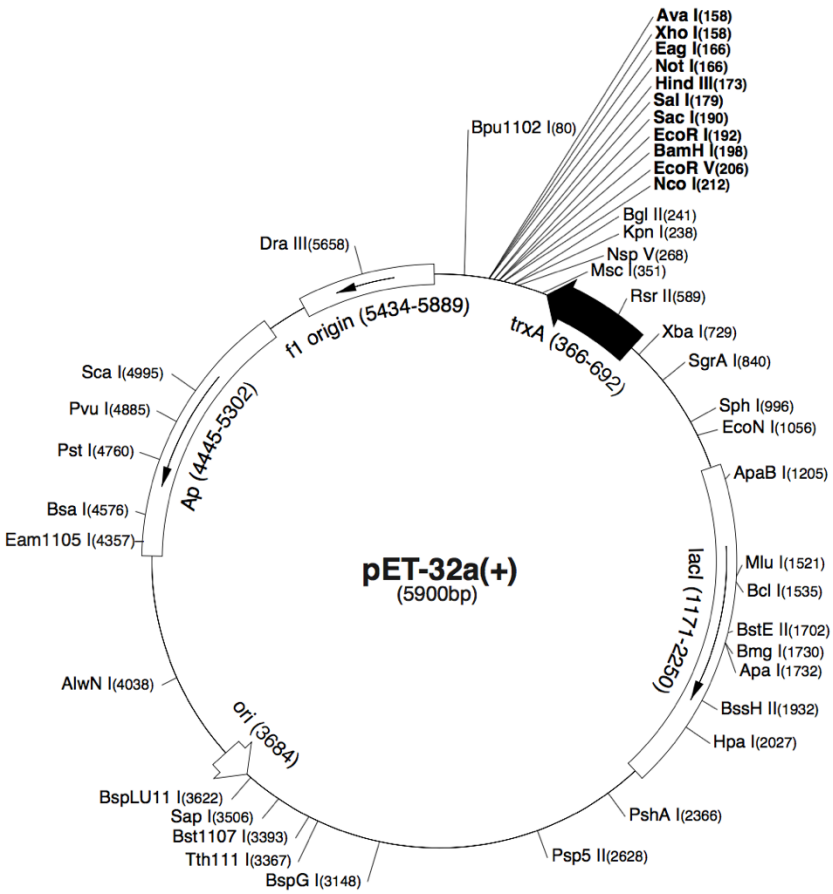

Figure 2. Sequence and map of pET 32a vector. A bacterial vector for expressing thioredoxin fusion proteins with an enterokinase site is depicted. This vector has variety enzyme sites and trx site for purifying target protein.

Table 1. The composition of medium. The chemical composition of the medium that were used in this study are listed below. LB medium was used for optimizing protein expression condition, and M9 minimal medium was used to generate isotopic-labelled protein sample for NMR investigation.

Medium	Composition (per Liter)
LB medium	1% Tryptophan, 0.5% Yeast extract, and 1% NaCl
M9 minimal medium (^{15}Nitrogen)	10 g K2HPO4, 13 g KH2PO4, 9 g Na2HPO4, 2.4 g K2PO4, 1x Trace elements, 0.2 mM CaCl2, 50 μg/ml Carbenicilin, 5 g Glucose, and 1 g $^{15}NH_4Cl$
M9 minimal medium (^{15}Nitrogen and ^{13}Carbon)	10 g K2HPO4, 13 g KH2PO4, 9 g Na2HPO4, 2.4 g K2PO4, 1x Trace elements, 0.2 mM CaCl2, 50 μg/ml Carbenicilin, 2 g U-$^{13}C_6$Glucose, and 1 g $^{15}NH_4Cl$

2.2. Multi-angle light scattering

*Xa*Cas2 were characterized by the size exclusion chromatography with inline multi-angle light scattering (SEC-MALS) experiments. 200 μM of *Xa*Cas2 was injected onto a WTC-030S column (Wyatt Technology) equilibrated with 10 mM HEPES, pH 7.4, and 100 mM NaCl. The chromatography system was connected to an 18-angle light-scattering detector (DAWN HELEOS II; Wyatt Technology), and a refractive index detector (Optilab T-rEX; Wyatt Technology). Data were collected every 1 s at a flow rate of 0.5 ml/min at 25°C. Data analysis was carried out using the software package ASTRA 6 to determine the molar mass and mass distribution of the sample.

2.3. Nuclease activity assay

The reaction mixture contained linearized pUC19 plasmid (100 ng) and *Xa*Cas2 protein (20 μM) in 20 mM HEPES, pH 7.4, and 100 mM NaCl. The metal-dependent nuclease activity was performed by incubation of the reaction mixture with 5 mM EDTA, or 4 mM of $MgCl_2$, $CaCl_2$, and $MnCl_2$ at 37°C for an hour. The reaction products were treated by protease K for half an hour and loaded onto a 0.7% (w/v) agarose gel for electrophoresis.

2.4. NMR Spectroscopy

The NMR sample contained 1.5 mM ^{13}C,^{15}N-*Xa*Cas2 in 50 mM sodium phosphate, pH 7.0, 100 mM NaCl, and 2 mM 1,4-dithiothreitol. The NMR spectra were collected on Bruker AVANCE 600, 800 and 900 MHz spectrometers equipped with a *z*-shielded gradient triple resonance cryoprobe at 35°C. Sequential and backbone assignments of ^{1}H, ^{15}N and ^{13}C resonances were achieved by three-dimensional triple resonance through-bond scalar correlation experiments (HNCOCACB, HNCACB, HNCO, HNCACO, HNCA and

HNCOCA) using transverse relaxation-optimized spectroscopy (TROSY) readouts. Residual $^{1}D_{NH}$ dipolar couplings were obtained by taking the difference in J splitting values measured in isotropic (water) and alignment media using 2D in-phase/antiphase $^{1}H–^{15}N$ heteronuclear single quantum coherence spectroscopy (HSQC). The alignment media employed were phage *pf1* (8 mg/ml; ASLA Biotech), and 250 mM NaCl was used to avoid line-broadening due to electrostatic interactions between the protein and the alignment media. ^{15}N-T_1, ^{15}N-T_2, and $^{1}H–^{15}N$ heteronuclear NOE measurements were carried out using pulse schemes described previously [23]. Delays of 10, 200, 400, 600, 800, 1000, and 1500 ms were used for the T_1 relaxation measurement, and 16.7, 33.4, 50.0, 66.8, 100.1, 150.2, and 233.7 ms were used for the T_2 relaxation measurement. For NMR titration, $^{1}H–^{15}N$ HSQC spectra were recorded for 0.3 mM of *Xa*Cas2 titrating with Mg^{2+} ions, and changes in the backbone amide chemical shifts were monitored. NMR spectra were processed using the NMRPipe [24] program and analyzed using the PIPP [25], the NMRView [26], and the Sparky [27] programs.

3. Results and Discussion

3.1. XaCas2 exhibits metal- and pH-dependent nuclease activity.

Previously, we have reported the crystal structure of C-terminal His_6-tagged *Xa*Cas2 and demonstrated its nuclease features. However, although it is weak, readily detectable interaction between His_6-tag and metal ions, especially Manganese ions used for *Xa*Cas2 metal preference nuclease activity, became concerned. Thus, we have cloned another *Xa*Cas2 with an N-terminal thioredoxin-His_6-tag followed by a TEV protease cleavage site to avoid producing any artifacts. Unusually, treating up to considerably excessive amount of TEV protease could not cleave off the His_6-tag from *Xa*Cas2. Alternatively, we have introduced a flexible Gly-Ser repeat linker between the TEV cleavage site and *Xa*Cas2, varying the length of the $(Gly\text{-}Ser)_n$ linker. Finally, the $(GS)_4$ linker yielded a complete removal of the His_6-tag. We assume that this unusual phenomenon is due to the inaccessible TEV cleavage site for TEV protease, located right before the first N-terminal secondary structure, β1 strand, in between two antiparallel strands (β3 and β4) and in nearby flexible loops (β3−α2 and β4−η1), thus, requiring longer linkers for efficient cleavage. Hence, *Xa*Cas2 in this study contains extra linker residues at the N-terminus (Figure 3).

$(GS)_4$-*Xa*Cas2

```
          10         20         30         40         50         60
GSGSGSGSGM MVLVSYDVST SSPGGDKRLR KVAKACRDLG QRVQFSVFEI EVDPAQWTAL

          70         80         90        100
RQRLCDLIDP DIDSLRFYHL GAKWEARVEH VGAKPSLDLK GPLIF
```

Figure 3. Sequence of *Xa*Cas2. The amino acid sequence of *Xa*Cas2 is indicated. Due to enabling the accessibility of restriction enzyme to the desired cleavage site, *Xa*Cas2 ended up carrying $(GS)_4$ linkers in N-terminus, though that does not affect to its functional activity. The unexpected linker is colored in red in the sequence.

To ensure that the newly constructed and purified *Xa*Cas2 possesses the same structural and functional characteristics compare to the previous one, we have conducted Multi-angle light scattering experiment and metal- and pH-dependent nuclease activity assays.

Multi-angle light scattering data showed that *Xa*Cas2 eluted as a single peak with the absolute molecular mass of 22.3 ± 0.3 kDa (Fig. 4A). The calculated molecular weight of *Xa*Cas2 was 11,492.1Da, and the data assured that *Xa*Cas2 formed a stable dimer in solution without any disruptions of the $(GS)_4$ linker.

Metal- and pH-dependent nuclease activity assays were also conducted to confirm the maintenance of functional features of *Xa*Cas2. The smeared band of the DNA substrate on the agarose gel, instead of various fragments, indicated that *Xa*Cas2 functions as a nonspecific exonuclease against dsDNA. Inconsistently, though we have suspected, the newly constructed *Xa*Cas2 indicated nuclease activity against double-stranded DNA in the presence of not only Mg^{2+} but also Mn^{2+} ions, while C-terminal His_6-tagged *Xa*Cas2 did not show the activity in the presence of Mn^{2+} ions. This result implies that although the intrinsic Mn^{2+} binding affinity of the His_6-tag is weak (K_D ~800 μM), two His_6-tags attached to the *Xa*Cas2 dimer possibly have increased the affinity of Mn^{2+} binding, interfering with the nuclease activity [28]. Note that the nuclease activity of *Xa*Cas2 is not originated neither from possibly co-purified trace nucleases nor endogenously bounded metal ions, since incubation of *Xa*Cas2 with the substrate did not show any nuclease activity. The pH-dependency of *Xa*Cas2's nuclease activity was coherent to the His_6-tag *Xa*Cas2, as active in the range of pH 7 to pH 10 but inactive at pH 6.

In conclusion, the newly constructed N-terminal $(GS)_4$-tagged *Xa*Cas2 forms a stable dimeric conformation in solution, exhibits Mg^{2+} and Mn^{2+} ions and pH-environmental dependencies for double-stranded DNA nuclease activity.

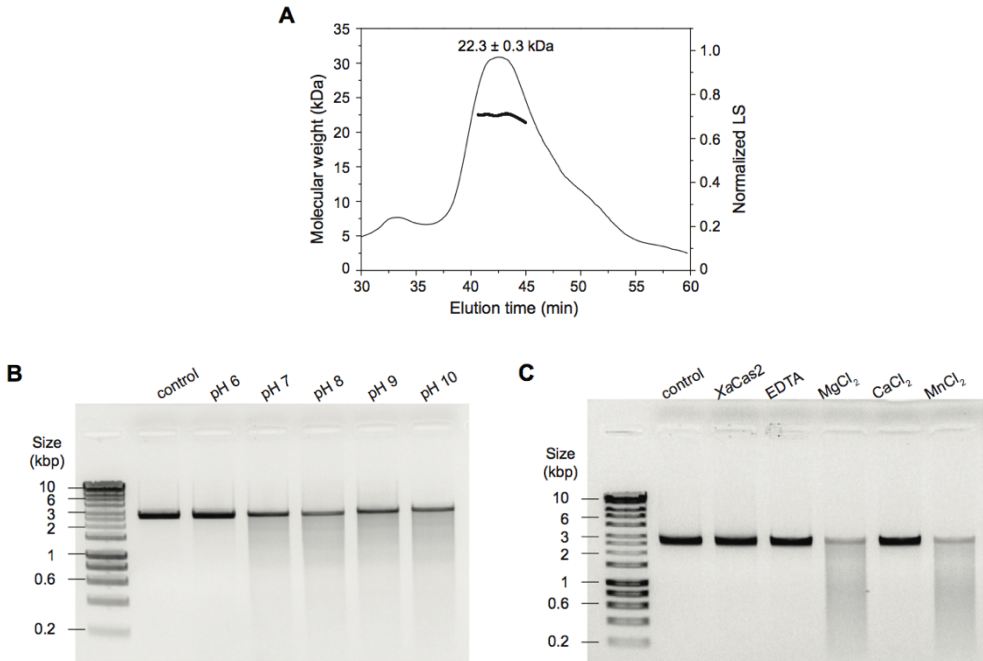

Figure 4. Multi-angle light scattering analysis and nuclease activity assay of *Xa*Cas2. (A) Light scattering data and absolute molecular mass of the *Xa*Cas2 dimer. (B) dsDNA nuclease activity of *Xa*Cas2 at varying pH in the presence of $MgCl_2$. The control denotes the substrate vector alone without *Xa*Cas2. (C) dsDNA nuclease activity of *Xa*Cas2 with different divalent metal ions. *Xa*Cas2 denotes the nuclease assay for fresh purified *Xa*Cas2 without addition of metal ions. EDTA denotes the assay for *Xa*Cas2 with a pre-treatment with EDTA. 4 mM of divalent metal ions were introduced for the assay.

3.2. XaCas2 adopts a catalytically inactive conformation in solution.

To compare the solution state of *Xa*Cas2 to the crystal structure, a series of 3D heteronuclear correlation NMR experiments were carried out. First, for the backbone assignment of *Xa*Cas2, chemical shifts for 87 out of 91 residues (excluding five prolines) and additional 3 linker residues were assigned and annotated in the 2D ^{1}H−^{15}N HSQC spectrum (Figure 5) (Table 2). The residues, Gly15 (located in the flexible loop region, β1−α1) and Trp 75, Glu 76, and Arg 78 (located in the hinge-region, 3_{10}-helix in the β4–β5 linker) were not be able to assign.

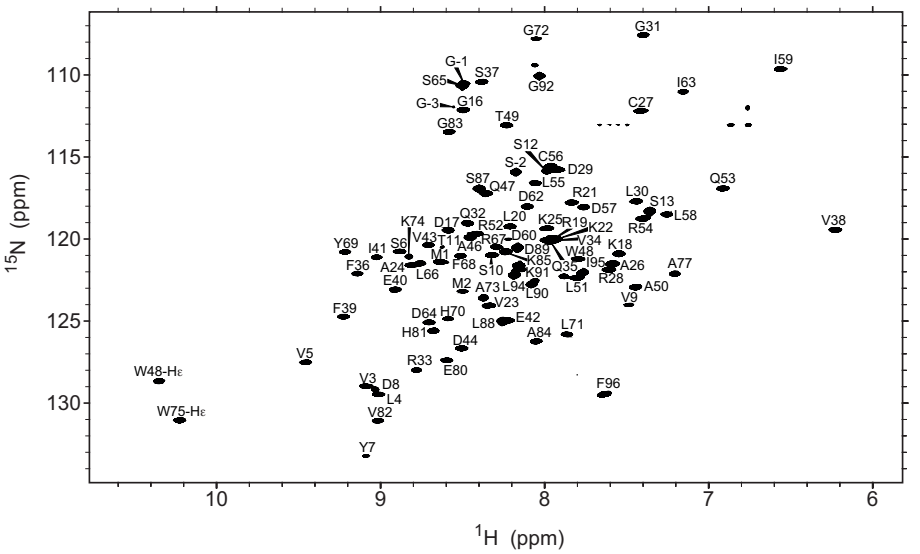

Figure 5. Backbone assignment of *Xa*Cas2. 2D ^{1}H–^{15}N HSQC spectrum of *Xa*Cas2 annotated by residue types and numbers. The assignment for the N-terminal $(GS)_4$ linker residues were annotated as G−1, S−2, and G−3. The assignment for the side chain NHε resonances of tryptophan indole groups were also annotated.

Table 2. Backbone HN, N, Cα, Cβ, and CO with side chain Hε and Nε chemical shifts of *Xa*Cas2 (in ppm). Chemical shifts of *Xa*Cas2 are listed by three-letter-code residue types and numbers. The assignment for the N-terminal $(GS)_4$ linker residues were annotated as G−1, S−2, and G−3. The assignment for the side chain NHε resonances of tryptophan indole groups were also listed.

Residue		HN	N	Cα	Cβ	CO	Hε	Nε
-3	GLY	8.554	112.0	45.52		174.2		
-2	SER	8.170	116.0	58.36	64.43	174.6		
-1	GLY	8.496	110.7	45.22		173.1		
1	MET	8.635	121.4	54.50	34.66	173.8		
2	MET	8.497	123.2	54.17	32.63	175.7		
3	VAL	9.092	129.0	60.47	34.90	173.8		
4	LEU	9.008	129.5	53.21	45.33	175.3		
5	VAL	9.455	127.5	61.38	33.78	174.6		
6	SER	8.880	120.8	56.32	62.53	174.8		
7	TYR	9.093	133.2	53.39		172.8		
8	ASP	9.015	129.3	52.87	41.38	174.6		
9	VAL	7.481	124.0	58.52	33.42	174.4		
10	SER	8.320	121.0	57.06	63.13	177.2		
11	THR	8.625	120.5	63.27	45.17	175.0		
12	SER	7.982	115.9	59.04	63.46	174.4		
13	SER	7.354	118.3	56.37	63.72	172.7		
14	PRO							
15	GLY							

16	GLY	8.492	112.1	47.96		174.7
17	ASP	8.583	119.5	57.14	40.22	178.3
18	LYS	7.542	120.9	59.35	32.28	179.2
19	ARG	7.957	120.1	61.72	30.06	178.7
20	LEU	8.206	119.3	58.62	40.68	178.3
21	ARG	7.831	117.8	60.25	29.99	179.5
22	LYS	7.925	120.0	60.25	32.81	179.7
23	VAL	8.332	124.1	67.09	31.83	177.3
24	ALA	8.812	121.6	55.09	17.59	180.6
25	LYS	7.983	119.4	59.58	32.21	178.5
26	ALA	7.580	121.5	54.45	17.80	179.8
27	CYS	7.413	112.2	62.03	28.13	176.3
28	ARG	7.604	121.9	59.32	30.26	176.9
29	ASP	7.907	115.8	55.83	40.67	176.6
30	LEU	7.434	117.7	54.39	44.00	176.4
31	GLY	7.394	107.6	46.73		172.0
32	GLN	8.466	119.1	54.40	32.58	174.3
33	ARG	8.778	128.0	55.76	30.72	175.4
34	VAL	7.969	120.1	62.42	32.21	175.4
35	GLN	7.971	120.1	54.62	35.57	174.5
36	PHE	9.137	122.1	63.06	39.77	176.8
37	SER	8.379	110.4	56.42	64.31	171.6
38	VAL	6.219	119.5	61.60	35.31	174.0
39	PHE	9.225	124.8	55.29	42.46	174.6
40	GLU	8.906	123.1	55.11	31.46	176.1
41	ILE	9.023	121.2	59.22	41.18	173.1

42	GLU	8.219	125.0	55.01	32.07	175.4		
43	VAL	8.706	120.4	58.90	36.16	174.3		
44	ASP	8.500	126.7	52.06	41.63	176.2		
45	PRO			66.16	32.17	178.8		
46	ALA	8.452	119.9	54.46	18.52	179.8		
47	GLN	8.352	117.2	56.03	29.80	177.5		
48	TRP	7.792	121.3	59.38	30.27	175.5	10.35	128.7
49	THR	8.231	113.1	67.00	68.17	176.7		
50	ALA	7.440	123.0	54.62	18.36	179.8		
51	LEU	7.800	122.4	57.55	41.42	177.4		
52	ARG	8.406	119.7	59.40	28.41	178.3		
53	GLN	6.910	116.9	58.57	28.30	176.9		
54	ARG	7.398	118.8	59.22	29.92	178.9		
55	LEU	8.053	116.6	57.96	41.63	178.8		
56	CYS	7.957	115.6	64.25	26.41	176.6		
57	ASP	7.757	118.1	55.92	40.91	177.0		
58	LEU	7.249	118.5	55.63	43.86	177.5		
59	ILE	6.557	109.7	60.55	39.81	174.8		
60	ASP	8.222	120.0	50.51	40.99			
61	PRO			64.07	31.89	176.0		
62	ASP	8.102	118.1	56.56	41.38	177.4		
63	ILE	7.149	111.0	61.55	41.21	173.9		
64	ASP	8.702	125.1	52.85	45.38	175.7		
65	SER	8.490	110.6	57.86	68.35	173.7		
66	LEU	8.756	121.5	53.03	46.57	174.1		
67	ARG	8.292	120.5	54.96	34.72	174.4		

68	PHE	8.512	121.0	53.68	41.01	175.1		
69	TYR	9.214	120.8	56.68	40.05	175.3		
70	HIS	8.594	124.8	56.87	30.30	175.9		
71	LEU	7.857	125.8	55.31	42.15	177.1		
72	GLY	8.049	107.8	44.44		173.0		
73	ALA	8.368	123.6	52.23	19.20	179.0		
74	LYS	8.824	121.1	58.38				
75	TRP						10.22	131.0
76	GLU							
77	ALA	7.203	122.1	52.59	18.70	177.2		
78	ARG							
79	VAL	7.879	122.3	59.41	34.91	174.7		
80	GLU	8.591	127.4	54.69	33.91	173.3		
81	HIS	8.699	126.0	55.91	34.50	174.9		
82	VAL	9.012	131.1	61.67	34.87	174.2		
83	GLY	8.581	113.5	44.24		173.9		
84	ALA	8.048	126.2	53.04	19.44	177.5		
85	LYS	8.235	120.8	53.85	32.54	174.6		
86	PRO			62.93	31.79	176.8		
87	SER	8.398	116.9	58.07	63.55	174.9		
88	LEU	8.254	125.0	55.34	42.41	177.0		
89	ASP	8.161	120.5	54.43	41.38	176.1		
90	LEU	8.082	122.8	55.34	42.25	177.1		
91	LYS	8.153	121.7	56.30	33.14	176.7		
92	GLY	8.027	110.1	44.67		171.6		
93	PRO			63.19	32.19	176.7		

94	LEU	8.181	122.2	55.19	42.29	176.4
95	ILE	7.763	122.1	60.78	38.89	174.6
96	PHE	7.629	129.4	59.04	40.47	180.1

Then, we measured the residual dipolar couplings (RDCs) of backbone amide groups ($^1D_{NH}$) in the *pf1* phage alignment medium to examine the backbone fold of *Xa*Cas2 in solution (Figure 6). The magnitude and orientation of the alignment tensors were obtained by the singular value decomposition (SVD) analysis for a best fit between observed RDCs and those calculated from the atomic coordinates of the crystal structure of *Xa*Cas2 (PDB code 5H1O).

Because from the crystal structure of dimeric *Xa*Cas2 exhibited conformational heterogeneity at the linker region (Ala73−Val79) between β4 and β5 strands, experimental RDCs did not agree with coordinates of individual subunits owing to large differences between observed and calculated RDCs from the β4–β5 linker region. The RDC R-factors for the fitting with individual A and B chains of the dimer were 31.1% and 28.0%, respectively (Figure 6A and 6B).

Yet, after perceiving that the measured RDCs were more likely to be weighted since the $^1H–^{15}N$ HSQC spectra represents only one set of backbone amide resonances for the β4–β5 linker residues, we carried out an ensemble fitting of RDCs assuming two conformers of the subunits were equally populated (Figure 6C). The agreement between observed and calculated RDCs significantly improved by the ensemble fitting, with the R-factor of 15.6%. Experimental RDCs from the β4–β5 linker region were in excellent agreement with the ensemble of two conformational states. In the course of the ensemble fitting of RDCs, the conformers can assume different alignment tensors, such that the weighted sum of their RDCs correspond with observed RDCs. We note that the C_2 symmetry axis of the dimer should coincide with one of the principal axes of the alignment tensor. Thus, in principle, individual alignment tensors obtained by the ensemble fitting would agree to each other in their magnitudes and orientations. The axial component (D_a) and the rhombicity (η) were 15.1 Hz and 0.40, respectively, for chain A, and 15.0 Hz and 0.41,

respectively, for chain B (Table 3). The orientation of the alignment tensors was also very similar between the two subunits, and slight differences in the Euler angles were attributed to imperfect symmetry in the crystal coordinates (Table 3).

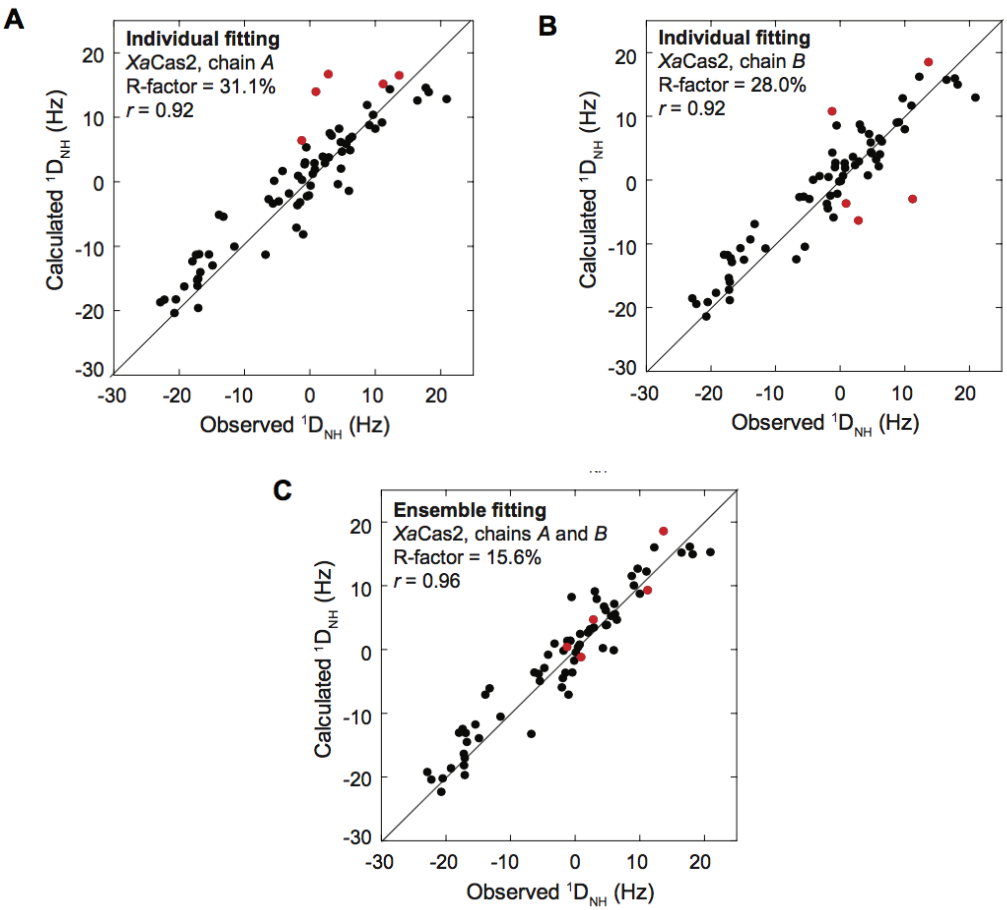

Figure 6. Comparison of observed $^{1}D_{NH}$ residual dipolar couplings for the backbone amide resonances of *Xa*Cas2 and those calculated from the atomic coordinates of crystal structure (PDB code 5H1O). (A) Crystal coordinates of chain A, and (B) crystal coordinates of chain B of the *Xa*Cas2 dimer were used for the SVD analysis. (C) Crystal coordinates of both A and B chains were used simultaneously with equal population. RDCs from the β4–β5 linker residues are highlighted in *red*. The RDC R-factors and the Pearson correlation coefficients (*r*) are shown in the panels.

Table 3. SVD analysis of $^{1}D_{NH}$ RDCs measured for chain A and chain B subunits of the *Xa*Cas2 dimer (PDB code 5H1O). Individual fitting employs the two coordinates separately, and ensemble fitting employs both coordinates simultaneously with equal population. The alignment tensor is described by five parameters: three Euler angles, which describe the orientation of the tensor in the atomic coordinate frame, the magnitude of the principal component of the tensor (D_a^{NH}), and the rhombicity (η). Same number of experimental RDC data were used for the analysis.

***Xa*Cas2 coordinates**	**Number of RDCs**	**Euler angles (°)**			**(D_a^{NH}) (*Hz*)**	**η**	**RDC R-factor (%)**
		ϕ	**θ**	**ψ**			
Individual fitting							
Chain A	70	124	5	100	–10.3	0.47	31.1
Chain B	70	128	7	91	–10.7	0.53	28.0
Ensemble fitting							
Chain A	70	15	89	155	15.1	0.40	15.6
Chain B	70	12	81	165	15.0	0.41	

Taken together, *Xa*Cas2 exhibited a catalytically inactive conformation in solution similar to the crystal structure. The β4–β5 linker region, however, exhibited a dynamic conformational ensemble, which was consistent with the NMR relaxation data in the following section.

3.3. The linker region and the metal binding site is highly dynamic

The innate backbone dynamics of *Xa*Cas2 in solution were measured via the ^{15}N R_1 and R_2 relaxation rates and $^{1}H-^{15}N$ heteronuclear NOE (Fig. 7). The crystallographically unstructured C-terminal region (Ser87–Phe96) exhibited significantly increased ^{15}N R_1 relaxation rates, and reduced ^{15}N R_2 relaxation rates and $^{1}H-^{15}N$ heteronuclear NOE values as typical of flexible regions and disordered tail [20]. The C-terminal region of Cas2 adopts well-defined β strands and provides the binding interface with Cas1 in the crystal structure of the Cas1–Cas2 complex, but are completely disordered in the absence of Cas1 [13,14]. The average backbone ^{15}N R_1 and ^{15}N R_2 relaxation rates of the structured region were 0.67 ± 0.06 s^{-1} and 23.2 ± 6.5 s^{-1}, respectively. From the ratio of the relaxation parameters (R_2/R_1), the rotational correlation time of *Xa*Cas2 was estimated as ~12.4 ns [29], which was typical of a globular protein with the size of a *Xa*Cas2 dimer (~23 kDa).

^{15}N R_2 relaxation rates significantly increased at the β4–β5 linker region that contained a 3_{10}-helix between Lys74 and Ala77. Ala73, Lys74, and Ala77 exhibited ^{15}N R_2 relaxation rates of 31.9 ± 0.5, 41.8 ± 1.0, and 32.7 ± 0.1, respectively. ^{15}N R_2 relaxation rates of Trp75, Glu76, Arg78, and Val79 in the linker region also exhibited a large increase, but could not be accurately determined due to severe line broadening. Collectively, the β4–β5 linker region exhibited extensive line broadening that is attributed to conformational

exchanges on the μs−ms timescales. It is remarkable that Tyr7 and Asp8 in the metal binding site also exhibited large ^{15}N R_2 relaxation rates of 47.6 ± 1.3 and 41.9 ± 1.8, respectively, which was comparable to those observed in the linker region. It has been postulated that Cas2 has a rigid-body hinge motion in the β4–β5 linker region to switch into the active metal-bound state with nuclease activity [16,17,20]. We speculate that the hinge motion might be responsible for the conformational dynamics in the β4–β5 linker region and the metal binding site.

*Xa*Cas2 binds to a divalent metal ion for the nuclease activity. We carried out the chemical shift perturbation experiment using the $^{1}H–^{15}N$ HSQC spectra to monitor the metal binding of *Xa*Cas2. When we titrated 0.2 mM ^{15}N-*Xa*Cas with up to 100 mM of the Mg2+ ion, we could not observe noticeable chemical shift changes in the backbone amide resonances. We also performed the isothermal titration calorimetry to measure the binding thermodynamics of the interaction *Xa*Cas2 and Mg^{2+} ions, but could not observe measurable reaction heats associated with the interaction. Both results indicate that *Xa*Cas2 only transiently binds to Mg^{2+} ions without tight association. We suppose that the catalytically active conformation of *Xa*Cas2 is energetically unfavorable due to a suboptimal dimer interface and only transiently formed to bind to metal ions and function as a nuclease. Our result is consistent with the observation that the nuclease activity requires relatively a high concentration (> 20 μM) of *Xa*Cas2 to monitor the reaction within hours. Taken together, we propose that *Xa*Cas2 predominantly adopts a conformation that functions as a scaffold protein for the acquisition of foreign genetic elements, and transiently switches into a metastable minor conformation that allows for the nuclease activity.

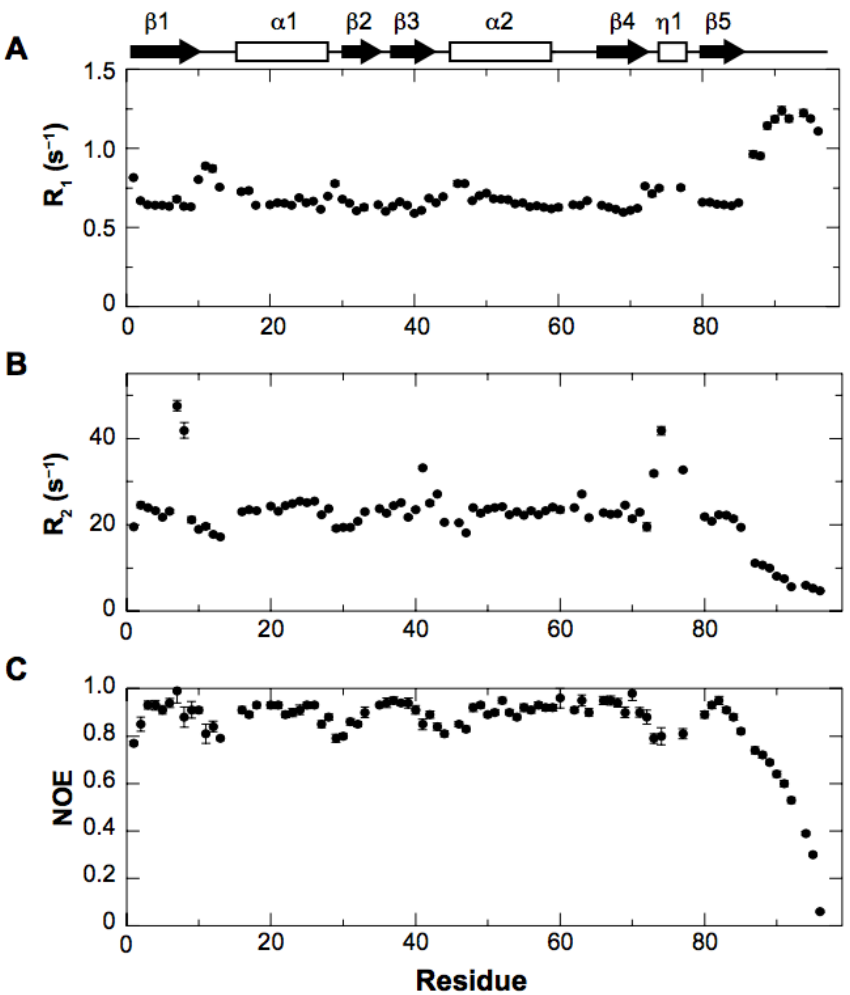

Figure 7. (A) ^{15}N R_1 relaxation parameters, (B) ^{15}N R_2 relaxation parameters, and (C) $^{1}H–^{15}N$ heteronuclear NOE data as a function of the residue number of *Xa*Cas2. The secondary structures of *Xa*Cas2 are shown on top as a schematic representation. η1 denotes the 3_{10}-helix in the β4–β5 linker between Lys74 and Ala77.

Table 4. ^{15}N R_1 and ^{15}N R_2 relaxation rates and ^{1}H–^{15}N heteronuclear NOE values with errors. Relaxation rates and hetNOE values of *Xa*Cas2 are listed by three-letter-code residue types and numbers. Excluding five prolines, 84 out of 91 residues were analyzed.

Residue		R1	R1 error	R2	R2 error	hetNOE	hetNOE error
1	MET	0.816327	0.002599	19.592476	0.128979	0.77	0.009
2	MET	0.669792	0.007268	24.557957	0.693557	0.85	0.030
3	VAL	0.643501	0.003984	23.992322	0.237736	0.93	0.017
4	LEU	0.642261	0.004579	23.326335	0.275324	0.93	0.021
5	VAL	0.641437	0.003090	21.748586	0.215688	0.91	0.018
6	SER	0.634115	0.008324	23.110700	0.368532	0.94	0.019
7	TYR	0.680735	0.016404	47.619048	1.249433	0.99	0.051
8	ASP	0.634115	0.008082	47.869800	1.798840	0.85	0.051
9	VAL	0.630915	0.013016	21.195422	0.655899	0.91	0.035
10	SER	0.803213	0.006032	18.964536	0.247082	0.91	0.013
11	THR	0.888889	0.017620	19.681165	0.553908	0.81	0.040
12	SER	0.873362	0.020900	17.796761	0.226775	0.84	0.022
13	SER	0.755858	0.004622	17.208742	0.160508	0.79	0.008
14	PRO						
15	GLY						
16	GLY	0.727802	0.010011	23.068051	0.353338	0.91	0.014
17	ASP	0.732601	0.011002	23.551578	0.159192	0.89	0.011

18	LYS	0.643087	0.001758	23.223409	0.070112	0.93	0.012
19	ARG						
20	LEU	0.645995	0.005508	24.396194	0.295206	0.93	0.014
21	ARG	0.657030	0.003156	23.132084	0.244003	0.93	0.014
22	LYS	0.652742	0.003622	24.443901	0.285607	0.89	0.013
23	VAL	0.643087	0.005955	24.937656	0.092661	0.90	0.017
24	ALA	0.688705	0.010814	25.516713	0.214864	0.91	0.020
25	LYS	0.657462	0.006397	25.188917	0.231586	0.93	0.014
26	ALA	0.666223	0.003276	25.464731	0.188700	0.93	0.010
27	CYS	0.615764	0.003412	22.361360	0.254015	0.85	0.015
28	ARG	0.698812	0.002554	23.758612	0.099911	0.88	0.012
29	ASP	0.778816	0.013647	19.168104	0.374765	0.79	0.017
30	LEU	0.678426	0.003176	19.417476	0.133849	0.80	0.012
31	GLY	0.654022	0.005775	19.425019	0.430158	0.86	0.013
32	GLN	0.606061	0.002391	20.802996	0.120741	0.85	0.011
33	ARG	0.627746	0.018324	22.993792	0.407639	0.90	0.022
34	VAL						
35	GLN	0.643501	0.003520	23.736055	0.114934	0.93	0.007
36	PHE	0.603865	0.005652	22.686025	0.322689	0.94	0.020
37	SER	0.635728	0.006547	24.431957	0.335469	0.95	0.014

38	VAL	0.664452	0.004296	25.125628	0.235474	0.94	0.011
39	PHE	0.640615	0.003890	21.795990	0.273638	0.94	0.021
40	GLU	0.591017	0.005205	23.568230	0.259956	0.91	0.018
41	ILE	0.609013	0.005489	33.244681	0.308353	0.85	0.022
42	GLU	0.685871	0.006633	25.050100	0.516439	0.89	0.015
43	VAL	0.656168	0.003518	27.159153	0.112856	0.84	0.015
44	ASP	0.694927	0.007147	20.593081	0.186169	0.81	0.012
45	PRO						
46	ALA	0.780031	0.017037	20.533881	0.228107	0.85	0.010
47	GLN	0.777605	0.004432	18.112661	0.139429	0.83	0.011
48	TRP	0.671592	0.005187	23.998080	0.250520	0.92	0.016
49	THR	0.700771	0.005647	22.711787	0.608674	0.93	0.011
50	ALA	0.718907	0.004987	23.691068	0.127408	0.89	0.010
51	LEU	0.681199	0.002339	23.946360	0.147944	0.90	0.011
52	ARG	0.680735	0.006071	24.271845	0.267462	0.95	0.013
53	GLN	0.676590	0.004115	22.411475	0.139632	0.90	0.012
54	ARG	0.650618	0.002269	23.078698	0.090546	0.88	0.011
55	LEU	0.656168	0.002786	22.212350	0.211664	0.92	0.016
56	CYS	0.630915	0.003809	23.353573	0.417223	0.91	0.014
57	ASP	0.640205	0.002574	22.381379	0.126233	0.93	0.013

58	LEU	0.627746	0.004926	23.304591	0.102104	0.92	0.014
59	ILE	0.619963	0.008802	24.067389	0.491195	0.92	0.014
60	ASP	0.627746	0.017299	23.557126	0.693673	0.96	0.042
61	PRO						
62	ASP	0.643915	0.002347	23.929170	0.137425	0.91	0.012
63	ILE	0.642261	0.010560	27.122322	0.492866	0.95	0.023
64	ASP	0.670691	0.007152	21.635656	0.134813	0.90	0.015
65	SER	1.071926	0.022636	8.223684	0.246169	0.65	0.005
66	LEU	0.641437	0.006130	22.867597	0.226950	0.95	0.018
67	ARG	0.628141	0.003622	22.441652	0.156628	0.95	0.019
68	PHE	0.616523	0.003497	22.583559	0.262659	0.94	0.018
69	TYR	0.596659	0.007547	24.612355	0.291374	0.90	0.022
70	HIS	0.611247	0.004483	21.473051	0.497979	0.98	0.030
71	LEU	0.621118	0.005671	22.899015	0.369153	0.90	0.020
72	GLY	0.761615	0.008527	19.592476	0.948147	0.88	0.030
73	ALA	0.712758	0.016155	31.877590	0.526382	0.79	0.021
74	LYS	0.749064	0.011615	41.841004	0.957616	0.80	0.036
75	TRP						
76	GLU						
77	ALA	0.927644	0.0450914	8.31947	0.593852	0.58	0.095

78	ARG						
79	VAL	0.751880	0.013229	32.701112	0.136878	0.81	0.022
80	GLU	0.661376	0.003696	21.929825	0.282779	0.89	0.015
81	HIS	0.661813	0.006395	20.876827	0.109396	0.93	0.013
82	VAL	0.648088	0.003226	22.416499	0.354765	0.95	0.017
83	GLY	0.643501	0.001764	22.256844	0.412641	0.91	0.012
84	ALA	0.637755	0.002034	21.431633	0.408331	0.88	0.013
85	LYS	0.656168	0.003914	19.474197	0.094811	0.82	0.011
86	PRO						
87	SER	0.964320	0.021016	11.164452	0.183228	0.74	0.015
88	LEU	0.953289	0.019811	10.667805	0.167289	0.72	0.012
89	ASP	1.143641	0.020534	9.940358	0.243074	0.69	0.009
90	LEU	1.185255	0.020089	8.084074	0.235268	0.64	0.012
91	LYS	1.240849	0.024789	7.501875	0.248187	0.60	0.012
92	GLY	1.188213	0.020189	5.659310	0.203697	0.53	0.012
93	PRO						
94	LEU	1.224290	0.022333	6.006006	0.190821	0.39	0.007
95	ILE	1.189768	0.013802	5.350455	0.134262	0.30	0.006
96	PHE	1.108647	0.005211	4.646840	0.099976	0.06	0.004

3.4. Model for the conformational transition

The crystal structure of *Xa*Cas2 presents as inactive state with 11.3 Å far distance between the Asp8 residues from each protomers, not able to accommodate a single Mg^{2+} ion (Fig. 8) [20]. However, the distance can be reduced by a rigid-body rotation (Fig. 8). We note that there is a domain swapping in the *Xa*Cas2 dimer such that the β5 strand of one dimer subunit forms a five-stranded β sheet with the other subunit. The rigid-body rotation in the dimer requires conformational changes at the β4–β5 linker, which can be described as a hinge motion of the linker region. It is well known that the dimerization of the domain-swapped dimer is often accompanied by the hinge loop dynamics [30,31]. The hinge loop dynamics may play a role in the conformational switch of *Xa*Cas2 for the nuclease activity.

The biological significance of the nuclease activity of Cas2 in the context of the CRISPR-Cas function has not been clear, since active site mutation of Cas2 does not abrogate the spacer acquisition by the Cas1-Cas2 complex [13]. In the Cas1-Cas2 complex, Cas2 acts as a scaffold protein to determine the length of the protospacer, while Cas1 functions as a nuclease for the spacer acquisition [14]. We note, however, that the nuclease activity of Cas2 has been linked to an infectious process of *Legionella pneumophila*, the primary pathogen of Legionnaires' disease [32]. Catalytically deficient Cas2 of *L. pneumophila* caused an impaired infectivity for amoebal host cells that served as a natural reservoir of the pathogen. Thus, the catalytic activity of Cas2 may be related to biological processes that are distinct from the CRISPR-Cas function of phage immunity.

In summary, we report the first solution structure and dynamics of *Xa*Cas2, which reveals conformational dynamics at the active site for metal binding and the linker region for inter-domain motions. Crystal structures have suggested a hinge motion at the linker region based on the conformational

heterogeneity between different Cas2 homologs or between Cas2 crystals at different crystallization conditions. We note, however, that the conformational heterogeneity can also originate from crystal packing that can strain and distort loop conformations [33,34]. It is thus of importance to validate that the loop flexibility is genuine in solution, and that conformations in the crystal structures represent the conformational ensemble in dynamic exchange. The solution structure of *Xa*Cas2 confirms that the catalytically inactive conformation persists in solution as was determined by the crystallography. RDCs and backbone dynamics at the active site, however, reveal a dynamic equilibrium of conformational states that may enable *Xa*Cas2 to handle the structural role as a scaffold protein and also the functional role as a nuclease. We speculate that the conformational dynamics at the hinge loop region and the metal binding site may allow *Xa*Cas2 to transiently access a minor state that exhibits the metal-dependent nuclease activity.

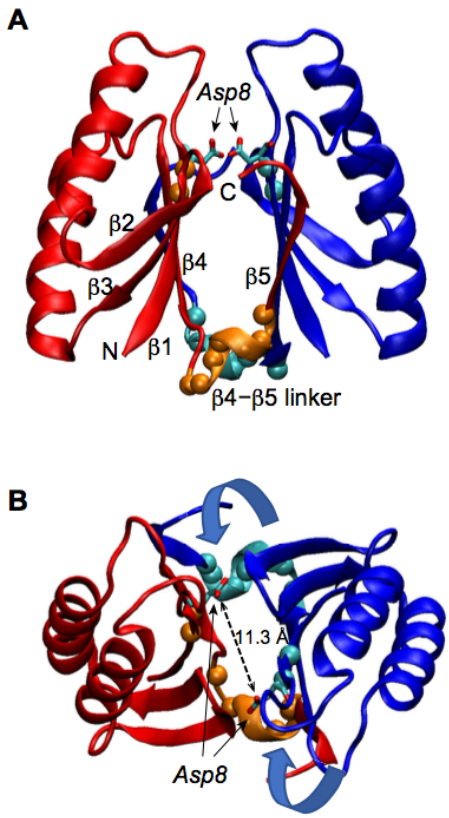

Figure 8. The three-dimensional structure of the *Xa*Cas2 dimer (PDB code 5H1O) in a ribbon diagram. (A) *Front* view and (B) *top* view of the *Xa*Cas2 dimer. The N- and C-terminus, and secondary structures are annotated in one subunit. The active site *Asp8* pair for the metal binding is shown as a *stick*, and the distance between the side chain carboxyl groups is denoted. The molecules are colored in *blue* and *red* for chain A and B, respectively. The β4–β5 linker regions are colored in *cyan* and *orange* for chain A and B, respectively. The backbone nitrogen atoms of residues with conformational dynamics shown in a space-filling model in *cyan* and *orange* for chain A and B, respectively. The putative hinge motion that brings the Asp8 pair to a close proximity is shown by *light blue* arrows.

4. References

[1] Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, *et al.* (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709–1712.

[2] Barrangou R and Marraffini LA (2014) CRISPR-Cas systems: Prokaryotes upgrade to adaptive immunity. *Mol Cell* **54**, 234–244.

[3] Marraffini LA and Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843–1845.

[4] Hatoum-Aslan A and Marraffini LA (2014) Impact of CRISPR immunity on the emergence and virulence of bacterial pathogens. *Curr Opin Microbiol* **17**, 82–90.

[5] Marraffini LA (2015) CRISPR-Cas immunity in prokaryotes. *Nature* **526**, 55–61.

[6] Fineran PC and Charpentier E (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**, 202–209.

[7] Heler R, Marraffini LA and Bikard D (2014) Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Mol Microbiol* **93**, 1–9.

[8] Brouns SJ, Jore MM, Lundgren M, Westra ER, Slijkhuis RJ, Snijders AP, *et al.* (2008) Small CRISPR RNAs guide antiviral defense inprokaryotes. *Science* **321**, 960–964.

[9] Jiang W, Samai P and Marraffini LA (2016) Degradation of phage transcripts by CRISPR-Associated RNases enables type III CRISPR-Cas immunity. *Cell* **164**, 710–721.

[10] Marraffini LA and Sontheimer EJ (2010) CRISPR interference: RNA-

directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**, 181–190.

[11] Koonin EV, Makarava KS and Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67–78.

[12] Marraffini LA and Sontheimer EJ (2009) Invasive DNA, chopped and in the CRISPR. Structure. **17**, 786–788.

[13] Nunez JK, Kranzusch PJ, Noeske J, Wright AV, Davies CW and Doudna JA (2014) Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat Struct Mol Biol* **21**, 528–534.

[14] Wang J, Li J, Zhao H, Sheng G, Wang M, Yin M, *et al.* (2015) Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* **163**, 840–853.

[15] Beloglazova N, Brown G, Zimmerman MD, Proudfoot M, Makarova KS, Kudritska M, *et al.* (2008) A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J Biol Chem***283**, 20361–20371.

[16] Nam KH, Ding F, Haitjema C, Huang Q, DeLisa MP and Ke A (2012) Double-stranded endonuclease activity in Bacillus halodurans clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J Biol Chem* **287**, 35943–35952.

[17] Ka D, Kim D, Baek G and Bae E (2014) Structural and functional characterization of Streptococcus pyogenes Cas2 protein under different pH conditions. *Biochem Biophys Res Commun* **451**, 152–157.

[18] Samai P, Smith P and Shuman S. (2010) Structure of a CRISPR-associated protein Cas2 from Desulfovibrio vulgaris. *Acta Crystallogr Sect F: Struct Biol Cryst Commun* **66**, 1552–1556.

[19] Jung TY, Park KH, An Y, Schulga A, Deyev S, Jung JH, *et al.* (2016) Structural features of Cas2 from Thermococcus onnurineus in CRISPR-cas system type IV. *Protein Sci* **25**, 1890–1897.

[20] Ka D, Hong S, Jeong U, Jeong M, Suh N, Suh JY and Bae E (2017) Structural and dynamic insights into the role of conformational switching in the nuclease activity of the Xanthomonas albilineans Cas2 in CRISPR-mediated adaptive immunity. *Struct Dynam* **4**, 054701.

[21] Blanch M, Legaz ME and Vicente C (2008) Xanthan production by *Xanthomonas albilineans* infecting sugarcane stalks. *J Plant Physiol* **165**, 366–374.

[22] Pieretti I, Royer M, Barbe V, Carrere S, Koebnik R, Couloux A, *et al.* (2012) Genomic insights into strategies used by Xanthomonas albilineans with its reduced artillery to spread within sugarcane xylem vessels. *BMC Genomics* **13**, 658.

[23] Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, *et al.* (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33**, 5984–6003.

[24] Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J and Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293.

[25] Garrett DS, Powers R, Gronenborn AM and Clore GM (1991) A common sense approach to peak picking in two-, three-, and four-dimensional spectra using automatic computer analysis of contour diagrams. *J Magn Reson* **195**, 214–220.

[26] Johnson BA (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* **278**, 313–352.

[27] Lee W, Tonelli M and Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327.

[28] Ching HY, Mascali FC, Bertrand HC, Bruch EM, Demay-Drouhard P, Rasia RM, *et al.* (2016) The use of Mn(II) bound to His-tags as genetically encodable

spin-label for nanometric distance determination in proteins. *J Phys Chem Lett* **7**, 1072–1076.

[29] Kay LE, Torchia DA and Bax A (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* **28**, 8972–8979.

[30] Heringa J and Taylr WR (1997) Three-dimensional domain duplication, swapping and stealing. *Curr Opin Struct Biol* **7**, 416–421.

[31] Gronenborn AM (2009) Protein acrobatics in pairs– dimerization via domain swapping. *Curr Opin Struct Biol* **19**, 39–49.

[32] Gunderson FF, Mallama CA, Fairbairn SG and Cianciotto NP (2015) Nuclease activity of *Legionella pneumophila* Cas2 promotes intracellular infection of amoebal host cells. *Infect Immun* **83**, 1008–1018.

[33] Rapp CS and Pollack RM (2005) Crystal packing effects on protein loops. *Proteins* **60**, 103–109.

[34] Thompsona HPG and Day GM (2014) Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chem Sci* **5**, 3173–3182.

Abstract in Korean

Xanthomonas albilineans Cas2 단백질의 수용액 구조 및 다이나믹스 분석을 통한 핵산분해효소 기능 이해

잔토모나스균 (*Xanthomonas*) 종의 CRISPR/Cas 면역 시스템에 관여하는 Cas2 단백질은 외부에서 침입한 DNA 및 RNA와 같은 유전물질을 인식하는 데 주로 Scaffold 역할만을 하는데 에도 불구하고, 핵산분해효소 기능을 소유하고 있다. 현재까지 규명 된 X-선 기반의 Cas2 구조들은 모두 효소 기능에 필요한 2가 금속 이온을 수용할 수 없는 비활성 상태로, 필요에 따라 변화가 용이한 Cas2의 변화 메커니즘을 설명하지 못한다. 본 연구에선, 잔토모나스 알빌리닌스 (*Xanthomonas albilineans*) Cas2의 수용액상 구조와 다이나믹스를 Residual dipolar couplings와 NMR relaxation 실험을 통해, 수용액상에서 또한 비활성 상태를 띄는 Cas2 구조를 규명하였으며, 금속 이온 결합 부위인 Asp8 아미노산 부근과 C-terminal 부근이 매우 역동적인 것을 관찰하였다. 이에 Cas2 단백질이 C-terminal 부근의 Hinge-bending motion을 원동력으로 때에 따른 구조적 변화를 보임을 직접적인 실험 결과를 통해 규명하였다. 본 연구는, 때에 따라 Cas2 의 기능을 제어하는 응용 기술의 기반을 구축한 연구이다.

주요어 : Cas2 | nuclease activity | structure and dynamics | NMR
학번 : 2015-22696