



# Structural Flexibility of Peripheral Loops and Extended C-terminal Domain of Short Length Substrate Binding Protein from *Rhodothermus marinus*

Ji-Eun Bae<sup>1,2</sup> · In Jung Kim<sup>3,7</sup> · Yongbin Xu<sup>4,5</sup> · Ki Hyun Nam<sup>6</sup>

Accepted: 19 February 2021 / Published online: 2 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

## Abstract

Substrate binding proteins (SBPs) bind to specific ligands in the periplasmic regions of cells and then bind to membrane proteins to participate in transport or signal transduction. Typically, SBPs consist of two  $\alpha/\beta$  domains and recognize the substrate by a flexible hinge region between the two domains. Conversely, the short-length SBPs are often observed in protein databases, which are located around methyl-accepting chemotaxis protein genes. We previously determined the crystal structure of *Rhodothermus marinus* SBP (named as RmSBP), consisting of a single  $\alpha/\beta$  domain; however, the substrate recognition mechanism is still unclear. To better understand the functions of short length RmSBP, we performed a comprehensive study, involving comparative structure analysis, computational substrate docking, and X-ray crystallographic data. RmSBP shares a high level of similarity in the  $\alpha/\beta$  domain region with other SBPs, but it has a distinct topology in the C-terminal domain. The substrate binding model suggested that conformational changes in the peripheral region of RmSBP was required to recognize the substrate. We determined the crystal structures of RmSBP at pH 5.5, 6.0, and 7.5. RmSBP showed structural flexibility in the  $\beta 1$ – $\alpha 2$  loop,  $\beta 5$ – $\beta 6$  loop, and extended C-terminal domains, based on the electron density map and temperature B-factor analysis. These results provide information that will further our understanding on the functions of the short length SBP.

**Keywords** Substrate-binding protein · SBP · ABC transport ·  $\alpha/\beta$  domain · *Rhodothermus marinus* · SARS-CoV-2

## 1 Introduction

Substrate binding proteins (SBPs) initially recognize their substrates in the periplasmic space and deliver the substrates to membrane-bound subunits that catalyze concentrative uptake into cells [1–3]. SBP is part of a family of proteins consisting of ATP-binding cassette (ABC)-transporters for substrate uptake, ion-gradient driven transporters, DNA binding proteins, as well as, prokaryotic and eukaryotic channels and receptors [2–5]. In 1999, SBPs were classified based on their sequence similarities and topological arrangements in their  $\beta$ -sheet region [6]. Recently, their classification has been updated into seven cluster groups based on a number of SBP structures deposited in Protein Data Bank (PDB), with each cluster having different structural characteristics [7]. These classified SBPs are involved in a unique molecular mechanism of the functioning of transporters, channels, and signal transducers [7]. In the protein database, SBPs vary in size from approximately 25–70 kDa [5, 8]. These SBPs have low sequence similarity but a highly conserved overall three-dimensional structural fold [5]. The core of SBPs

✉ Ki Hyun Nam  
structures@postech.ac.kr

<sup>1</sup> School of Life Sciences, KNU Creative BioResearch Group, Kyungpook National University, Daegu 41566, Republic of Korea

<sup>2</sup> KNU Institute for Microorganisms, Kyungpook National University, Daegu 41566, Republic of Korea

<sup>3</sup> Division of Biotechnology, Korea University, Seoul 02841, Republic of Korea

<sup>4</sup> Department of Bioengineering, College of Life Science, Dalian Minzu University, Dalian 116600, Liaoning, China

<sup>5</sup> Key Laboratory of Biotechnology and Bioresources Utilization of Ministry of Education, Dalian Minzu University, Dalian 116024, China

<sup>6</sup> Department of Life Science, Pohang University of Science and Technology, Pohang 37673, Republic of Korea

<sup>7</sup> Present Address: Department of Biotechnology and Enzyme Catalysis, Institute of Biochemistry, University of Greifswald, 17487 Greifswald, Germany

consist of two structural  $\alpha/\beta$  domains which are connected by a flexible hinge region [8]. This shows that the unique architecture depends on the classified SBP cluster [7]. Substrate binding occurs at the flexible hinge region located between the two domains of SBP, stabilizing the closed form of the tightly packed protein with its substrate buried at the interface [5]. Typically, SBP is present in four structural states during the process of substrate recognition: (i) open-unliganded, (ii) open-liganded, (iii) closed-unliganded, and (iv) closed-liganded [5].

*Rhodothermus marinus* is a thermohalophilic bacterium that grows optimally at 65 °C [9]. We previously characterized the short length *Rhodothermus marinus* SBP (named as RmSBP) consisting of 138 amino acids, excluding the signal peptide [10]. This SBP gene is located around methyl-accepting chemotaxis protein (MCP) gene, which is composed of a single peptide, transmembrane, HAMP, and methyl-accepting transducer regions [10]. A similar feature of SBP-MCP gene cluster from *R. marinus* is also found in *Rhodothermus profundus*, *Rhodothermaceae bacterium* RA, *Salinibacter ruber* strain DSM 13855, and *Salinibacter ruber* strain M8 [10], indicating that short length SBP often exists in the nature along with its counterpart MCP protein. We previously determined the crystal structure of RmSBP at pH 4.5 and identified the presence of a single  $\alpha/\beta$  domain [10]. RmSBP had a high structural similarity with C-terminal domain of *Streptococcus pneumoniae* SBP (PDB code: 3LFT, r.m.s. deviation of 2.3 Å for 149 C $\alpha$ -atoms, named as SpSBP) and *Vibrio cholerae* serotype O1 SBP (3LKV, 2.5 Å for 149 C $\alpha$ -atoms, named as VcSBP). SpSBP and VcSBP structures interact with L-tryptophan and L-phenylalanine amino acids, respectively. The residues that recognize these amino acids are not conserved in RmSBP [10]. Although the structural features of RmSBP have been analyzed, the mechanism by which they recognize substrates is still unknown.

To better understand the substrate recognition of short length RmSBP, we performed a study combining comparative structural analysis, computational substrate docking, and X-ray crystallographic data. We described the topology between RmSBP and other SBPs and modeled potential substrate binding sites. The crystal structures of RmSBP at pH 5.5, 6.0, and 7.5 were determined at 1.5, 1.8, and 1.9 Å resolution, respectively. The structural flexibility present in peripheral  $\beta 1$ – $\alpha 2$  loop and  $\beta 5$ – $\beta 6$  loop of RmSBP, as well as, in extended C-terminal regions was observed. Our results provide the beginning framework to understand the molecular functions of short length SBP.

## 2 Materials and Methods

### 2.1 Comparative and Computational Analysis

The crystal structure of RmSBP at pH 4.5 (PDB code 5Z6V) was used as a starting point for the homolog search and substrate prediction study. The obtained homolog models were searched and evaluated using Phyre2 server [11]. TM (template model) scores were also obtained using Phyre2 server. The putative substrate binding site were predicted using 3DLigandSite server [12].

### 2.2 Protein Expression and Purification

Detailed protocols for cloning and expression of proteins have been reported in a previous study [10]. Briefly, *RmSBP* gene, excluding the signal peptide, was cloned into pET28 vector and expressed in *E. coli* BL21 (DE3) cells. Purified recombinant RmSBP was obtained by a two-step purification process using a Ni–NTA affinity column and size exclusion chromatography. The final purified protein was stored in 10 mM Tris–HCl, pH 8.0 and 200 mM NaCl.

### 2.3 Crystallization

Purified RmSBP were concentrated to 20 mg/mL using Centricon (Millipore, 10 kDa cutoff). Crystallization screens were performed using the sitting-drop vapor diffusion method at 20 °C using commercial crystallization kits. Briefly, 0.3  $\mu$ L protein solution was mixed with 0.3  $\mu$ L precipitant solution and equilibrated against 70  $\mu$ L precipitant solution. Microcrystals were obtained by following 3 different conditions: (i) 0.1 M Bis-Tris, pH 5.5, 0.2 M MgCl<sub>2</sub>, and 25% (w/v) polyethylene glycol 3350, (ii) 0.1 M MES, pH 6.0 and 1.26 M ammonium sulfate, and (iii) 0.1 M HEPES, pH 7.5 and 25% (w/v) polyethylene glycol 3350. Suitable crystals for X-ray diffraction were obtained using the sitting-drop vapor diffusion method at 20 °C by mixing 1.5  $\mu$ L protein solution and 1.5  $\mu$ L precipitant solution and then equilibrated against 200  $\mu$ L reservoir solution with crystallization solutions mentioned above.

### 2.4 Diffraction Data Collection

X-ray diffraction data for RmSBP crystals were collected at 100 K on beamline 7A at Pohang Light Source II (PLS-II, Korea) [13]. All crystals were equilibrated in a cryoprotectant solution containing reservoir supplemented with 20% (v/v) ethylene glycol and then flash-cooled in a stream of liquid nitrogen. For the amino acid soaking experiment, RmSBP crystals were soaked in a cryoprotectant solution

supplemented with an amino acid mixture 1 min before data collection. The diffraction images were indexed, integrated, and scaled with the HKL2000 package [14]. The data collection statistics are listed in Table 1.

## 2.5 Structure Determination

The initial phases of RmSBPs were solved using the molecular replacement method by Phaser-MR in Phenix [15] with selenium-derived RmSBP at pH 4.5 (PDB code: 5Z6V) [10] as a search model. Manual model building was performed with COOT program [16]. Model refinement was performed with Refmac5 [17] and Phenix refinement in Phenix [18]. The geometry of refined model was evaluated using Mol-Probity server [19]. The structure refinement statistics are listed in Table 1. Figures were generated with the PyMOL [20]. Structure factors and coordinates have been deposited in the Protein Data Bank under PDB codes 6K1W (pH 5.5), 6K1X (pH 6.0) and 6K1Y (pH 7.5).

## 3 Results

### 3.1 Computational Analysis of RmSBP

All SBP structures reported to date recognize their substrates by their flexible hinge region present between the two structural domains [7], whereas RmSBP has a single domain [10]. We previously suggested that these RmSBPs can recognize their substrates alone or with the help of other partner proteins [10]. In both the cases, we hypothesized that structural changes in the peripheral regions of RmSBP were required to recognize the substrate molecule. To better understand the substrate recognition of RmSBP, we performed comparative analysis and substrate docking studies using a previously reported crystal structure of SBP (PDB code 5Z6V) as the initial model structure. The analysis using Phyre2 server provided 19 models which similar to RmSBP and also provided the expected 19 substrate binding sites. Among them, 9 models (PDB codes: 3LFT, 2QH8, 5ER3, 4RS3, 4KZK,

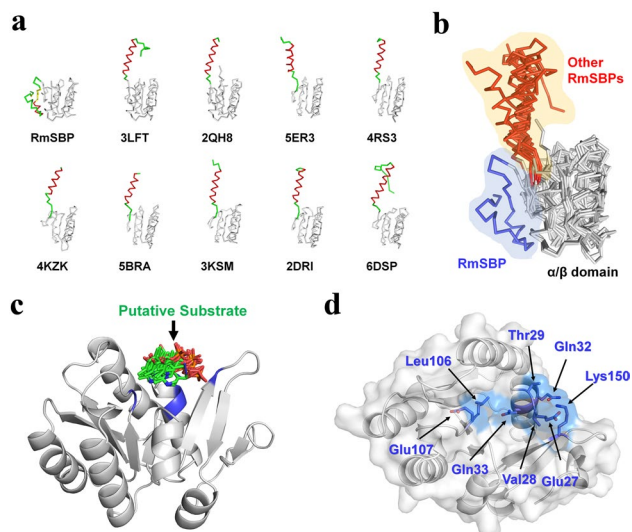
**Table 1** Data collection and refinement statistics

Data collection	pH 5.5	pH 6.0	pH 7.5
Resolution	39.29–1.50 (1.53–1.50)	20.0–1.8 (1.87–1.8)	50.0–1.90 (1.93–1.90)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Unique reflections	21,337	12,659	10,511
Unit cell parameter			
a, b, c (Å)	46.42, 48.61, 58.11	34.91, 35.72, 115.57	46.54, 48.60, 57.82
Completeness (%)	99.2 (97.7)	90.5 (95.0)	95.5 (91.3)
Multiplicity	5.4 (4.9)	4.8 (4.9)	7.6 (4.2)
<i>I</i> / $\sigma$ ( <i>I</i> )	40.12 (5.57)	27.60 (6.20)	44.74 (4.09)
R <sub>merge</sub>	0.065 (0.349)	0.118 (0.425)	0.101 (0.480)
R <sub>pim</sub>	0.031 (0.170)	0.057 (0.198)	0.034 (0.245)
CC <sub>1/2</sub>	1.000 (0.888)	0.994 (0.904)	0.995 (0.499)
Refinement			
Resolution	37.29–1.50	14.94–1.80	37.20–1.90
R <sub>work</sub> <sup>a</sup>	0.185	0.190	0.178
R <sub>free</sub> <sup>b</sup>	0.218	0.220	0.228
Average B factor (Å <sup>2</sup> )			
Protein	22.15	22.95	35.58
R.m.s.deviations			
Bonds	0.019	0.007	0.014
Angles	1.983	1.203	1.680
Ramachandran			
Preferred	98.66	98.7	98.7
Allowed	1.34	1.30	1.3
PDB	6K1W	6K1X	6K1Y

Values in the parentheses refers to the highest resolution shell

<sup>a</sup>R<sub>work</sub> =  $\sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$ , where  $F_{\text{obs}}$  and  $F_{\text{calc}}$  are the observed and calculated structure-factor amplitudes respectively.

<sup>b</sup>R<sub>free</sub> was calculated as R<sub>work</sub> using a randomly selected subset (10%) of unique reflections not used for structure refinement.



**Fig. 1** Computational analysis of RmSBP. **a** Comparative analysis on topologies of RmSBP and other SBPs.  $\alpha/\beta$  domain is indicated by a grey ribbon.  $\alpha$ -helix,  $\beta$ -strands and loops in the C-terminal domain are represented by red, yellow, and green ribbons, respectively. **b** Superimposition of RmSBP with other structural homolog SBPs. The C-terminal domain of RmSBP and other SBPs are represented by blue and red ribbons, respectively. **c** Computational prediction of the substrate binding to RmSBP. Total 15 PLP molecules were used as model substrates and have been placed on  $\alpha/\beta$  domain of RmSBP. A cartoon representation of the predicted substrate recognition loops is also shown here (colored blue). **d** Predicted substrate recognition residues are shown as blue sticks (Color figure online)

**Table 2** Models used for structural comparison

	Confidence	Sequence identity	RMSD	TM-score
5Z6V	100			1.00
3LFT	99.8	19	2.018	0.66
2QH8	99.8	23	2.024	0.68
5ER3	98.1	13	2.174	0.49
4RS3	98.1	16	2.150	0.47
4KZK	98.1	13	1.760	0.52
5BRA	98.0	14	2.177	0.48
3KSM	97.9	12	2.122	0.52
2DRI	97.9	16	1.924	0.51
6DSP	97.8	15	2.130	0.52

5BRA, 3KSM, 2DRI, and 6DSP) with TM-scores of  $>0.52$  were used in this study, which were all either amino acid or sugar binding SBPs (Fig. 1a and Table 2). These models consisted of  $\alpha/\beta$  fold structures and had low sequence identity of 12–23% with RmSBP. The superimposition of RmSBP with other SBPs showed similarities in the core  $\alpha/\beta$  domain consisting of 73–119 C $\alpha$  atoms (with r.m.s. deviation of 1.753–2.133 Å), whereas the topology of C-terminal

residues showed differences in conformation (Fig. 1a). In RmSBP structure,  $\beta_6$ -strand in extended C-terminal domain formed an antiparallel  $\beta$ -sheet with  $\beta_5$ -strand of the  $\alpha/\beta$  domain core. In addition,  $\alpha_5$ - and  $\alpha_6$ -helices of RmSBP flanked their  $\alpha/\beta$  domain. In contrast, the C-terminal regions of other SBPs were present upward in the direction of the substrate binding site (Fig. 1a and b), which further help in the binding with its partner  $\alpha/\beta$  domain for target substrate recognition. RmSBP not only showed structural similarity to an SBP from a non-thermophilic bacteria as per the Phyre2 server, but also showed structural similarity with the C-terminal domain of an SBP from the thermophilic *Aeropyrum pernix* (Supplementary Fig. 1). Therefore, the extended C-terminal domain of RmSBP had significantly distinct topology against other typical SBPs. Next, the prediction of substrate binding site was performed using 3DLigandSite software, which displayed 15 binding sites on the surface of  $\alpha/\beta$  domain using PLP (pyridoxal phosphate) as a model substrate. Results showed that N-terminus (Glu27, Val28, and Thr28),  $\alpha_1$ -helix (Gln32 and Gln33),  $\beta_3$ – $\alpha_4$  loop (Leu106 and Glu107), and  $\beta_5$ – $\beta_6$  loop (Lys150) residues in RmSBP were predicted to be substrate binding sites (Fig. 1c). Among these residues, the conformational changes in side chains of Gln32 and Gln33 are probable, but large conformation changes in the whole protein were difficult since the main chain was present as a stable helix formation. Leu106 and Glu107 residues, on the other hand, were located in a sharp turn of  $\beta_3$ – $\alpha_4$  loop, thus limiting large conformational changes. Based on the computational docking study, we considered that non-structural N-terminus and  $\beta_5$ – $\beta_6$  loop of RmSBP could show certain amount of structural flexibility.

### 3.2 Crystal Structures of RmSBPs

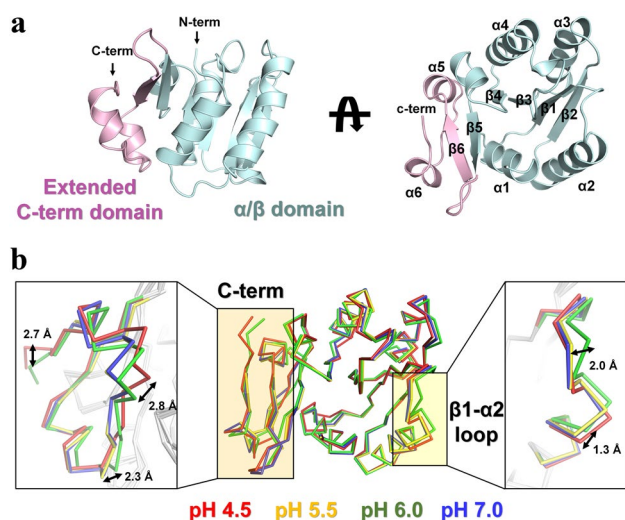
Our computational substrate docking study suggested that RmSBP might have structural flexibility in the peripheral loop region of  $\alpha/\beta$  domain and C-terminal domain. However, in the previously reported structure of RmSBP at pH 4.5, the peripheral loops and  $\alpha/\beta$ -fold of RmSBP exhibited a highly rigid structure (see below). As a result, there was no experimental evidence to prove the structural flexibility of RmSBP by computational analysis. As a proof for computational analysis results, we performed an extended crystallographic study to observe structural flexibility in the peripheral loops on the substrate recognition surface of RmSBP. We obtained RmSBP crystals at pH 5.5, 6.0, and 7.5 with different crystallization conditions. Crystals of RmSBP at pH 5.5 and 7.5 belonged to the orthorhombic space group  $P2_12_12_1$ , with a similar unit-cell dimension of approximately  $a=46$  Å,  $b=48$  Å and  $c=58$  Å and occupying one molecule in the asymmetric unit (Table 1). Crystal of RmSBP at pH 6.0 belonged to the orthorhombic space

group  $P2_122_1$ , with unit cell dimension of 34 Å, 35 Å, and 115 Å and occupying one molecule in the asymmetric unit (Table 1). In contrast, the previously reported RmSBP crystal at pH 4.5 belonged to the monoclinic space group  $C2_1$  [10]. The structures of RmSBP at pH 5.5, 6.0, and 7.5 were refined up to 1.5 Å, 1.8 Å, 1.9 Å resolutions, respectively, and produced  $R_{\text{work}}/R_{\text{free}}$  of 18.5%/21.8%, 19.0%/22.0%, and 17.8%/22.8%, respectively. All RmSBP structures at pH 5.5, 6.0, and 7.5 were composed of six  $\alpha$ -helices and six  $\beta$ -strands, and formed an  $\alpha/\beta$  fold with an extra domain at C-terminal region (Fig. 2a). Detailed structural topology information describing RmSBP structure has been previously reported [10]. Here, we described a novel finding for the peripheral flexible regions of RmSBP. Previously, we had classified RmSBP structure into a single  $\alpha/\beta$  domain [10]. However, in the present study, RmSBP was newly classified into an  $\alpha/\beta$  domain ( $\alpha 1$ – $\alpha 4$  and  $\beta 1$ – $\beta 5$ ) and a C-terminal extended domain ( $\alpha 5$ ,  $\alpha 6$ , and  $\beta 6$ ) through comparative analysis with other homologous SBP structures (Figs. 1a and 2a).

For all RmSBP structures obtained at pH 5.5, 6.0, and 7.5, electron density maps of core of  $\alpha/\beta$  domain of RmSBP was well defined, but the peripheral loop and C-terminus were disordered or structurally flexible. In RmSBP-pH 5.5, the electron density map for four amino acids (Asp151, Ala152, Glu153, and Gly154) in the  $\beta 4$ – $\beta 5$  loop and two residues (Asp177 and Arg178) in the  $\alpha 6$ -helix region on the C-terminal domain were disordered. In RmSBP-pH 6.0, the electron density map for three residues (Asp151, Ala152, Glu153) at  $\beta 4$ – $\beta 5$  loop were disordered. In RmSBP-pH

7.5, although all the residues were fitted into the electron density map without any disorder, however, structural flexibility with a high B-factor value was observed (see below). The superimposition of crystal structures of RmSBP at pH 5.5, pH 6.0, and pH 7.5 with a previously reported crystal structure of RmSBP at pH 4.5 showed similarity in all C $\alpha$  atoms (with r.m.s. deviation of 0.7196–1.3262), but two significantly different conformations were observed. In the  $\beta 1$ – $\alpha 2$  loop region, C $\alpha$  atoms at pH 6.0 in the loop portion of the  $\beta 1$ – $\alpha 2$  loop were shifted by about 2.0 Å compared to C $\alpha$  atoms in structures obtained at other pHs, and by about 1.3 Å in the  $\alpha 2$ -helix region (Fig. 2b). On the other hand, the extended C-terminal domain was not structurally aligned (Fig. 2b) and was shifted by 2.8 Å in the  $\alpha 5$ -helix region, and by 2.3 Å and 2.7 Å in the loop and C-terminal region, respectively (Fig. 2b).

Subsequently, we performed temperature B-factor analysis on all RmSBP structures (Fig. 3). At pH 4.5, RmSBP showed a rigid fold with a low B-factor value, except in the N-terminus (Fig. 3a). At pH 5.5, B-factor of RmSBP was relatively high at  $\beta 1$ – $\alpha 2$  loop region present on  $\alpha/\beta$  domain, and C-terminal domain was not built due to lack of electron density map in that region (Fig. 3a). At pH 6.0, the overall  $\alpha/\beta$  domain of RmSBP showed high rigidity, but it also displayed high B-factor value at the  $\alpha 6$ -helix region in C-terminal domain (Fig. 3a). At pH 7.5, there were no disordered regions in the electron density map of RmSBP, which was similar to RmSBP structure at pH 4.5, but a relatively high flexibility was observed in  $\beta 1$ – $\alpha 2$  loop and C-terminal domain (Fig. 3a). The analysis of normalized C $\alpha$  atom B-factor values for the four RmSBP domains showed that residues in  $\beta 1$ – $\alpha 2$  loop,  $\beta 5$ – $\beta 6$  loop and C-term helix had a relatively higher flexibility when compared with other residues (Fig. 3b). On the other hand, in RmSBP structure, the portion of the electron density map with disordered or relatively high B-factor area did not have a structural change in the proportion of the acid or basic pH concentrations. Next, we analyzed the electrostatic surfaces of the structures of RmSBP at pH 5.5, 6.0, and 7.5 (Fig. 3c). The surface charges on the C-terminal region of RmSBP are observed differently because there is some amount of disorder in that region. On the other hand, the putative substrate binding site has a negative charge in common, and no specific differences were found between the three structures (Fig. 3c). As a result, the potential substrate binding sites are well preserved, while the peripheral loops of RmSBP are shown to be flexible.

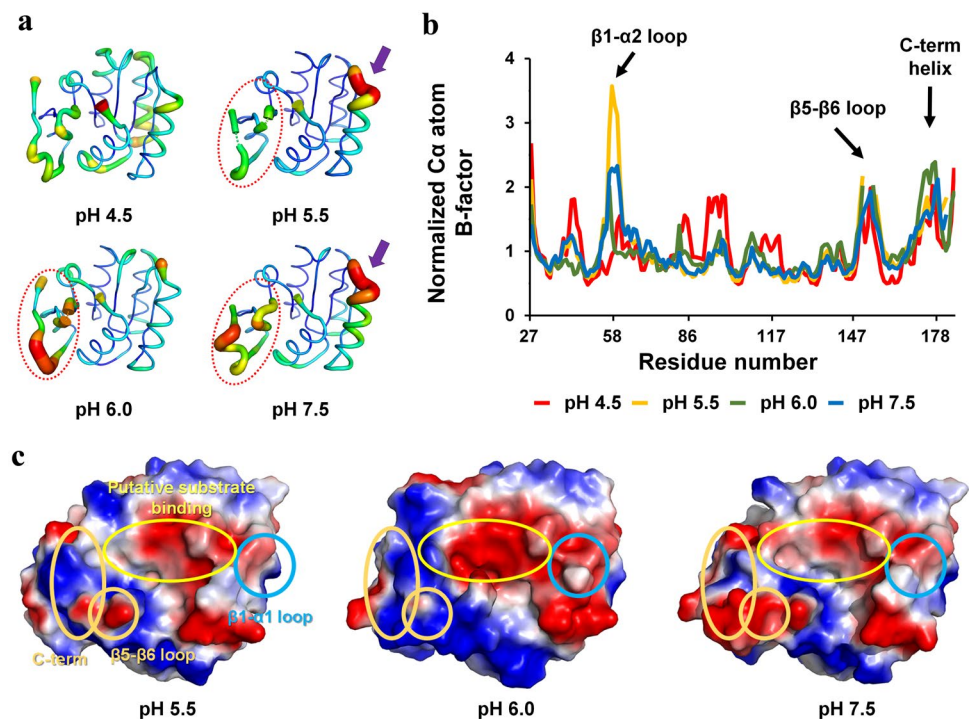


**Fig. 2** Crystal structures of RmSBPs. **a** The overall structure of RmSBP (pH 7.5) consists of an  $\alpha/\beta$  core domain (cyan) and an extended C-terminal domain (pink). **b** Superimposition of RmSBPs at pH 4.5 (red), 5.5 (yellow), 6.0 (green), and 7.5 (blue) is shown here. Conformational differences of RmSBP are observed at  $\beta 1$ – $\alpha 2$  loop (yellow transparency) and C-terminal domain (orange transparency) (Color figure online)

## 4 Discussion

We performed a comparative, computational and structural analysis of RmSBP for comprehensive understanding of the molecular functions involved in this short length SBP.

**Fig. 3** Analysis of flexible region of RmSBPs. **a** B-factor representation of RmSBPs at pH 4.5, 5.5, 6.0, and 7.5. **b** A plot showing normalized B-factor values of C $\alpha$  atoms of RmSBPs at pH 4.5 (red), 5.5 (orange), 6.0 (green), and 7.5 (blue). **c** Electrostatic surface of structures of RmSBP at pH 5.5, 6.0, and 7.5. Putative substrate binding sites commonly exhibited a negative charge (Color figure online)

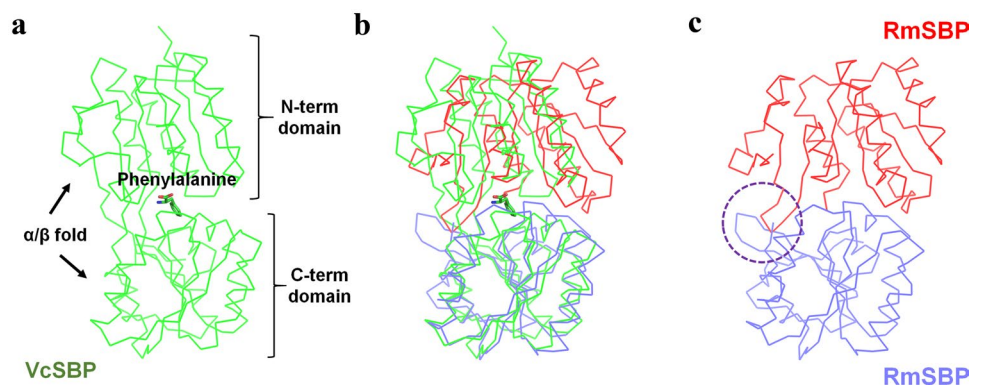


We previously described that RmSBP consisted of a single  $\alpha/\beta$  domain, however, in this study, we divided the RmSBP structure into a  $\alpha/\beta$  domain and an extended C-terminal domain based on the comparative structural analysis. In particular, the C-terminal domain represented a unique topology in which RmSBP was distinguished from other SBPs. In the substrate binding model, four substrate binding sites were predicted at positions essential for the conformational changes required to recognize the substrate. However, since previously determined RmSBP at pH 4.5 exhibited a highly rigid structure and a low B-factor value, there was no experimental evidence on whether the actual RmSBP was structurally flexible. To observe the structural flexibility of RmSBP, we crystallized and determined three crystal structures of RmSBP through three new crystallization conditions that

had not been reported previously. The structural flexibility of  $\beta 1$ – $\alpha 2$  loop,  $\beta 5$ – $\beta 6$  loop, and C-terminal helix region of RmSBP might provide the initial framework for structural studies on short length SBP, as well as other SBPs.

The N- and C-terminal domains of SpSBP and VcSBP have an  $\alpha/\beta$  fold in common (Fig. 4a). Therefore, we superimposed the structures to determine the similarity between the  $\alpha/\beta$  fold of RmSBP and the N-terminal domain of SpSBP and VcSBP (Fig. 4b). However, unlike the C-terminal domain of VcSBP, no similarity was found in the N-terminal domain (Fig. 4b). These results suggest that there is no amino acid sequence similarity; of note, it was difficult to find structural similarity due to differences in the length and direction of the  $\beta$ -stand and the length of the helix in  $\alpha/\beta$  fold. On the other hand, when the structure

**Fig. 4** Comparison and superimposition of RmSBP and VcSBP. **a** Ribbon representation of VcSBP consisting of two  $\alpha/\beta$  domains. **b** Superimposition of two RmSBPs to N- and C-terminal  $\alpha/\beta$  domain of VcSBP. **c** Dimeric model of RmSBP. Two RmSBP molecules superimposed to N- and C-terminal domains of VcSBP. The molecule overlapped region in dimeric RmSBP model is indicated by dot-circle



of RmSBP was superimposed on the N- and C-terminal domains of VsSBP, molecular overlap occurred between the two RmSBPs in dimeric formation (Fig. 4c). This said, it is considered that even if the RmSBP exists as a homodimer, it will have a different conformation from those of existing SBPs; therefore, large structural changes will be required for substrate recognition. In fact, RmSBP showed structural similarity to the C-terminal domains of both SpSBP and VcSBP. In their crystal structures, SpSBP and VcSBP are complexed with tryptophan and phenylalanine, respectively; of note, these amino acids interact with Asn and Ile at the same position in the structure. Meanwhile, in RmSBP, Glu107 and Ser109 were placed in the same positions. As a result, we expect that even if RmSBP recognizes amino acids, they will be different types of amino acid substrates. Here, to investigate whether RmSBP was able to recognize amino acids, the RmSBP crystal was soaked in a cryoprotectant solution containing an amino acid mixture and diffraction data was collected. However, no electron density map, suggestive of an amino acid, was found at the RmSBP potential substrate-binding. This result indicates that RmSBP does not recognize amino acids or does not have a high affinity for them; other partner molecules may be needed for the recognition of amino acids. Of note, the amino acid binding sites of SpSBP and VcSBP have a hydrophobic surface charge, whereas the potential substrate-binding site of RmSBP is a negatively charged surface. Therefore, these data suggest that RmSBP recognizes a charged substrate distinct from those recognized by SpSBP or VcSBP.

Although our study provides important information for understanding the structural properties of a short length SBP, further biochemical experiments on a variety of potential SBP substrates need to be performed to understand their exact biological functions. In this regard, not only identifying the substrates for RmSBP proteins, but also an in-depth study of their partner proteins and their functional relevance with the MCP-bound RmSBP, is required. The crystal structures of RmSBP, in this study, will definitely differ from that of RmSBP at 65 °C. To better understand the molecular flexibility of RmSBP, we believe that the equilibrium molecular dynamics simulations at growth temperatures are definitely worthwhile. Of note, RmSBP shows sequence similarity of 20% with the ORF1ab polyprotein of SARS-CoV-2 (strain SARS-CoV-2\_HKU-SZ-001\_2020), and was, therefore, recently used as a template for modeling structures [21]. Although the sequence similarity of amino acids is not high between RmSBP and ORF1ab polyprotein in SARS-CoV-2, our structural results for RmSBP will help to understand the model structure of ORF1ab of SARS-CoV-2.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10930-021-09970-z>.

**Acknowledgements** We thank the beamline staff at the MX beamlines at PLS-II at Pohang Accelerator Laboratory for assistance in data collection. This research was funded by the National Research Foundation of Korea (Grant Nos. NRF-2017R1D1A1B03033087 and NRF-2017M3A9F6029736). This work was supported by MSIT and PAL (Korea).

**Author Contributions** JEB performed cloning, and protein preparation. IJK performed protein preparation, and crystallization. YX analyzed the structures, and revised manuscript. KHN performed data collection, structure determination, and wrote the manuscript.

## Declarations

**Conflict of interest** The authors declare no conflict of interest.

## References

1. Maqbool A, Horler RS, Muller A, Wilkinson AJ, Wilson KS, Thomas GH (2015) The substrate-binding protein in bacterial ABC transporters: dissecting roles in the evolution of substrate specificity. *Biochem Soc Trans* 43(5):1011–1017. <https://doi.org/10.1042/BST20150135>
2. van der Heide T, Poolman B (2002) ABC transporters: one, two or four extracytoplasmic substrate-binding sites? *EMBO Rep* 3(10):938–943. <https://doi.org/10.1093/embo-reports/kvf201>
3. Higgins CF (1992) ABC transporters: from microorganisms to man. *Annu Rev Cell Biol* 8:67–113. <https://doi.org/10.1146/annurev.cb.08.110192.000435>
4. Tam R, Saier MH Jr (1993) Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol Rev* 57(2):320–346
5. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B (2010) A structural classification of substrate-binding proteins. *FEBS Lett* 584(12):2606–2617. <https://doi.org/10.1016/j.febslet.2010.04.043>
6. Fukami-Kobayashi K, Tateno Y, Nishikawa K (1999) Domain dislocation: a change of core structure in periplasmic binding proteins in their evolutionary history. *J Mol Biol* 286(1):279–290. <https://doi.org/10.1006/jmbi.1998.2454>
7. Scheepers GH, Lycklama ANJA, Poolman B (2016) An updated structural classification of substrate-binding proteins. *FEBS Lett* 590(23):4393–4401. <https://doi.org/10.1002/1873-3468.12445>
8. Quijcho FA, Ledvina PS (1996) Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol Microbiol* 20(1):17–25. <https://doi.org/10.1111/j.1365-2958.1996.tb02484.x>
9. Bjornsdottir SH, Blondal T, Hreggvidsson GO, Eggertsson G, Petursdottir S, Hjorleifsdottir S, Thorbjarnardottir SH, Kristjansson JK (2006) *Rhodothermus marinus*: physiology and molecular biology. *Extremophiles* 10(1):1–16. <https://doi.org/10.1007/s00792-005-0466-z>
10. Bae JE, Kim IJ, Kim KJ, Nam KH (2018) Crystal structure of a substrate-binding protein from *Rhodothermus marinus* reveals a single alpha/beta-domain. *Biochem Biophys Res Commun* 497(1):368–373. <https://doi.org/10.1016/j.bbrc.2018.02.086>
11. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc* 10(6):845–858. <https://doi.org/10.1038/nprot.2015.053>
12. Wass MN, Kelley LA, Sternberg MJ (2010) 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acids Res* 38:W469–473. <https://doi.org/10.1093/nar/gkq406>

13. Park SY, Ha SC, Kim YG (2017) The protein crystallography beamlines at the Pohang light source II. *BioDesign* 5(1):30–34
14. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326
15. Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D* 66(Pt 2):213–221. <https://doi.org/10.1107/S0907444909052925>
16. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60(Pt 12):2126–2132. <https://doi.org/10.1107/S0907444904019158>
17. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D* 67(Pt 4):355–367. <https://doi.org/10.1107/S0907444911001314>
18. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D* 68:352–367. <https://doi.org/10.1107/S0907444912001308>
19. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D* 66(Pt 1):12–21. <https://doi.org/10.1107/S0907444909042073>
20. The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.
21. Shanker AK, Bhanu D, Alluri A, Gupta S (2020) Whole-genome sequence analysis and homology modelling of the main protease and non-structural protein 3 of SARS-CoV-2 reveal an aza-peptide and a lead inhibitor with possible antiviral properties. *New J Chem* 44(22):9202–9212. <https://doi.org/10.1039/d0nj00974a>

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