

# Expression and Purification of BsaXI Restriction Endonuclease and Engineering New Specificity From BsaXI Specificity Subunit

Sonal Gidwani<sup>†‡</sup>, Daniel Heiter<sup>‡</sup> and Shuang-yong Xu<sup>\*</sup>

New England Biolabs, Inc., Ipswich, MA, United States

#### **OPEN ACCESS**

#### Edited by:

Gülay Özcengiz, Middle East Technical University, Turkey

#### Reviewed by:

Iwona Mruk, University of Gdańsk, Poland Dmitri Nikitin, Lomonosov Moscow State University, Russia

> \*Correspondence: Shuang-yong Xu

#### <sup>†</sup>Present address:

xus@neb.com

Sonal Gidwani, Calder Biosciences, Inc., Brooklyn, NY, United States

<sup>‡</sup>These authors have contributed equally to this work

#### Specialty section:

This article was submitted to Microbiotechnology, a section of the journal Frontiers in Microbiology

Received: 02 March 2022 Accepted: 24 March 2022 Published: 19 May 2022

#### Citation:

Gidwani S, Heiter D and Xu S-y (2022) Expression and Purification of BsaXI Restriction Endonuclease and Engineering New Specificity From BsaXI Specificity Subunit. Front. Microbiol. 13:8884355. doi: 10.3389/fmicb.2022.888435 It is stated that BsaXI is a Type IIB restriction endonuclease (REase) that cleaves both sides of its recognition sequence 5' N9 AC N5 CTCC N10 J 3' (complement strand 5' 1N7 GGAG N5 GT N121 3'), creating 3-base 3' overhangs. Here we report the cloning and expression of bsaXIS and bsaXIRM genes in Escherichia coli. The BsaXI activity was successfully reconstituted by mixing the BsaXI RM fusion subunit with the BsaXI S subunit and the enzyme complex further purified by chromatography over 6 columns. As expected, the S subunit consisted of two subdomains encoding TRD1-CR1 [target recognition domain (TRD), conserved region (CR)] for 5' AC 3', and TRD2-CR2 presumably specifying 5' CTCC 3'. TRD1-CR1 (TRD2-CR2 deletion) or duplication of TRD1 (TRD1-CR1-TRD1-CR2) both generated a new specificity 5' AC N5 GT 3' when the S variants were complexed with the RM subunits. The circular permutation of TRD1 and TRD2, i.e., the relocation of TRD2-CR2 to the N-terminus and TRD1-CR1 to the C-terminus generated the same specificity with the RM subunits, although some wobble cleavage was detected. The TRD2 domain in the BsaXI S subunit can be substituted by a close homolog ( $\sim$ 59% sequence identity) and generated the same specificity. However, TRD2-CR2 domain alone failed to express in E. coli, but CR1-TRD2-CR2 protein could be expressed and purified which showed partial nicking activity with the RM subunits. This work demonstrated that like Type I restriction systems, the S subunit of a Type IIB system could also be manipulated to create new specificities. The genome mining of BsaXI TRD2 homologs in GenBank found more than 36 orphan TRD2 homologs, implying that quite a few orphan TRD2s are present in microbial genomes that may be potentially paired with other TRDs to create new restriction specificities.

Keywords: Type IIB restriction endonuclease, BsaXI specificity, RM fusion and S subunits, circular permutation of BsaXI S subunit, TRD1-CR1

## INTRODUCTION

It is stated that *Bsa*XI is a Type IIB *Bcg*I-like restriction endonuclease (REase) that cleaves outside of its recognition sequence 5'  $\downarrow$ N9 AC N5 CTCC N10 $\downarrow$  3' (complement strand 5'  $\downarrow$ N7 GGAG N5 GT N12 $\downarrow$  3'), creating 3-base 3' overhangs (Roberts et al., 2003; Pingoud et al., 2016). It was originally found in the *Bacillus stearothermophilus* 25B strain [New England Biolabs catalog

1

(NEB)<sup>1</sup>; The Restriction Enzyme Database (REBASE), new strain name Geobacillus sp. 25B] (Roberts et al., 2015). It is known that BcgI consisted of three subunits in the form of [RM]2 S in which the RM subunit is a fusion of endonuclease and N6mA methyltransferase (MTase) (Kong et al., 1993; Kong, 1998). It is also similar to a certain extent to Type IIG/IIB REases that form a single R-M-S fusion protein whose activity is sometimes stimulated by the presence of S-adenosylmethionine (SAM) (Jurenaite-Urbanaviciene et al., 2007). However, some BcgI-like REase activities are independent of ATP and SAM. The domain organization of the S subunit of BcgI-like REases (Kong, 1998) is similar to the HsdS (S) subunit of Type I restriction systems [HsdR2/HsdM2/HsdS or (R2/M2/S)] in the linear order of TRD1-CR1-TRD2-CR2 (TRD, target recognition domain, CR, conserved region) (Wilson and Murray, 1991; Murray, 2000; Loenen et al., 2014). In Type I HsdM2/HsdS complex structure, the CR1 and CR2 form two long α-helixes termed coiled-coil which interacts with the HsdM subunits to form a four-helix bundle and recognize two bipartite DNA sequences (Liu et al., 2017). A structure model of Type I S subunit in complex with the R-M subunits has also been published previously, revealing two TRD domains of the HsdS subunit interacting with [HsdR/HsdM]2 subunits (Kennaway et al., 2012). It has been shown that by site-directed mutagenesis of the S domain in the single chain MmeI-like endonucleases (R-M-S fusion), new specificities could be easily engineered by a few amino acid (aa) substitutions (Morgan and Luyten, 2009). Since MmeI methylase domain shared the same specificity domain, the MTase target is also altered simultaneously with the restriction specificity. Specificity domain TRDs swapping among three Type IIG/B enzymes with related recognition sequences (AloI, PpiI, TstI) also created new restriction specificities (Jurenaite-Urbanaviciene et al., 2007). Engineering new specificity from Type IIP REases proved to be more challenging (Lanio et al., 1998; Samuelson and Xu, 2002; Zhu et al., 2003; Samuelson et al., 2006), unless creating relaxed activity (i.e., increased star activity) or shortened recognition sequences (Heitman and Model, 1990; Guan et al., 2010). However, by site-directed mutagenesis, one could isolate enzyme variants with significantly reduced star activity ("highfidelity" HF mutants) of KpnI, BamHI, and EcoRI (SYX et al. unpublished results) (Vasu et al., 2013) (US patents number 8,673,610 on BamHI-HF, US patent number 9,249,396 on EcoRI-HF). In some sequenced bacterial genomes, some HsdR and HsdM can potentially partner with a few specificity subunits to create multi-specificity Type I restriction systems as the HsdS genes are in the same operon as the R-M genes (REBASE).

Most of the successful specificity (HsdS) engineering studies have been carried out with Type I restriction systems (Fuller-Pace et al., 1984; Nagaraja et al., 1985). In Type I REase EcoR124I, the N-terminal domain of the HsdS subunit (C-terminus deletion) can still interact with the HsdR2/HsdM2 subunits and generated a new specificity with symmetric recognition [WT EcoR124I (GAA N6 RTCG); EcoR124I HsdS<sub>NT</sub> N-terminal domain (GAA N7 TTC)] (Abadijieva et al., 1993; Youell and Firman, 2008). The length of a repeated sequences (two *vs.* three copies) in the long *a*-helical region (CR1) of the S subunit in EcoR124I and Eco124/3 (EcoR124II) was shown to dictate the length of the nonspecific DNA sequences between the two bipartite sites (GAA N6 RTCG vs. GAA N7 RTCG) (Gubler and Bickle, 1991). Since the domain architecture of the S subunit of Type IIB restriction systems is analogous to that of Type I system, we hypothesize that truncation of TRDs in the S subunit may also create new specificity. Our goal is to define the functional domain of BsaXI TRDs and the boundary of TRD and CR, and the minimal length of CR sequences. In this work, we present the results of cloning and expression of BsaXI RM and S genes in E. coli. The BsaXI enzyme complex can be purified by mixing cell extracts containing RM and S subunits or the RM and S subunits can be purified independently and two subunits can be mixed together to reconstitute BsaXI restriction activity. The TRD1-CR1 was expressed and purified which was reconstituted with the purified RM subunits to create a new specificity (5' AC N5 GT 3'). A part of the long  $\alpha$ -helix in CR1 (42 aa) could be deleted: Deletion of 8-aa, 15-aa, 21-aa, and 32-aa residues in CR1 created partial activity when the deletion variants were complexed with RM subunits. The TRD1 duplication in TRD1-CR1-TRD1-CR2 also showed partial activity. We also demonstrated that CR1 can be replaced by CR2 in TRD1-CR2 and generated a partial activity. The circular permutation of BsaXI TRDs in TRD2-CR2-TRD1-CR1 (i.e., relocation of TRD2 to the N-terminus and TRD1 to the C-terminus) produced the same specificity as the WT enzyme, although cleavage at the 3' side of CTCC and GT is somewhat imprecise and wobbly (5' ↓N9 AC N5 CTCC N10- $13\downarrow 3'$ ) (complement strand 5'  $\downarrow$ N7 GGAG N5 GT N12-13 $\downarrow$ 3'). Soluble TRD2-CR2 protein could not be expressed in E. coli. However, CR1-TRD2-CR2 variant protein could be expressed and purified, which showed low DNA nicking activity when it is reconstituted with the RM subunits. In addition, we showed that the BsaXI TRD2 can be substituted by a close homolog (59% aa sequence identity) and the chimeric S subunit created the same specificity as the WT enzyme. However, substitution with a distant TRD2 homolog failed to produce an active S variant. The protein homolog search using BlastP in GenBank found very few standalone BsaXI TRD1 homologs in microbial genomes, but there are at least 36 orphan (standalone) BsaXI TRD2 homologs, implying TRD2-CR2 in bacteria are probably available to pair up with other functional TRDs to create new restriction systems. This work provided more understanding of the S subunit of Type IIB restriction system and demonstrated a direct strategy to engineer new specificity from Type IIB systems.

### MATERIALS AND METHODS

The restriction enzymes,  $2 \times$  Phusion DNA polymerase PCR master mix, T4 DNA ligase, pBR322, pUC19, pTYB1, chitin beads, and Gibson DNA assembly/cloning kit were supplied by NEB. A bacterial genomic DNA (gDNA) was purified from *B. stearothermophilus* 25B strain (NEB strain collection, new strain name *Geobacillus* sp. 25B) by a phenol–chloroform extraction method as described before (Lunnen et al., 1988). The Big-Dye Sanger sequencing kit was purchased from Thermo–Fisher/ABI and plasmid DNA was sequenced with the protocol recommended by the manufacturer. The amplified PCR

<sup>&</sup>lt;sup>1</sup>rebase.neb.com/rebase/sublist.B.html

fragments of *bsaXIRM* and *bsaXIS* genes were cloned into intein-CBD expression vector pTYB1 (*NdeI-XhoI* cut, NEB) Xu and Evans (2003). and the inserts were sequenced to verify the correct coding sequences. The plasmids were transferred into T7 expression host C2566 (NEB) by transformation. DNAStar (Lasergene 14) and Geneious software were used for DNA sequence editing and cut site determination.

### **BsaXI Enzyme Complex Purification**

C2566 [pTYB1-bsaXIRM] and C2566 [pTYB1-bsaXIS] cells were cultured in 1 L flasks of LB + ampicillin and grown at 30°C until OD600 reached 1.3. After sitting at room temperature for 30 min, cultures were induced with 0.4 mM IPTG and grown overnight at 18°C. The cells were harvested by centrifugation and stored at -80°C. The frozen cell pellets from 3 L of BsaXI RM culture and 6 L of BsaXI S culture were all thawed (preliminary experiment indicated that the expression level of RM subunit is higher than the S subunit), resuspended in chitin column buffer (500 mM NaCl, 20 mM Tri-HCl, pH 8.5, 1 mM EDTA, and 0.1% Triton X-100 or Tween 20), lysed through 2 passes at 30 kpsi in a Dyhydromatics HL60 cell disruptor, and centrifuged to remove cell debris. The crude supernatants were loaded through two separate 40-ml chitin resin (NEB S6651) columns by gravity, washed with 300 ml of chitin column buffer, and then cleaved over 3 days with 50 mM DTT chitin cleavage/elution buffer. All chitin eluates were pooled together in 450 ml, diluted to 300-mM NaCl with addition of 300-mL zero salt column buffer and flowed through 20-ml diethylaminoethanol (DEAE) resin to remove nucleic acids. The DEAE flow-through and wash were again diluted to 150-mM NaCl with zero salt column buffer, applied to a Heparin TSK column (11 ml) and eluted with increasing NaCl gradient. The pooled active fractions were dialyzed overnight against 100-mM NaCl, 20-mM Tris-HCl, pH 8, 1 mM DTT, 0.1 mM EDTA, and 5% glycerol, passed through a Source 15S column (21 ml), then applied to 22 ml Source 15Q and eluted with increasing NaCl gradient (0.1 to 1 M). The pooled active fractions were dialyzed overnight against 50% glycerol (NEB Diluent B) and applied to a Superdex75 SEC column (1,787 ml). The pooled active fractions were concentrated by a second run over Heparin TSK. The final yield was  $\sim 10.0$  mg of pure BsaXI REase ([RM]2 + S). One BsaXI unit (U) is defined as the amount of protein required for complete digestion of 1-µg phage  $\lambda$  DNA in CutSmart buffer at 37°C for 1 h.

# Separate Purification of *BsaXI* RM and S Subunits

A 2 L of IPTG-induced cells containing RM-intein-CBD or 6 L of IPTG-induced cells of S-intein-CBD were resuspended in 60 ml and 180 ml of chitin column buffer and the suspensions sonicated to lyse the cells. After the removal of the cell debris by the centrifugation at 15,000 rpm at 4°C for 30 min, clarified cell lysates were loaded onto 20-ml chitin columns by gravity flow (1 column for RM-intein-CBD, 2 columns for S-intein-CBD). The CBD-tagged enzymes in flow-through were reloaded 3 times and the columns were washed with 10 column volumes of chitin buffer ( $\sim$ 200 ml). A 20 ml of cleavage buffer (chitin column

buffer + 50 mM DTT) was added to the top of chitin column. Approximately 2 ml of cleavage buffer passed through the column and the flow-through was discarded. The DTT-catalyzed intein cleavage reaction continued at 4°C for 2-3 days. The eluted proteins from chitin columns were diluted in a low salt buffer to reach 0.1-M NaCl concentration, which was subsequently loaded onto 5-ml Hi-Trap Heparin column (GE Healthcare). After the extensive washing with low salt, the RM or S protein was eluted with a salt gradient (0.1–1-M NaCl). Eluted peak fractions were analyzed by SDS-PAGE, concentrated in protein concentrators and protein was resuspended in enzyme storage buffer (0.2 M NaCl, 20 mM Tris-HCl, pH 7.5, 10 mM DTT, 0.5 mM EDTA, 50% sterile glycerol) to be stored at  $-20^{\circ}$ C.

### The BsaXI Specificity Subunit Domain Organization and Amino Acid Sequence of S Subunit Variants

The synthetic genes (gene blocks) encoding S variants were purchased from IDT, which carry 24-27 bp overlapping vector sequences at the 5' and 3' ends. They were assembled into pTYB1 using Gibson assembly kit and the assembled DNA was transferred into T7 expression strain C2566 (T7 Express, NEB) by transformation. To construct 6xHis-tagged S expression clone, A PCR fragment containing *bsaXIS* gene was cloned into pET21b (*NdeI* and *XhoI* cut, Novagen) to achieve C-terminal 6xHis tag of the targe protein. The pET21-*bsaXIS* plasmid was transferred into T7 Express (C2566) for expression. The *BsaXI* S (6xHis) tagged protein was purified from a nickel agarose beads column (10 ml, NEB) by gravity flow using a protocol supplied by NEB.

WT: TRD1-CR1-TRD2-CR2 (54.98 kDa)

TRD1 and TRD2: The aa residues are shown in blue and green, respectively. The CR1 and CR2, underlined aa residues. The boundary of CR1 (42-aa) and CR2 (42-aa) are predicted by structured-guided Phyre2 aa sequence alignment (Kelley et al., 2015). The PIPP residues shown in red indicates the boundary between TRD2 and CR2 (see **Supplementary Figure 1** for the sequence alignment). The boundary between TRD1 and CR1 is less clear-cut as CR1 may include 1–2 additional aa residues depending on the secondary structure prediction software; similarly, the boundary between CR1 and TRD2 has a 5-aa sequence (HFNIN) that is not part of either domain (see **Supplementary Figure 1**).

TFASEPSVRF	DFNYMKSVTP
LFEVVPSTVP	TLDESEPFKY
EVFPVTLSFE	DRDELNEDLF
PERGNILISA	IRPYLNKIVL
FTKAFIQIKP	LINSRILYYA
AVSRQGKGYP	TLKEDDLKTI
LAKEEELISN	IDALEKDIKE
<u>IVDEVFSS</u> HF	NINMVELMAL
SSISSLNSTI	RYSYRWNKMK
DCIEPLGKYI	LSSNNGWSPE
LGQEHLEFDG	VLNVSPTKAT
FIQEGDLFIS	RGNTVDLVGL
DIIYPDLYIR	LKIDEKVIHK
	TFASEPSVRF LFEVVPSTVP EVFPVTLSFE PERGNILISA FTKAFIQIKP AVSRQGKGYP LAKEEELISN IVDEVFSSHF SSISSLNSTI DCIEPLGKYI LGQEHLEFDG FIQEGDLFIS DIIYPDLYIR

SANKLIAVYI DIPPMERQIE IVSKIERQIG AQUELKQIE EKROIEVII EVEKES ADDIEKQIE EKROIEVII EVEKES ADDIEKQIE EKROIEVII EVEKES TRDI-CRI (TRD2-CR2 deletion, 27.06 kDa) TRDI-CRI (TRD2-CR2 deletion, 27.06 kDa) TRDI-CRI (TRD2-CR2 deletion, 27.06 kDa) TRDI-CRI TRAEFSKE INDIEKS TRDI-CRI TRAEFSKE DENVMKSUTP ALLORANGE TRAEFSKE DENVMKSUTP ALLORANGE VEFYTISEE DENDLANDLF THEYYTYKS LEFVYSTUF ILDESEFFKY KLIEKOFIL FERKMILSA INFYLANIUL KRIEKOFIL FERKMILSA INFYLANIUL TEFYTYKS LFFVYSTVP TIDESEFFKY COULD ADE EXPRESSION DENNKSTF INFYLANIUL FERKMILSA INFYLANIUL KRIEKOFIL FERKMILSA INF	KYLALLENS	F FGRLYFKYV	S KGKNOTMVKT	TKEDDKTDTY	FTKAFIOIKP	LINSRILYYA
AQMETERQIE EXENDIEVII PETARS OFSKVIDNI LAREEELISN IDALEKD TROI-CRI (TRO2-CR2 deletion, 77.06 kDs) In the predicted long a-helix at the C-terminus, 22.79 kDa) at the C-terminus, 22.79 kDa) ALECONDER ADDRESS TRAGEPSYRF DENYMKSYFF ALECONDER ALECONDER ADDRESS TRAGEPSYRF DENYMKSYFF ALECONDER ADDRESS TRAGEPSYRF DENYMKSYFF ALECONDER ADDRESS TRAGEPSYRF DENYMKSYFF ALECONDER ADDRESS TRAGEPSYRF DENYMKSYFF ALECONDER ADDRESS TRAGEPSYRF DENYMKSYFF ADDRESS TRAGEPSYRF DENYMKSYFF TEDJ-CRL-ISBAG (15-as deletion in the PERVILSS IF ADSRESSYFF DENYMKSYFF TRDJ-CRL-ISBAG (15-as deletioN IN TRAGEPSYFF ADDRESSYFFF DENELSDIFF ADDRESSYFFF DENELSDIFF	SSNELLNYY	L PIPPMEEOI	E IVGKIEEOIG	LRTIFSEKIN	AVSROGKGYP	TLKEDDLKTI
TEDI-CRI (TED2-CR2 deletion, 27.06 kDa) MGLIQRNES TFASEPSVRE DENYMKSVTP TEDSTTYKS LPEVVPSTVT TLD2SEPERY MGLIQRNES TFASEPSVRE DENYMKSVTP ABIGHVSKNG EVPFVTLSFE DRDLNRDLF TEDSTTYKS LPEVVPSTVT TLD2SEPERY MGLIQRNES TFASEPSVRE DENYMKSVTP ABIGHVSKNG EVPFVTLSFE DRDLNRDLF TREDTKTYK SLPEVVPSTVT TLD2SEPERY MGLIQRNES TFASEPSVRE DENYMKSVTP TREDTKTYK SLPEVVPSTVT TLD2SEPERY MGLIQRNES TFASEPSVRE DENYMKSVTP TRD1-CR1-57saA (15-aa deletion in TRD1 GRI with CR2, 27.33 kDa) MGLIQRNES TFASEPSVRE DENYMKSVTP AEGGVVSKNG EVPFVTLSFE DRDLNDLF MGLIQRNES TFASEPSVRE DENYMKSVTP AEGGVVSKNG EVPFVTLSFE DRDLNDLF MGLIQRNES TFASEPSVRE DENYMKSVTP MGLIQRNES TFASEPSVRE DENYMKSVTP TREDTTKIS LPEVVFYVT LDESEPFYX ATTERVTYKS LPEVVFYVT TLDESEPFYX ATTERVTYKS LPEVVFYTY TLDESEPFYX ANGROSKE VFYTLSFE DROLLADLF MVFINAL MSGNVFYTTRE TASEPSVRE DENYMKSVTP TTERVTYKS LPEVVFYTY TLDESEPFYX ANGROSKE VFYTLSFE DROLLADLF MVFINAL MSGNVFYTTRE TASEPSVRE DENYMKSVTP TTERVTYKS LPEVVFYTY TLDESEPFYX ANGROSKE VFYTLS	AONEIEKOI	E EKRNOIRVI	I EETARS	OFSKKVIDNL	LAKEEELISN	IDALEKD
TRDI-CRI (TRD2-CRI Geletion, 7:06 kba) MGLIGRNPS TPASEPSYRF DENYUKSYTP at the C-terminus, 22.78 kba) TIEDYTYKS LPEWYDSTVF TLDESEPFKY MGLIGRNPS TEASEPSYRF DENYUKSYTP ALIGHYSNG EVFYTLSED BRUELNBULF TLDESEPFKY MGLIGRNPS TEASEPSYRF DENYLNSYTP ARISKOGTEL PERKNILSA TRYLNRIVL ARIGHYSNG EVFYTTSE DEDELIADIT KRITKGDIFL PERKNILSA TRYLNRIVL ARIGHYSNG EVFYTTSE DEDELIADIT LKEDKTDY FTKAFIGIKE LINSKILYA ARIGHYSNG EVFYTLSED BRUELNBUFF TRD1-CRI LARSEELSN IDALEKDEKTI IKEDKKTI IKEDKKTDY FTKAFIGIKE LINSKILYA ARIGHYSNG EVFYTLSED BRUELNBUFF TRD1-CRI (TRD2 deletion, replacing CRI with CR2, 77.53 Kba) MGLIGRNPS TPASEPSYRF DENYUKSVTP ARIGHYSNG EVFYTLSED BRUELNBUFF TLDESEFFKY COUL of the entire predicted MGLIGRNPS TPASEPSYRF DENYLWKSVTP ND1 GFRAFUDA MGLIGRNPS TPASEPSYRF DENELFF MGLIGRNPS TFASEPSYRF DENELFF MGLIGR	<u></u> <u>k</u> <u>k</u>			<u></u>		
TRDI-CRI (TRD2-CR2 deletion, 27.06 kDa) in the predicted long a-helix MCLIQRANES TFASESPAYE DEWYMENTP THEOREM STP. ACCOUNT AND A CONTRACT AND				TRD1-CR1-32	aa $\Delta$ (32-aa d	eletion
MGLIGENNES TRASEPSNE DENYMMSVTP at the C-terminus, 22.78 Noa) THENYTYNS LEVVPENT TIDESERFY MGLIGENES FASEPSNE DENYMSVTP ALIGUYSKNG EVEPYTISE OBDELNEDLF THENYTYNS LEVVPENT LINSELLYA ALIGUYSKNG EVEPYTISE OBDELNEDLF THENTYNS LEVVPENT INFOLMANTIN FRAFIGIRF LINSELLYA KXIEKGDIFL PERGNILISA TRPNINKIVL ALIGUYSKNG EVEPYTISE OFSKVIDAL LAREELISN IDALEDNIE INFOLMANTIN COSKNIDAL LAREELISN IDALEDNIE TROI-CR2 (TRO2 deletion, replacing TROI-CR2 (TRO2 deletion in TROI ALIGUYSKNG EVEPYTISE DEDELNEDLF MGLIGENNES TRASEPSNEP DENYMSVTP THEVYTYKS LEVVPENTP TLDESEPFKY COULD NE SETARS TROI-CR1-8aaA (8-aa deletion in the TROI-CR1-8aaA (8-aa deletion in the TROI-CR1-7ROI (TRAFTGIKP LINSRITYA ALIGHYSKNG EVEPVTISEE DENNESSERFY THESYTYKS LEVVPENTP TIDESERFKY THESYTYKS LEVVPENTP TIDESERFKY THESYTYKS LEVVPENTP TIDESERFKY THESYTYKS LEVVPENTP TIDESERFKY THESYTYKS LEVVPENTP TIDESERFKY ALIGHYSKNG VEFYTISE TROI-CR1-8aaA (15-aa deletion in the TROI-CR1-7ROI-CR2 (TROI DUPINES TRAFTGIKKNINI, LAREELISN IDALENDIK KKIEKGDIFL PERGNILISA THEYLKNINI KKIEKGDIFL PERGNILISA THEYLKNINI	TRD1-CR1 (T	RD2-CR2 delet	tion, 27.06 kDa)	in the pred	icted long $\alpha$ -	-helix
TTEEYTYTKS LFEVUESTVP TLDESEPFKY MGLIQREFS TFASSESVER DFWYMKSVTP ARIGHVSKNG EVPFVTLSFE DRDELMEDIF KKTERGOTEL PERGNILISA TEPVINKTVL ARIGHVSKNG EVPFVTLSFE DREINEDIF KKTERGOTEL PERGNILISA TEPVINKTVT ARIGHVSKNG EVPFVTLSFE DREINEDIF KKTERGOTEL PERGNILISA TEPVINKTVT TERYTYKS LFEVYBSTVP TLDESEFFXY ADGLIQRENFS TFASESVKF DENYMKSVTP TERYTYKS LFEVVBSTVP TLDESEFFXY ADGLIQRENFS TFASESVKF DENYMKSVTP KKTERGOTEL PERGNILISA TEPVINKTVL KKTERGOTEL PERGNILISA TEPVINKTVT TERYTYKS LFEVVPSTVP TLDESEFFXY ALGGNVSKNG EVPFVTLSFE DREINEDIF KKTERGOTEL PERGNILISA TEPVINKSVTP TERYTYKS LFEVVPSTVP TLDESEFFXY TERYTYKS TRAFFSVFF DFNYKSVTP TRENTYKS TRAFFSVFF DFNYKSVTP TRENTYKS TRAFFSVFF TERTSFYFY TERYTYKS TERFSVFF TERTSFYFY TERYTYKS TERFSVFF TERTSFYFY TERYTYKS TERFSVFFY TTER TERTSFYFY TERYTYKS TERFSVF	MGLIQRRNFS	TFASEPSVRF	DFNYMKSVTP	at the C-te	rminus, 22.78	3 kDa)
ARIGNSKNG EVFENTLSFE DBDELMEDLF TTERYTYKS LFEVNESTVF TLDESSEFRY KKIEKGOTL PERGNILASA IAPVINKUL ARIGNSKNG EVFENTLSFE DBDELMEDLF IKEDGKTDIY FIKAFIQIKP LINSRLIYA KKIEKGOTL FERGNILASA IAPVINKUL LAKTESEKIN AVSRQGKGYP TLKEDDLKTI IKEDGKTDIY FIKAFIQIKP LINSRLIYA (SFKKVIDNL LAKEELISN IDALEKDIKE IRTIFSEKIN AVSRQGKGYP TLKEDDLKTI IKEDGKDIY TRDI-CR2 (TRD2 deletion, replacing TRDI-CR2 (TRD2 deletion) TEEYIYKS LFEVVPSTVP TLDESEFFKY KOLIGARNES TRASEPSVRF DENUMKSVTP MGLIGARNES TRASEPSVRF DENUMKSVTP TRDI-CR1-8aA (8-aa deletion in the predicted long a-helix at the C-termins, 25.55 kDa MGLIGARNES TRASEPSVRF DENUMKSVTP TGEVYTKS LFEVVPSTVP TLDESEFFKY KIEKGDIFL FERGNILISA IRPYLNKUT KIEKGDIFL FERGNILISA IRPYLNKUT KIEKGDIFL FERGNILISA IRPYLNKUT TRDI-CR1-8aA (8-aa deletion in the predicted long a-helix at the C-termins, 25.55 kDa MGLIGARNES TRASEPSVRF DENUMKSVTP MGLIGARNES TRASEPSVRF DENUMKSVTP MGLIGARNES TRASEPSVRF DENUMKSVTP KIEKGDIFL PERGNILISA IRPYLNKUT KKIEKGDIFL PERGNILISA INTAK KKIEKGDIFL PERGNILISA INTAK KKIEKGDIFL ACKEDUKTI KKIEKGDIFL PERGNILISA INTAK KKIEKGDIFL PERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK KKIEKGDIFL PERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK KKIEKGDIFL FERGNILISA INTAK K	TTEEYYTYKS	LFEVVPSTVP	TLDESEPFKY	MGLIQRRNFS	TFASEPSVRF	DFNYMKSVTP
KKIEKGDIFL     PERCHILISA     IPPLIMENUL       AKEDDKIDIY     PERCHILISA     IPPLIMENUL       IRTEDKIDIY     PERCHILISA     IPPLIMENUL       IRTEDKIDIY     PERCHILISA     IPPLIMENUL       IRTEDKIDIY     PERCHILISA     IPPLIMENUL       IRTESEKIN     ANSOCREY     TINSHIYYA       OFSKVIDNI     LAKEBELISN     IDALEKDIKE     IPPLIMENUL       REGIORARE     IUDEVFSSHF     NIN     OFSKVIDNI       REGIORARE     IUDEVFSSHF     NIN     OFSKVIDNI       ALGURAKKE     IUDEVFSSHF     NIN     OFSKVIDNI     IAKEEE       TRDI-CR2     (TRO-CR2)     CR02     deletion     in TRDI       CR1     DUGURANS     TRASEZSKF     DENNIKSVTP     cold not be expressed in E. coli       TREDKTDIY     FTASEPSKF     DENNIKSVTP     cold not be expressed in E. coli     in RevinksvTP       KKIEKGDIFL     FRAMINISA     TEPVINKIV     TEFEVYTYKS     IFFWPTLSF     DENNIKSVTP       KKIEKGDIFL     FRAMINISA     TEPVINKSVTP     TINSHIYYA     AEIGHVSKNG     EVEVTINS     TERVINKVL       KKIEKGDIFL     FRAMINISA     TEPVINKVL     TEFEVYTYKS     IFFWPTLSFF     DENNIKSVTP       TRDI-CR1-RDA     (As a deletion in the     TPDI-CR1-TRDI-CR2     TRDI-KNIVA       PREGIONARS     FF	AEIGHVSKNG	EVFPVTLSFE	DRDELNEDLF	TTEEYYTYKS	LFEVVPSTVP	TLDESEPFKY
IKEDEKTDIY FIKAPIGIKP LINSKILYYA KKEKEDIFL PERCNILISA INPULNKIUL IKRIPSEKI AVSROCKOYP TLKEDDLKTI KKEKEDIY FIKAPIGIKP LINSKILYYA OPSKKVIDNI LAKEEELISN IDALEKDIKE INTIPERIN AVSROCKOYP TLKEDDLKTI OPSKKVIDNI LAKEEE KRIQRSKKE IVDEYFSSHF NIN OPSKKVIDNI LAKEEE TRDI-CR2 (TRD2 deletion, replacing TRDI-CR1-57aaA (15-aa deletion in TRDI CR1 with CR2, 27.53 kba) plus 42-aa deletion of the entire predicted MGLIGRNFS TFASEPSVRF DFNYMKSVTP long a-heiix at the C-terminus, protein could not be expressed in <i>E. coli</i> ) AFIGWSKNG EVPYVT.SFE DRDELMENLF MGLIGRNFS TFASEPSVF DENYMKSVTP KKIEKGDIFL PERCNILISA IRPYLNKIVL TEEYYYKS LFEVVPSTVP TLDESEPFKY ALKDEKTDIY FIKAPIGIKP LINSKILYYA AEIGUVSKNG EVFYVTISZE DENUKSVTP KKIEKGDIFL PERCNILISA IRPYLNKIVL TEEYTYKS LFEVVPSTVP TLDESEPFKY ALKDEKTDIY FIKAPIGIKP LINSKILYYA AEIGUVSKNG EVFYVTISZE DENUKSVTP VRDI-CR1-BasA (8-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kba) MGLIGRNFS TFASEPSVRF DENYMKSVTP MGLIGRNFS TFASEPSVRF DENYMKSVTP KKIEKGDIFL PERCNILISA IRPYLNKIVL KKIEKGDIFL PERCNILISA IRPYLNKIVT KKIEKGDIFL PERCNILISA IRPYLNKIVL KKIEKGDIFL PERCNILISA IRPYLNKIVL KKIEKGDIFL PERCNILISA IRPYLNKIVL KKIEKGDIFL PERCNILISA IRPYLNKIVL KKIEKGDIFL PERCNILISA IRPYLNKIVI KKIEKGDIFL PERCNILISA IRPYLNKIVI KKIEKGDIFL PERCNILISA IRPYLNKIVI KKIEKGDIFL	KKIEKGDIFL	PERGNILISA	IRPYLNKIVL	AEIGHVSKNG	EVFPVTLSFE	DRDELNEDLF
LRTEPSEKIN AVSROCKGYP TLKEDDLKTI IKEDDKTDIY FTKAFIGIKP LINSKLLYYA QFSKKVIDNL LAREEELISN IDALKDIKE UNDEVFSSHF NIN QFSKKVIDNL LAREEEL TRDI-CR2 (TRD2 deletion, replacing TRDI-CR2 (TRDE VERVENT) AEGRYSKNG EVEPVILSE DRDELMEDLF MCLIORRNES TRASEPSVRF DENVMSKYTP TRDI-CR1-BasA (8-aa deletion in the TRDI-CR1-BasA (8-aa deletion in the TRDI-CR1-TRDI-CR2 (TRDI duplication, a.k.a., 2XTRDI with CR1 and CR2, 5.55 Kba) MCLIORRNES TRASEPSVRF DENVMSKYTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHYSKNG EVEPVILSFE DRDELMEDLF AEIGHYSKNG EVEPVILSFE DRDELMEDLF AEIGHYSKNG EVEPVILSFE DRDELMEDLF AEIGHYSKNG EVEPVILSFE DRDELMET IKEDDKTDIY FTKAFIQIKP LINSKLYYA IKEDDKTDIY FTKAFIQIKP LINSKLYYA IKEDCR1F PERGNILISA IRPYLNKIVL KKIEKCDIFL PERGNILISA IRPYLNKIVL KKIEKCDIFL PERGNILISA IRPYLNKIVL KKIEKCDIFL PERGNILISA IRPYLNKIVL KKIEKCDIFL PERGNILISA IRPYLNKIVL IKEDCKTPF PERGNELISA ISPYLNKIVL IKEDCKTPF PERGNELISA ISPYLNKIVL IKEDCKTPF PERGNELISA ISPYLNKIVL IKEDCKTPF TASEPSVRF DYNMKSYTP TTESYTYKS TRDI-CR1-15aaA (15-aa deletion in the UFYVILSFE DRDELNEDLF TRASEPSVRF DYNMKSYTP TTESYTYKS LFEVPSSVF PINMKSYTP TRDI-CR1-15aAA (15-aa deletion in the IKEIQESKE IKEIQ	IKEDDKTDIY	FTKAFIQIKP	LINSRILYYA	KKIEKGDIFL	PERGNILISA	IRPYLNKIVL
OPSKKVIDNL         LAKEREELISN         IDALEKDIKE         LRTIFSEKIN         AVSRQGKOP         TLKEDDLKTI           LKSIQRSKE         IVDEVYSSHF         NIN         OPSKKVIDNL         LAKEEE           TRD1-CR2 (TRD2 deletion, replacing         TRD1-CR1-57aaA (15-aa deletion in TRD1 plus 42-aa deletion of the entire predicted           OKGLIGRNNFS         TFASEPSVR         DENYMKSVTF         long α-helix at the C-terminus, protein           ALEGAVSKNG         EVEVPTUSEE         DRDINKSVTF         could not be expressed in R. coli)           ALEGAVSKNG         EVEVPTUSEE         DRDINKSVTF         COULGARNES         TFASEPSVF           AKIEKGOFTL         PERGNILISA         REPLINKIVL         TERETYTKS         LEVEVPSTVP           ARIGHVSKNG         EVEVPTLSEE         DRDINKSVTF         MGLIQARNES         DRNMKSVTP           IKEDDKTDIY         FTKAFOIKA         AVSRQGKOP T         T         T           RTD1-CR1-SaaA (6-aa deletion in the         TRD1-CR1-TRD1-CR2 (TRD1 duplication, a.k.a., 2XTRD1 with CR1 and CR2, 54.11 kDa)         S4.11 kDa)           MSLIQRANFS         TFASEPSVRF         DENYMKSVTP         TGDSEPFRY         TEREPTYTKS           ARIGHVSKNG         EVEVPTUSED         DRDELMEDLF         ALGHVSKNG         EVEVPTUSED           VERKVIDNI         ARKENOIKVI         ERKNOKVIT	LRTIFSEKIN	AVSRQGKGYP	TLKEDDLKTI	IKEDDKTDIY	FTKAFIQIKP	LINSRILYYA
LKSIQRSKKEIVDEVPESSHFNINOPSKVIDNLLAKEEETRDI-CR2 (TRD2 deletion, replacingTRDI-CR1-57aaA (15-aa deletion in TRDICR1 with CR2, 27.53 kDa)plus 42-aa deletion of the entire predictedMGLIQRNPSTFASEPSVRFDENYMKSVTPDAIGNISMONEVEPVILSTPDENYMKSVTPALGUNSKONEVEPVILSTPREDENELIFKKIEKGDIFLPERGNILISAIRPYLNKIVLTTERYYTYKSLFEVVPSTVPTLDESEPFKYKKIEKGDIFLPERGNILISAIRPYLNKIVLTTERYYTYKSLFEVVPSTVPTLDESEPFKYKKIEKGDIFLEKKNORYPTLKEDDLKIIKKIEKGDIFLEKKNORYPTLKEDDLKIIKKIEKGDIFLEKKNORYPTLKEDDLKIIGYSKKVIDNLDETPYMEQUEIVCKIERCIGKKEDKKIDNEKKNORVENIEKKNORVENIPredicted long a-helix at the C-terminus, a.k.a., 2XTRDI with CR1 and CR2, 54.11 kDa)S4.11 kDa)MGLIQRNPSTFASEPSVRFDENYMKSVTPTTREYTYTKSLFEVVPSTVPTLDESEPFKYTTREYTYTKSLFEVVPSTVPTLDESEPFKYKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRPYLMKIVLKKIEKGDIFLPERGNILISAIRP	QFSKKV <u>IDNL</u>	LAKEEELISN	<u>IDALEKDIKE</u>	LRTIFSEKIN	AVSRQGKGYP	TLKEDDLKTI
TRD1-CR2 (TRD2 deletion, replacingTRD1-CR1-57aaA (15-aa deletion in TRD1CR1 with CR2, 27.53 kDa)plus 42-aa deletion of the entire predictedMGLIQRRNFS TFASEPSVRF DFNYMKSVTPcould not be expressed in E. coli)ALIGHVSKNG EVFPVILSFE DRDELNEDLFMGLIQRRNFS TFASEPSVRF DFNYMKSVTPKITEKOTIYF FIKAFIGIKP LINSKLIYAAEIGHVSKNG EVFPVILSFE DRDELNEDLFRKTEKODIFL PERGNILISA IRPYLNKIVLTTEFYYTYKS LFEVVPSTVP TLDESEPFKYAQNEIEKQIE EKKNQIRVII EETARSIKEDDKTDIY PTKAFIGIKP LINSKLIYAAQNEIEKQIE EKKNQIRVII EETARSIKEDKTDIY PTKAFIGIKP LINSKLIYAAQNEIEKQIE LONG a-heix at the C-terminus, 25.56 kDa)TRD1-CR1-TRD1-CR2 (TRD1 duplication, a.k.a., 2XTRD1 with CR1 and CR2, 55.6 kDa)MGLIQRRNFS TFASEPSVFF DENYMKSVTPMGLIQRRNFS TFASEPSVFF DENYMKSVTPMGLIQRRNFS TRASEPSVFF DENYMKSVTPTTEFYTYKK LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDELNEDLF AEIGHVSKNG EVFPVTLSFE DRDELNEDLFKKIEKCDIFL PERCNILISA IRPYLNKIVLKKIEKCDIFL PERCNILISA IRPYLNKIVL KKIEKCDIFL PERCNILISA IRPYLNKIVL KKIEKCDIFL PERCNILISA IRPYLNKIVL KKIEKCDIFL PERCNILISA IRPYLNKIVL KKIEKCDIFL PERCNILISA INPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRLYYA AEIGHVSKNG EVFPVTLSFE DRDELNEDLF AEIGHVSKNG EVFPVTLSFE DRDELNEDLF AKIEKCDIFL PERCNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRLYYA AEIGHVSKNG EVFPVTLSFE DRDELNEDLFTRD1-CR1-15aaA (15-aa deletion in the Predicted long α-helix at the C-terminus, 24.69 kDa) MGLIQRRNFS TFASEPSVRF DFNYMKSVTP TTEEYTYKS LFEVPSTVP TLDESEPFKY ASIGHVSKNG EVFPVTLSFE DRDENDELF FTKAFIDIKP LINSRLYYA IKEDDKTDIY FTKAFICIKF IKEDDKTDIY FTKAFICIKF IKEDDKTDIY FTKAFICIKF IKEDDKTDIY FTKAFICIKF IKEDDKTDIY FTKAFICIKF IKEDKTDIY PICKELSPNFY DITYNKSVTF TTEYTYKS LFEVP	<u>LKSIQRSKKE</u>	<u>IVDEVFSS</u> HF	NIN	QFSKKV <u>IDNL</u>	LAKEEE	
IRDI-CR2IRDI-CR2IRDI-CR1-STARAIRDI-CR1-STARAIRDI-CR1-STARAMGLIQRRMFSIFASEPSVRFDENYMKSVTPIng a-helix at the C-terminus, proteinTTEEYYTYKSIFEVYTYKSIFEVYTYVFYIDESEPFKYALGHVSKNGEVFYVILSPEDROLINEDLFMGLIQRNPSTRASEPSVRFIKEDDKTDIYFTKAFDIKTIIKEDKTDIYFTKAFDIKTIKINSTILYYAALGHVSKNGEVFYVILSPEDROELMEDLFKKIEKGDIFLPERGNILISAIRDI-CR1-StarA(S-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kba)IRTI-SEKINAVSROGKGYPMGLIQRNFSTFASEPSVRFDFNYMKSVTPTTEEYYTYKSIFEVVPSTVPTTEEYYTYKSIFEVVPSTVPTLDESEPFKY ALGHVSKNGARk.a., 2xTRD1 with CR1 and CR2, 54.11 kba)MGLIQRNFSTFASEPSVRFDFNYMKSVTPTTEEYYTYKSIFEVVPSTVPTLDESEPFKY AEIGHVSKNGEVFVVISFEDROELMEDLF AEIGHVSKNGKKIEKGDIFLPERONILISAREYLNKIVL KKIEKGDIFLREGNISSANGEVFVVISFEDROELDEDLF AEIGHVSKNGIKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYIKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYFTKAFIQIKPINSRILYYAIKEDDKTDIYFTKAFIQIKPINSRILYYAIKEDDKTDIYIKEDDKTDIYFTKAFIQIKPINSRILYYAIKEDDKTDIYFTKAFIQIKPINSRILYYAIKEDDKTDIYINSRILYYAIKEDDKTDIYIKEDDKTDIYFTKAFIQIKPINSRILYYAIKEDDKTDIYINSRILYYAIKEDDKTDIYFTKAFIQIKP					(15)	-l-t Tool
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	CD1 with CD	RDZ GELECION,	replacing	IRDI-CRI-57	$da\Delta$ (15-aa u	the entire predicted
	MCITOPPNES	Z, ZI.JJ KDA)	DENVMEQUED	long «-boli	v at the C-te	rminus protoin
THENTITIGCould not be transessed in D. ScoreALEGIVSKNGEVPFVTLSFEDRDELMEDLFKKLEKKODTPLPERGNILISAIRPYLMKIVLTREDOKTUYFTKAFIQIKPLINSRILYVAARIGOVSKNGEVPFVTLSFEDRDELMEDLFKKLEKKODTPLPERGNILISAIRPYLMKIVLQFSKKVIDNLLETPPMERQIEIVGKIERQIGAQMEIEKQIEEKRNQIRVIIEETARSTRD1-CR1-8aaA(8-aa deletion in theTRD1-CR1-TRD1-CR2Predicted long a-helix at the C-terminus,a.k., 2/XRD1 with CR1 and CR2,25.56 kDa)S4.11 kDa)S4.11 kDa)MGLIQRRNFSTFASEPSVRFDFNYMKSVTPTEREYTYKSLFEVPSVPVTDESEPFKYTEEYTYKSLFEVPSVPVTDESEPFKYAEIGHVSKNGEVPFVTLSFEDRDLMEDLFAEIGHVSKNGEVPFVTLSFEDRDLMEDLFAEIGHVSKNGEVPFVTLSFEDRDLMEDLFAEIGHVSKNGEVPFVTLSFEDRDLMEDLFAEIGHVSKNGEVPFVTLSFEDRDLMEDLFAEIGHVSKNGEVPFVTLSFEDRDLMEDLFAEIGHVSKNGEVPFVTLSFEDRDLMEDLFKKEKGDIFLPERGNILISAIRPYLMKIVLIKEDDKTDIYFTKAFIQIKPINSRILYAAIKEDDKTDIYFTKAFIQIKPINSRILYAAIKEDDKTDIYFTKAFIQIKPINSRILYAAIKEDDKTDIYFTKAFIQIKPINSRILYAAIKEDDKTIDYFTKAFIQIKPINSRILYAAIKEDDKTIDYFTKAFIQIKPINSRILYAAIKEDDKTIDYFTKAFIQIKPINSRILYAAIKEDDKTIDYFTKAFIQIKPINSRILYAAIKEDDKTIDY <t< td=""><td>TTEEVVTVKQ</td><td>IFASEFSVAF</td><td>TIDESEDEKY</td><td>could not b</td><td>at the c-le</td><td><math>E_{in} = Coli</math></td></t<>	TTEEVVTVKQ	IFASEFSVAF	TIDESEDEKY	could not b	at the c-le	$E_{in} = Coli$
$ \begin{array}{l c c c c c c c c c c c c c c c c c c c$	AFICHVSKNC	FVFPVTLSFF	DEDELITINE	MGLIORRNES	TFASEPSVRF	DENYMKSVTP
NUMBER ININSTITUTEINSTITUTEINEDDATIONFINALIZAABIGHUSANGEVEPUTISEEDRDELNEDLELRTIFSEKINAVSRQCKGYPTLKEDDLKTIKIEKGDIFLPERSNILLSAIFPUINKIVLQFSKKUDNLLPIPPMEEQLEIVGKIEEQIGIKEDDKTDIYFIKAPIQIKPLINSRILYAAQMEIEKQIEEKRNQIRVIIEETARSLRTIFSEKINAVSRQCKGYPTTRDI-CRI-8aaA(8-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kDa)TRDI-CRI-TRDI-CR2 (TRDI duplication, a.k.a., 2XTRDI with CRI and CR2, 54.11 kDa)MGLIQRNPSTFASEPSVFPDENYMKSVTPMGLIQRNPSTFASEPSVFPTTEEYYTYKSLFEVVPSTVPTLDESEPFKYAEIGHUSKNGEVEPVTLSFEDRDELNEDLFAKIEKGDIFLPERGNILISAIRFYLNKIVLKKIEKGDIFLPERGNILISAIRFYLNKIVLIKEDDKTDIYFTKAFIQIKPLINSRILYAIKEDDKTDIYFTKAFIQIKPLINSRILYAIKEDDKTDIYFTKAFIQIKPLINSRILYAIKEDDKTDIYFTKAFIQIKPINSRILYAIKEDDKTDIYFTKAFIQIKPLINSRILYAIKEDDKTDIYTIESYTIKSIKEDDKTDIYFTKAFIQIKPLINSRILYAIKEDDKTDIYTIESYTIKSIKEJORSKKEIVDEVFSKYDROENNESTTAEFESKINNGLIQRNPSTTRDI-CRI-15aaA(15-aa deletion in the predicted long \$\mathbf{a}\$ hepvinktviEKSIQRSKEIVDEVFSHFMGLIQRNPSTTASEPSVRFDENYMKSVTPFTKAFIQIKPLINSRILYATTREYTYKSLFEVPSTVPTLDESEPFKYAVSRQKGYPTIENYTIKSRGILORNPSTFASEPSVFFDENYMKSVTP <t< td=""><td>KKIEKCDIEI.</td><td>PERGNILISA</td><td>TRPYLNKTVI</td><td>TTEEYYTYKS</td><td>LFEWVPSTVP</td><td>TLDESEPFKY</td></t<>	KKIEKCDIEI.	PERGNILISA	TRPYLNKTVI	TTEEYYTYKS	LFEWVPSTVP	TLDESEPFKY
InternationInternationINTIFERINGINTIFEDINGINTIFERINGINTIFEDINGINTIFERINGINTIFEDINGINTIFERINGINTIFEDINGINTIFERINGINTIFEINTIFERINGINTIFE <td>TKEDDKTDIY</td> <td>FTKAFIOIKP</td> <td>LINSRILYYA</td> <td>AEIGHVSKNG</td> <td>EVEPVTLSEE</td> <td>DRDELNEDLE</td>	TKEDDKTDIY	FTKAFIOIKP	LINSRILYYA	AEIGHVSKNG	EVEPVTLSEE	DRDELNEDLE
OFSKKUTONLLPTPPMECUEUVGKTEEQIG EKRNQIRVIIINSRILYA EETARSAQNEIEKQIEEKRNQIRVIIEETARSINTERSETTRD1-CR1-8aaA (8-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kDa)TRD1-CR1-RD1-CR2 (TRD1 duplication, a.k.a., 2xTRD1 with CR1 and CR2, 54.11 kDa)MGLIQRNFS TFASEPSVRFDFNYMKSVTP MGLIQRNFS TFASEPSVRFTLDESEPFKY AEIGHVSKNG EVFPVTLSFETRD1-CR1-RD1-CR2 (TRD1 duplication, a.k.a., 2xTRD1 with CR1 and CR2, 54.11 kDa)KEIEKOFIFPERVINISIFASEPSVRFDFNYMKSVTP MGLIQRNFS TFASEPSVRFDFNYMKSVTP TTDESEPFKYAEIGHVSKNGEVFPVTLSFEDRDELNEDLF AEIGHVSKNG EVFPVTLSFEDRDELNEDLF AEIGHVSKNG EVFPVTLSFEDRDELNEDLF KKIEKGDIFLKKIEKGDIFLPERVILISAIRPYLNKIVL IKEDDKTDIYFTKAFIQIKPLINSRILYYA LKIEKDLKTI UFSKVIDNLLAKEEELISNIAKSORKEIDALEKDIKE ISSURSKEOFSKVIDNLLAKEEELISNIDALEKDIKE IDALEKDIKE UFSVTVPTIDALEKDIKE VEFSVTVPTTRD1-CR1-15aa(15-aa deletion in the Predicted long α-helix at the C-terminus, 24.69 kDa)PERGNILISA IRPYLNKIVL PERGNILISA IRPYLNKIVLIKEDDKTDIY FTKAFIQIKPINSRILYYA AEIGHVSKNG EVFPVTLSFEDFNYMKSVTPTRD2-CR2 (ATRD1-CR1, POEN MGLIQRNFSIDALEKDIKE VERVSTUPPIPPMEEQLE IVGSKVIDNL AEIGHVSKNG EVFPVTLSFERCNQIRVISSISSLNSTI RYSRWNKK LIQKIYMSSISSLNSTI RYSRWNKK LIQKIYMSSISSLNSTI RYSRWNKK LIQKIYMSSISSLNSTI RYSRWNKK LIQKIYM DCIEPIGKYITRD1-CR1-21aa(21-aa deletion IDALEKDIKE IKSDDKDIN LKSIDALEKDIKE ISSNUSSPSSSISSLN	LRTIFSEKIN	AVSROGKGYP	TIKEDDIKTI	KKIEKGDIFL	PERGNILISA	TRPYLNKTVL
AQNETERQIEEXERCITEAQNETERQIEEXENDITIETARSTRD1-CR1-8aaA (8-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kDa)TRD1-CR1-TRD1-CR2 (TRD1 duplication, a.k.a., 2xTRD1 with CR1 and CR2, 54.11 kDa)MGLIQRRNFS TFASEPSVRF DPNYMKSVTPMGLIQRRNFS TFASEPSVRF DPNYMKSVTPTEEYYTYKS LFEVVPSTVP TLDESEPFKYAEIGHVSKNG EVFPVTLSFE DRDELNEDLFAEIGHVSKNG EVFPVTLSFE DRDELNEDLFRKIEKGDIFL PERGNILISA IRPYLNKIVLKKIEKGDIFL PERGNILISA IRPYLNKIVLKKIEKGDIFL PERGNILISA IRPYLNKIVLRKIEKGDIFL PERGNILISA IRPYLNKIVLLKEDDKTD1Y FTKAFIQIKP LINSRILYYAIKEDDKTD1Y FTKAFIQIKP LINSRILYYAILKEDDLKTILKSIQRSKKEOFSKKVIDNL LAKEEELISN IDALEKDIKE ILASIGRSKKEOFSKKVIDNL LAKEEELISN IDALEKDIKE ILSIGRSKKEDENUMKSVTP TTEEYYTYKSTRD1-CR1-15aaA (15-aa deletion in the predicted long α-helix at the c-terminus, 24.69 kDa)PERGNILISA IRPYLNKIVL IKEDDKTDIYMGLIQRRNFS TFASEPSVRF DRDLINEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVLPERGNILISA IRPYLNKIVL IKEDDKTDIYMGLIQRRNFS TFASEPSVRF DRDLINEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVLPERGNILISA IRPYLNKIVL IKEDDKTDIYMGLIQRRNFS TFASEPSVRF DRDLINEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVLTRD2-CR2 (ATRD1-CR1, protein could not be expressed in QFSKKVIDNL LAKEELISN IDALEKDIKE ISSNSMSPE SVVGEGIPI LGQEHLEFDGTRD1-CR1-21aaA (21-aa deletion in the predicted long α-helix at the C-terminus, 23.99 kDa)SGNTVDGL RCNTUNGL ACVVETEV TI DISEPFKY KCNOTWKI SSNELLNYL NGLIQRNFS TFASEPSVRF PONYMKSVTPTINDEKVIHK KYIALLFNSF FRIVEYS SNELNYL IKDEKVIHK KYIALLFNSF FRIVEYSTRD1-CR1-21aaA (21-aa deletion in the predicted long α-helix at the C-terminus, 23.	OFSKKVIDNI	I.P.T.P.P.MEEOLE	TVGKIEEOIG	TKEDDKTDTY	FTKAFIOIKP	LINSRILYYA
TRD1-CR1-8aaA (8-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kDa)TRD1-CR1-TRD1-CR2 (TRD1 duplication, a.k.a., 2xTRD1 with CR1 and CR2, 54.11 kDa)MGLIQRNFS TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDENEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYYA IRTIFSEKIN AVSRQGKGYP TLKEDDLKTI QFSKKVIDNL LAKEEELISN IDALEKDIKE UCFSKVIDNL LAKEEELISN IDALEKDIKE Predicted long $\alpha$ -helix at the C-terminus, 24.69 kDa)TRD1-CR1-15aaA (15-aa deletion in the PERGNILISA IRPYNKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AVSRQGKGYP TLKEDDLFF KKIEKGDIFL PERGNILISA IRPYNKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDENNEDLF KKIEKGDIFL PERGNILISA IRPYNKSVTP AEIGHVSKNG EVFPVTLSFE DRDENNEDLF KKIEKGDIFL PERGNILISA IRPYNKIVL KKIEKGDIFL PERGNILISA IRPYNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYYA IRTIFSEKIN AVSRQGKGYP TLKEDDLKTI QFSKKVIDNL LAKEEELISN IDALEKDIKE DESEPFKY ALSE AUGURANFS TFASEPSVRF DENNMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AUGURANFS TFASEPSVRF DRDENNEDLF KKIEKGDIFL PERGNILISA IRPYNKIVL KKIEKGDIFL PERGNILISA IRPYNKIVL KKIEKGDIFL PERGNILISA IRPYNKIVL IKEDDKTDY FTKAFIQIKP LINSRILYYA IKEDDKTDY FTKAFIQIKP LINSRILYYA IKEDDKTDY FTKAFIQIKP LINSRILYYA IKEDDKTDY FTKAFIQIKP LINSRILYYA IKEDDKTDY FTKAFIQIKP LINSRILYYA IKEDDKTDY FTKAFIQIKP LINSRILYYA IKEDCKTDY FTAFIQIKP LINSRILYYA IKEDCKTDY FTAFICAL ACVECAUNCH I IKEDCKTDY FTAFICAL ACVECAUNCH I IKEDCKTDY FTAFICAL ACVECAUNCH I IKEDCKTDY FTAFICAL ACVECAUNCH I IKEDCKTDY FTAFICAUNCH I INSRILYYA IKEDCKT	AONEIEKOIE	EKRNOIRVII	EETARS	LRTIFSEKIN	AVSROGKGYP	Т
TRD1-CR1-8aA (8-aa deletion in the predicted long a-helix at the C-terminus, 25.56 kDa)TRD1-CR1-TRD1-CR2 (TRD1 duplication, a.k.a., 2xTRD1 with CR1 and CR2, 54.11 kDa)MGLIQRNFS TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDELNEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYYA LRTIFSEKIN AVSRQCKGYP TLKEDDLKTI CFSKKVIDNL LAKEEELISN IDALEKDIKE LKSIQRSKKE TRD1-CR1-15aaA (15-aa deletion in the predicted long $\alpha$ -helix at the C-terminus, 24.69 kDa)CPSKKVIDNL DALEKDIKF PERGNILISA IRPYLNKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY ASIGNVKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY ASIGNKSKE LKSIQRSKKE TRASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY ASIGNKSKG EVFPVTLSFE DRDELNEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYYA SIGNUSSKNG EVFPVTLSFE DRDELNEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL AEIGHVSKNG EVFPVTLSFE DRDELNEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL ACVECK2 (ATRD1-CR1, protein could not be expressed in QFSKKVIDNL ACVECK2 (ATRD1-CR1, protein could not be expressed in MVELMAL DSQRRVDVGL SSISSLNSTI RVSTRNNKK LIQCYLVRDI DCIEPLGXYI TRSTSRNNKK LIQCYLVRDI DCIEPLGXYI TRSTSRNNKK LIQCYLVRDI DCIEPLGXYI TRSTSRNNKK KIQCYLVRDI DCIEPLGXYI TRSTSRNNKK KIQCYLVRDI DCIEPLGXYI TRSTSRNNKK KIQCYLVRDI DCIEPLGYIVIR MGLIQRRNFS TFASEPSVRF DFNYMKSVPF TTEEYTYKS LFEVVPSTVP TLDESEPFKY ACGNTUVKI SSNELLMYYL PIPPMEEQLE ACINGKSVPF TRSESPSFKY DFNYMKSVPF TTEEYTYKS LFEVVPSTVP TLDESEPF	_~_~~	~			~	
predicted long a-helix at the C-terminus, 25.56 kDa) MGLIQRRNFS TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDELNEDLF AEIGHVSKNG EVFPVTLSFE DRDELNEDLF AEIGHVSKNG EVFPVTLSFE DRDELNEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYVA LRTIFSEKIN AVSRQGKGYP TLKEDDLKTI QFSKKVIDNL LAKEEELISN IDALEKDIKE LKSIQRSKKE TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDELNEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL TRDI-CR1-15aa $\Delta$ (15-aa deletion in the LFEVVPSTVP TLDESEPFKY AEIGHVSKNG predicted long $\alpha$ -helix at the C-terminus, 24.69 kDa) MGLIQRNFS TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDELMEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYYA LRTIFSEKIN AVSRQGKGYP TLKEDDLKTI RTD1-CR1-21aa $\Delta$ (21-aa deletion in the predicted long $\alpha$ -helix the C-terminus, 23.99 kDa) MGLIQRNFS TFASEPSVRF DFNYMKSVTP TRD1-CR1-21aa $\Delta$ (21-aa deletion in the predicted long $\alpha$ -helix the C-terminus, 23.99 kDa) MGLIQRNFS TFASEPSVRF DFNYMKSVTP TRD1-CR1-21aa $\Delta$ (21-aa deletion in the predicted long $\alpha$ -helix the C-terminus, 23.99 kDa) MGLIQRNFS TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AEIGHVSKNG EVFPVTLSFE DRDELMEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL ETAKS	TRD1-CR1-8a	a $\Delta$ (8-aa del	etion in the	TRD1-CR1-TR	D1-CR2 (TRD1	duplication,
25.56 kDa) 54.11 kDa) MGLIQRRNFS TFASEPSVRF DFNYKSVTP TLDESEPFKY ALEGHVSKNG EVFPVTLSFE DRDELMEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKEDDKTDIY FTKAFIQIKP LINSRILYYA LATKFSEKIN AVSRQGKGYP TLKEDDLKTI QFSKKVIDNL LAKEEELISN IDALEKDIKE LKSIQRSKKE TRD1-CR1-15aa $\Delta$ (15-aa deletion in the predicted long $\alpha$ -helix at the C-terminus, 24.69 kDa) MGLIQRRNFS TFASEPSVRF DRYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY ALEGHVSKNG EVFPVTLSFE DRDELMEDLF KKIEKGDIFL PERGNILISA IRPYLNKIVL IKSIQRSKKE LKSIQRSKKE TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AVSRQGKGYP TLKEDDLKTI RTD1-CR1-15aa $\Delta$ (15-aa deletion in the predicted long $\alpha$ -helix at the C-terminus, 24.69 kDa) MGLIQRRNFS TFASEPSVRF DRYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY AUSRQGKGYP TLKEDDLKTI IKEDDKTDIY FTKAFIQIKP LINSRILYA LKEKGDIFL PERGNILISA IRPYLNKIVL LKEKGGIFL PERGNILISA IRPYLNKIVL LKEDKTDIY FTKAFIQIKP LINSRILYYA LRTIFSEKIN AVSRQGKGYP TLKEDDLKTI QFSKKVIDNL LAKEEELISN IDALEKDIKE LKSD WV TI TLDESEPFKY AUSRQGKGYP TLKEDDLKTI QFSKKVIDNL DALEKDIKE LKSD WV TI TLESEPFKY AUSRQGKGYP TLKEDDLKTI QFSKKVIDNL JAKEEELISN IDALEKDIKE LKSD KKIEKGDIFL PERGNILISA IRPYLNKIVL KKIEKGDIFL PERGNILISA IRPYLNKIVL LKEDDKTDIY FTKAFIQIKP LINSRILYYA LRTIFSEKIN AVSRQGKGYP TLKEDDLKTI TRD2-CR2 ( $\Delta$ TRD1-CR1, protein could not be expressed in QFSKKVIDNL LAKEEELISN IDALEKDIKE LKS WVELMAL MGLIQRNFS TFASEPSVRF DFNYMKSVTP TRD2-CR2 ( $\Delta$ TRD1-CR1, protein could not be expressed in QFSKKVIDNL LAKEEELISN IDALEKDIKE LKS WVELMAL RYSYRWNKK LIQKYLYRDI DCIEPLGKYI TRD2-CR2 ( $\Delta$ TRD1-CR1, protein could not be expressed in QFSKKVIDNL LAKEEELISN IDALEKDIKE LKS WVELMAL RYSYRWNKK LIQKYLYRDI DCIEPLGKYI TRD2-CR2 ( $\Delta$ TRD1-CR1, protein could not be expressed in QFSKKVIDNL LAKEEELISN IDALEKDIKE LKS WVELWAL RYSYRWNKK LIQKYLYRDI DCIEPLGKYI RYSYRWNKK LIQKYLYRDI DCIEPLGKYI RYSYRWNKK LIQKYLYRDI DCIEPLGKYI SSISSLNSTI RYSYRWNKK LIQKYLYRDI DCIEPLGKYI SSISSLNSTI RYSYRWNKK SSISSLNSTI RYSYRWNKK SYPYTLSFE DRDELMEDLF KKIEKGOIFL PERGNILISA IRPYLNKIVL ETAKS	predicted lo	ong a-helix a	at the C-terminus,	a.k.a., 2xT	RD1 with CR1	and CR2,
MGLIQRRNFSTFASEPSVRFDFNYMKSVTPMGLIQRRNFSTFASEPSVRFDFNYMKSVTPTTEEYYTYKSLFEVVPSTVPTLDESEPFKYTTEEYYTYKSLFEVVPSTVPTLDESEPFKYAEIGHVSKNGEVFPVTLSFEDRDELNEDLFAEIGHVSKNGEVFPVTLSFEDRDELNEDLFKKIEKGDIFLPERCNILISAIRPYLNKIVLKKIEKGDIFLPERCNILISAIRPYLNKIVLIKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYFTKAFIQIKPLINSRILYYALRTIFSEKINAVSRQGKGYPTLKEDDLKTILRTIFSEKINNAVSRQGKGYPTLKEDDLKTIQFSKKVIDNLLAKEEELISNIDALEKDIKEQFSKKVIDNLLAKEEELISNIDALEKDIKELKSIQRSKETFASEPSVRFDFNYMKSVTPTTEEYYTYKSTEASEPSVRFNGLIQRRNFSTRD1-CR1-15aaΔ(15-aa deletion in theLEVVPSTVPTLDESEPFKYAEIGHVSKNGpredicted long α-helix at theEVFPVTLSFEDRDELNEDLFKKIEKGDIFLC-terminus, 24.69 kDa)PERGNILISAIRPYLNKIVLIKEDDKTDIYMGLIQRRNFSTFASEPSVRFDFNYMKSVTPTLKEDDLKTIQFSKKVIDNLAEIGHVSKNGEVFPVTLSFEDRDELNEDLFPIPPMEEQLEIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERCNILISAIRPYLNKIVLEETARSIRPSYRNSIRPSYRNSIKEDDKTDIYFTKAFIQIKPLINSRILYALSNNGWSPESVGGGGIPILGQEHLEFDGIn the predicted long α-helixVLNVSPTKATTKKNNMENFFIQEGDLFISAin the predicted long α-helixVLNVSPTKATTKKNNMENFFIQEGDLFISin the predicted long α-helixVLNVSPTKA	25.56 kDa)			54.11 kDa)		
$\begin{array}{llllllllllllllllllllllllllllllllllll$	MGLIQRRNFS	TFASEPSVRF	DFNYMKSVTP	MGLIQRRNFS	TFASEPSVRF	DFNYMKSVTP
AEIGHVSKNGEVFPVTLSFEDRDELNEDLFAEIGHVSKNGEVFPVTLSFEDRDELNEDLFKKIEKGDIFLPERGNILISAIRPYLNKIVLKKIEKGDIFLPERCNILISAIRPYLNKIVLIKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYFTKAFIQIKPLINSRILYYAQFSKKVIDNLLAKEEELISNIDALEKDIKEQFSKKVIDNLLAKEEELISNIDALEKDIKELKSIQRSKKEIVDEVFSSKFNGLIQRNFSTTEEYYTKSTRD1-CR1-15aaΔ(15-aa deletion in theEFEVPSTVFDENVMKSVTPTTEEYYTKSTRD1-CR1-15aaΔ(15-aa deletion in theEVFPVTLSFEDRDELNEDLFKKIEKGDIFLC-terminus, 24.69 kDa)PERCNILISAIRPYLNKIVLIKEDDKTDIYMGLIQRNFSTFASEPSVRFDFNYMKSVTPFTKAFIQIKPLINSRILYYATTEEYYTYKSLFEVVPSTVPTLDESEPFKYAVSRQGKGYPAVSRQKWIDNLAEIGHVSKNGEVFPVTLSFEDRDELNEDLFPTPPMEEQLEIVGKIEEQIGKKIEKGDIFLPERGNILISAIRPYLNKIVLEKRNQIRVIIEETARSIKEDDKTDIYFTKAFIQIKPLINSRILYYATRD2-CR2 (ATRD1-CR1, protein could not be expressed in OFSKKVIDNLDSQRVDVGLSSISSLNSTI RYSYRNKMKKLKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRNKMKKIGQENLEFOGII II IPDLEFIGLKSTRD1-CR1-21aaA(21-aa deletionLSSNNGWSPESVVGGEGIPILGQEHLEFDGLKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRNKMKKLIQKYLYRDIDCIEPLGKYILKSMCLIQRRNFSTFASEPSVRFDRVMKSVTPKTNNWSPFTIQEGDLFISat th	TTEEYYTYKS	LFEVVPSTVP	TLDESEPFKY	TTEEYYTYKS	LFEVVPSTVP	TLDESEPFKY
KKIEKGDIFLPERGNILISAIRPYLNKIVLKKIEKGDIFLPERGNILISAIRPYLNKIVLIKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYFTKAFIQIKPLINSRILYYALRTIFSEKINAVSRQGKGYPTLKEDDLKTILRTIFSEKINAVSRQGKGYPTLKEDDLKTIQFSKKVIDNLLAKEEELISNIDALEKDIKEQFSKKVIDNLLAKEEELISNIDALEKDIKELKSIQRSKKEQFSKKVIDNLLAKEEELISNIDALEKDIKEIDALEKDIKEC-terminus, 24.69kaLFVPVTLSFEDRDLNEDLFKKIEKGDIFLC-terminus, 24.69PENYMKSVTPFTKAFIQIKPINSRILYALRTIFSEKINMGLIQRNFSTFASEPSVRFDFNYMKSVTPFTKAFIQIKPLINSRILYALRTIFSEKINMGLIQRNFSTFASEPSVRFDFNYMKSVTPFTKAFIQIKPLINSRILYALRTIFSEKINAEIGHVSKNGEVPFVTLSFEDRDELMEDLFPIPPMEEQLEIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERGNILISAIRPYLNKIVLETARSKIEKQIFLKEDDKTDIYFTKAFIQIKPLINSRILYAKIEKQIFETARSIKEDDKTDIYFTKAFIQIKPILKEDOKTDICOLEPLGKYIGFSKKVIDNLLAKEEELISNIDALEKDIKEE< coli)	AEIGHVSKNG	EVFPVTLSFE	DRDELNEDLF	AEIGHVSKNG	EVFPVTLSFE	DRDELNEDLF
IKEDDKTDIYFTKAFIQIKPLINSRILYYAIKEDDKTDIYFTKAFIQIKPLINSRILYYALRTIFSEKINAVSRQGKGYPTLKEDDLKTILRTIFSEKINAVSRQGKGYPTLKEDDLKTIQFSKKVIDNLLAKEEELISNIDALEKDIKEQFSKKVIDNLLAKEEELISNIDALEKDIKELKSIQRSKKEUDEVFSSHFNGLIQRNFSTRD1-CR1-15aa $\Delta$ (15-aa deletion in theLFEVVPSTVPTLDESEPFKYAEIGHVSKNGpredicted long $\alpha$ -helix at theEVFVTLSFEDRDELNEDLFKKIEKGDIFLC-terminus, 24.69 kDa)PERGNILISAIRPYLNKIVLIKEDDKTDIYMGLIQRNFSTFASEPSVRFDFNYMKSVTPFTKAFIQIKPLINSRILYATTEEYYTYKSLFEVVPSTVPTLDESEPFKYAVSRQGKGYPTLKEDDLKTIAEIGHVSKNGEVFPVTLSFEDRDELNEDLFPTPPMEEQLEIVGKIEEQIGAQNEIEKQIFPERGNILISAIRPYLNKIVLEKRNQIRVIIEETARSIKEDDKDIYFTKAFIQIKPLINSRILYYASISSLNSTIRYSRVIDNLLAKEEELISNIDALEKDIKEE. coli)LKSMVELMALDSQRVDVGLSSISSLNSTIRYSRWNKKKLIQKYLYRDIDCIEPLGKYIIKEDD-CR1-21aa $\Delta$ (21-aa deletionLSSNNGWSPESVVGEGEIPIin the predicted long $\alpha$ -helixVLNVSPTKATTKTKNNMENFFIQEDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIYPDLYIRMGLIQRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFTTEYSTYKSLFEVVPSTVPTLDESEPFKYKGKNQTWKISSNELLNYLMGLIQRNFSTFASEPSVRFDFNYMKSVTP </td <td>KKIEKGDIFL</td> <td>PERGNILISA</td> <td>IRPYLNKIVL</td> <td>KKIEKGDIFL</td> <td>PERGNILISA</td> <td>IRPYLNKIVL</td>	KKIEKGDIFL	PERGNILISA	IRPYLNKIVL	KKIEKGDIFL	PERGNILISA	IRPYLNKIVL
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	IKEDDKTDIY	FTKAFIQIKP	LINSRILYYA	IKEDDKTDIY	FTKAFIQIKP	LINSRILYYA
$\begin{array}{llllllllllllllllllllllllllllllllllll$	LRTIFSEKIN	AVSRQGKGYP	TLKEDDLKTI	LRTIFSEKIN	AVSRQGKGYP	TLKEDDLKTI
$ \begin{array}{llllllllllllllllllllllllllllllllllll$	QFSKKV <u>IDNL</u>	LAKEEELISN	IDALEKDIKE	QFSKKV <u>IDNL</u>	LAKEEELISN	IDALEKDIKE
$\begin{tabular}{lllllllllllllllllllllllllllllllllll$	LKSIQRSKKE			<u>lksiqrskke</u>	<u>IVDEVFSS</u> HF	NGLIQRRNFS
$ \begin{array}{llllllllllllllllllllllllllllllllllll$				TFASEPSVRF	DFNYMKSVTP	TTEEYYTYKS
predicted long $\alpha$ -helix at the EVFPVTLSFE DRDELNEDLF KKIEKGDIFL C-terminus, 24.69 kDa) PERGNILISA IRPYLNKIVL IKEDDKTDIY MGLIQRRNFS TFASEPSVRF DFNYMKSVTP FTKAFIQIKP LINSRILYYA LRTIFSEKIN AEIGHVSKNG EVFPVTLSFE DRDELNEDLF PIPPMEEQLE IVGKIEEQIG AQNEIEKQIE KKIEKGDIFL PERGNILISA IRPYLNKIVL EKRNQIRVII EETARS IKEDDKTDIY FTKAFIQIKP LINSRILYYA LRTIFSEKIN AVSRQGKGYP TLKEDDLKTI TRD2-CR2 ( $\Delta$ TRD1-CR1, protein could not be expressed in QFSKKVIDNL LAKEEELISN IDALEKDIKE E. coli) LKS MVELMAL DSQRRVDVGL SSISSLNSTI RYSYRWNKMK LIQKYLYRDI DCIEPLGKYI TRD1-CR1-21aa $\Delta$ (21-aa deletion LSSNNGWSPE SVVGGEGIPI LGQEHLEFDG in the predicted long $\alpha$ -helix VLNVSPTKAT TKTKNNMENF FIQEGDLFIS at the C-terminus, 23.99 kDa) RGNTVDLVGL ACVVETEVTE DIIYPDLYIR MGLIQRRNFS TFASEPSVRF DFNYMKSVTP LKIDEKVIKK KYLALLFNSF FGRLYFKYVS TTEEYYTYKS LFEVVPSTVP TLDESEPFKY KGKNQTMVKI SSNELLNYYL PIPPMEEQLE AEIGHVSKNG EVFPVTLSFE DRDELNEDLF IVGKIEEQIG AQNEIEKQIE EKRNQIRVII KKIEKGDIFL PERGNILISA IRPYLNKIVL EETARS	TRD1-CR1-15	aa $\Delta$ (15-aa d	eletion in the	LFEVVPSTVP	TLDESEPFKY	AEIGHVSKNG
C-terminus, 24.69 kDa)PERGNILISAIRPYLNKIVLIKEDDKTDIYMGLIQRRNFSTFASEPSVRFDFNYMKSVTPFTKAFIQIKPLINSRILYYALRTIFSEKINTTEEYYTYKSLFEVVPSTVPTLDESEPFKYAVSRQGKGYPTLKEDDLKTIQFSKKVIDNLAEIGHVSKNGEVFPVTLSFEDRDELNEDLFPIPPMEEQLEIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSIRPYLNKIVLIKEDDKTDIYIKEDDKTDIYFTKAFIQIKPLINSRILYYATRD2-CR2 (ATRD1-CR1, protein could not be expressed in QFSKKVIDNLDSQRRVDVGLSSISSLNSTI RYSYRWNKMKLIQKYLYRDILKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRWNKMKLIQKYLYRDIDCIEPLGKYITRD1-CR1-21aaΔ(21-aa deletion in the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSEETARS	predicted lo	ong α-helix a	it the	EVFPVTLSFE	DRDELNEDLF	KKIEKGDIFL
MGLIQRRNFSTFASEPSVRFDFNYMKSVTPFTKAFIQIKPLINSRILYYALRTIFSEKINTTEEYYTYKSLFEVVPSTVPTLDESEPFKYAVSRQGKGYPTLKEDDLKTIQFSKKVIDNLAEIGHVSKNGEVFPVTLSFEDRDELNEDLFPIPPMEEQLEIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERGNILISAIRPYLNKIVLEKRNQIRVIIEETARSIKEDDKTDIYFTKAFIQIKPLINSRILYYATRD2-CR2 (ΔTRD1-CR1, protein could not be expressed in E. coli)QFSKKVIDNLLAKEEELISNIDALEKDIKEE. coli)LKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRWNKMKLIQKYLYRDITRD1-CR1-21aaΔ(21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARS	C-terminus,	24.69 kDa)		PERGNILISA	IRPYLNKIVL	IKEDDKTDIY
TTEEYYTYKSLFEVVPSTVPTLDESEPFKYAVSRQGRGYPTLREDDLKTIQFSKKVIDNLAEIGHVSKNGEVFPVTLSFEDRDELNEDLFPIPPMEEQLEIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERGNILISAIRPYLNKIVLEKRNQIRVIIEETARSIKEDDKTDIYFTKAFIQIKPLINSRILYYATRD2-CR2 (ΔTRD1-CR1, protein could not be expressed in E. coli)LKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRWNKMKTRD1-CR1-21aaΔ(21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSEETARS	MGLIQRRNFS	TFASEPSVRF	DFNYMKSVTP	FTKAFIQIKP	LINSRILYYA	LRTIFSEKIN
AEIGHVSKNGEVFPVTLSFEDRDELNEDLFPIPPMEEQLEIVGKIEEQIGAQNEIEKQIEKKIEKGDIFLPERGNILISAIRPYLNKIVLEKRNQIRVIIEETARSIKEDDKTDIYFTKAFIQIKPLINSRILYYATRD2-CR2 (ΔTRD1-CR1, protein could not be expressed in QFSKKVIDNLDSQRRVDVGLSSISSLNSTI RYSYRWNKMKLKSIDALEKDIKEE. coli)LKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRWNKMKLIQKYLYRDIDCIEPLGKYITRD1-CR1-21aaΔ (21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSEETARS	TTEEYYTYKS	LFEVVPSTVP	TLDESEPEKY	AVSRQGKGYP	TLKEDDLKTI	QFSKKVIDNL
KKIEKGDIFLPERGNILISAIRPYLNKIVLEKKNQIKVIIEEIAKSIKEDDKTDIYFTKAFIQIKPLINSRILYYATRD2-CR2 (ΔTRD1-CR1, protein could not be expressed in QFSKKVIDNLLRTIFSEKINAVSRQGKGYPTLKEDDLKTITRD2-CR2 (ΔTRD1-CR1, protein could not be expressed in E. coli)LKSMVELMALDSQRRVDVGLSSISSLNSTI RYSYRWNKMKTRD1-CR1-21aaΔ(21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long α-helixVLNVSPTKATTKTKNNMENFat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKMGLIQRRNFSLFEVVPSTVPTLDESEPFKYKGKNQTMVKIAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIFLPERGNILISAIRPYLNKIVLEETARS	AEIGHVSKNG	EVFPVTLSFE		PIPPMEEQLE	IVGKIEEQIG	AQNEIEKQIE
IKEDDRIDTYFIRAFIQIRPLINSRILYYALRTIFSEKINAVSRQGKGYPTLKEDDLKTITRD2-CR2 (ΔTRD1-CR1, protein could not be expressed inQFSKKVIDNLLAKEEELISNIDALEKDIKEE. coli)LKSMVELMALDSQRRVDVGLSSISSLNSTIRYSYRWNKMKLIQKYLYRDIDCIEPLGKYITRD1-CR1-21aaΔ(21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSFGRLYFKIE	KKIEKGDIFL	PERGNILISA	IRPYLNKIVL	<u>EKRNQIRVII</u>	<u>EETARS</u>	
LRTIFSERINAVSRQGRGYPTEREDURTITRD2-CR2 (ATRD1-CR1, protein could not be expressed in $LKS$ LKSIDALEKDIKEE. coli)TRD1-CR1-21aa $\Delta$ (21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long $\alpha$ -helixVLNVSPTKATTKTKNNMENFat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLDLYSKALKUDEKVIHKKYLALLFNSFFGRLYFKYSLFEVVPSTVPTLDESEPFKYAEIGHVSKNGEVFPVTLSFEDRDELNEDLFKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSEETARS	IKEDDKIDIY	FIKAFIQIKP	LINSRILYYA			
LKSMVELMALDSQRRVDVGLSSISSLNSTITRD1-CR1-21aaΔ(21-aa deletionLSSNNGWSPESVVGGEGIPILGQEHLEFDGin the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSFGRLYFKYE	LRIIF SEKIN	AVSRQGKGIP	ILKEDDLKII	$I RD2-CR2 (\Delta I E coli)$	RDI-CRI, protein	n could not be expressed in
LKSMVELMALDSQRRVDVGLSSISSENSITRYSYRWNKMKLIQKYLYRDIDCIEPLGKYITRD1-CR1-21aaΔ (21-aa deletionLSSNNGWSPESVVGGEGIPIin the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYAEIGHVSKNGEVFPVTLSFEDRDELNEDLFKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSEKRNQIRVII	VL SVUA TONT	TAVEFETION	IDALENDIKE		DCODDUDUCI	COTOCINCTI
TRD1-CR1-21aa $\Delta$ (21-aa deletionLSSNNGWSPESVVGGEGIPILGQEHLEFDGin the predicted long $\alpha$ -helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSE	<u>C71L</u>			MAGADMMKWK WARTUWAT	DOGUKADAGT	DCIEDICKAI
InstructureInstructureInstructureInstructureInstructurein the predicted long α-helixVLNVSPTKATTKTKNNMENFFIQEGDLFISat the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSE	TRD1-CR1-21	aaA (21-22 d	eletion	LSSNNCWSPF	SVVCCECTPT	LGOEHLEFDG
at the C-terminus, 23.99 kDa)RGNTVDLVGLACVVETEVTEDIIYPDLYIRMGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSE	in the nred	icted long of-	-helix	VINVSPTKAT	TKTKNNMENE	FIOEGDLEIS
MGLIQRRNFSTFASEPSVRFDFNYMKSVTPLKIDEKVIHKKYLALLFNSFFGRLYFKYVSTTEEYYTYKSLFEVVPSTVPTLDESEPFKYKGKNQTMVKISSNELLNYYLPIPPMEEQLEAEIGHVSKNGEVFPVTLSFEDRDELNEDLFIVGKIEEQIGAQNEIEKQIEEKRNQIRVIIKKIEKGDIFLPERGNILISAIRPYLNKIVLEETARSE	at the C-ter	rminus, 23 90	) kDa)	RGNTVDLVGL	ACVVETEVTE	DITYPDLYIR
TTEEYYTYKS     LFEVVPSTVP     TLDESEPFKY     KGKNQTMVKI     SSNELLNYYL     PIPP <u>MEEQLE</u> AEIGHVSKNG     EVFPVTLSFE     DRDELNEDLF     IVGKIEEQIG     AQNEIEKQIE     EKRNQIRVII       KKIEKGDIFL     PERGNILISA     IRPYLNKIVL     EETARS     EKRNQIRVII	MGLIORRNES	TFASEPSVRF	DENYMKSVTP	LKIDEKVIHK	KYLALLENSE	FGRLYFKYVS
AEIGHVSKNG     EVFPVTLSFE     DRDELNEDLF     IVGKIEEQIG     AQNEIEKQIE     EKRNQIRVII       KKIEKGDIFL     PERGNILISA     IRPYLNKIVL     EETARS	TTEEYYTYKS	LFEVVPSTVP	TLDESEPFKY	KGKNOTMVKT	SSNELLINYYI.	PIPPMEEOLE
KKIEKGDIFL PERGNILISA IRPYLNKIVL <u>EETARS</u>	AEIGHVSKNG	EVFPVTLSFE	DRDELNEDLF	IVGKIEEOIG	AQNEIEKOIE	EKRNQIRVII
	KKIEKGDIFL	PERGNILISA	IRPYLNKIVL	EETARS	_ <u>,                                     </u>	<b>`</b>

```
CR1-TRD2-CR2 (\DeltaTRD1, 33.16 kDa)
```

MLLAKEEELI	<u>SNIDALEKDI</u>	<u>KELKSIQRSK</u>
KEIVDEVFSS	HFNINMVELM	ALDSQRRVDV
GLSSISSLNS	TIRYSYRWNK	MKLIQKYLYR
DIDCIEPLGK	YILSSNNGWS	PESVVGGEGI
PILGQEHLEF	DGVLNVSPTK	ATTKTKNNME
NFFIQEGDLF	ISRGNTVDLV	GLACVVETEV
TEDIIYPDLY	IRLKIDEKVI	HKKYLALLFN
SFFGRLYFKY	VSKGKNQTMV	KISSNELLNY
YLPIPP <u>MEEQ</u>	LEIVGKIEEQ	<u>IGAQNEIEKQ</u>
IEEKRNOIRV	LIEETARS	

Circular permutation of TRDs: A rearranged TRD domains: TRD2-CR2-TRD1-CR1 (54.85 kDa)

```
MVELMAL.
           DSORRVDVGL SSISSLNSTI
RYSYRWNKMK LIQKYLYRDI
                       DCIEPLGKYI
LSSNNGWSPE SVVGGEGIPI LGQEHLEFDG
VLNVSPTKAT TKTKNNMENF FIOEGDLFIS
RGNTVDLVGL ACVVETEVTE DIIYPDLYIR
LKIDEKVIHK KYLALLFNSF FGRLYFKYVS
KGKNQTMVKI
           SSNELLNYYL PIPPMEEQLE
IVGKIEEQIG AQNEIEKQIE
                       EKRNQIRVII
EETARS
GLIORRNFS
           TFASEPSVRF
                       DFNYMKSVTP
TTEEYYTYKS
           LFEVVPSTVP
                       TLDESEPFKY
           EVFPVTLSFE
AEIGHVSKNG
                       DRDELNEDLF
KKIEKGDIFL
           PERGNILISA
                       IRPYLNKIVL
IKEDDKTDIY
           FTKAFIQIKP
                       LINSRILYYA
                       TLKEDDLKTI
           AVSROGKGYP
LRTIFSEKIN
            LAKEEELISN
                       IDALEKDIKE
QFSKKVIDNL
LKSIQRSKKE
           IVDEVFSSHF
                       NIN.
```

### Replacing *Bsa*XI TRD2 With a Target Recognition Domain Homolog From Another Bacterium

CspC0110I TRD homolog (partial) was found in GenBank by BlastP search with 59% aa sequence identity to *Bsa*XI TRD2 (see below for GenBank accession number). It was annotated as a Type I HsdS partial sequence in *Cyanothece* sp. CCY0110.

Hypothetical Type I restriction enzyme EcoEI specificity protein (S protein, partial)

WP\_008279016, 273 aa, linear, BCT 05-JUN-2013.

hypothetical Type I restriction enzyme EcoEI specificity protein (S protein), partial [*Cyanothece* sp. CCY0110].

WP\_008279016.1, GI:495554437. ORGANISM: Cyanothece sp. CCY0110 (Bacteria; Cyanobacteria; Oscillatoriophycideae;

Chroococcales; Cyanothece).

Protein: 1.273.

Product: hypothetical Type I restriction enzyme, EcoEI specificity protein (S protein).

Note: EcoKI restriction-modification system protein HsdS; Provisional.

1 enikfiirsk kedsiiinei fgkefnisls dissidtnsk ihinlsdiyf gnnnirnsfr 61 wnkiqliqkf iyknidciql lgnfiiktkn

```
gwsppsiegg ngipilggeh fgsdgvlkil
121 ptkftedtkn nienyfigeg dffvsrgnti
dlvalasvve eeisedilfp dlyikvklde
181 tvidkkylal lfnsffgrly fkyvskgkng
tmvkissrel vnfvlpipdi kkgkkivegi
241 tdkidegski nkkiekniak inliieesin gng
TRD1-CR1-[CspC0110I TRD]-CR2 (55.04 kDa)
MGLIORRNFS TFASEPSVRF DFNYMKSVTP
TTEEYYTYKS LFEVVPSTVP TLDESEPFKY
AEIGHVSKNG EVFPVTLSFE DRDELNEDLF
KKIEKGDIFL PERGNILISA IRPYLNKIVL
IKEDDKTDIY FTKAFIQIKP LINSRILYYA
LRTIFSEKIN AVSROGKGYP TLKEDDLKTI
QFSKKVIDNL LAKEEELISN IDALEKDIKE
<u>LKSIQRSKKE</u> <u>IVDEVFSSHF</u><u>N</u> isls
dissidtnsk ihinlsdiyf gnnnirnsfr
wnkiqliqkf iyknidciql lgnfiiktkn
gwsppsiegg ngipilgqeh fgsdgvlkil
ptkftedtkn nienyfigeg dffvsrgnti
dlvalasvve eeisedilfp dlyikvklde
tvidkkylal lfnsffgrly fkyvskgkng
tmvkissrel ynfyl
PIPP MEEQLE IVGKIEEQIG AQNEIEKQIE
EKRNQIRVII EETARS
```

Chimeric S subunit TRD1-CR1-[CstTRD]-CR2

The *Bsa*XI TRD2 and Cst TRD2 share 28% as sequence identity. The Cst S subunit is annotated as a Type I

specificity protein in GenBank (accession number WP\_005530828).

LOCUS: WP\_005530828 371 aa

DEFINITION: type I restriction modification DNA specificity protein

SOURCE: Corynebacterium striatum

ORGANISM classification: Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

Corynebacterineae; Corynebacteriaceae; Corynebacterium.

```
1 mastakhdwp mvrlgdvchl kygkalkkee
rvagefpvfg sagsvgshve anfvgpvsvv
61 grkgsagfve wssgncwiid tafgvfpkse
eqvdsrwlyw llkdlrlgrl qkhaavpgis
121 kadvveekfl lppldeqrri aaildevdea
lfrvnqslgd llqlkqelft dlflrieres
181 tiigeylest qygtsdkane nvgipilrmg
nvsyngeidl sdlkyvelda sdrekyslka
241 gdllfnrtns kdlvgktavv pelqeeytya
gylircrvnd kavpeyisgf lnsvlgkkil
301 rntakaivgm aninanelkr lpipqaslde
qqefasltsr iddvesqmkr qrkllqelqe
361 slstrafqee l
```

Chimeric S subunit: TRD1-CR1-[Cst TRD]-CR2 (no protein expression) MGLIQRRNFS TFASEPSVRF DFNYMKSVTP TTEEYYTYKS LFEVVPSTVP TLDESEPFKY

AEIGHVSKNG	EVFPVTLSFE	DRDELNEDLF
KKIEKGDIFL	PERGNILISA	IRPYLNKIVL
IKEDDKTDIY	FTKAFIQIKP	LINSRILYYA
LRTIFSEKIN	AVSRQGKGYP	TLKEDDLKTI
QFSKKVIDNL	LAKEEELISN	IDALEKDIKE LKSIQRSKKE
<b>IVDEVFSSHF</b>	<u>N</u> IN MVE	
lppldeqrri	aaildevdea	lfrvnqslgd
llqlkqelft	dlflrieres	
tiigeylest	qygtsdkane	nvgipilrmg
nvsyngeidl	sdlkyvelda	sdrekyslka
gdllfnrtns	kdlvgktavv	pelqeeytya
gylircrvnd	kavpeyisgf	lnsvlgkkil
rntakaivgm	aninanelkr	<u>lpippMEEQLE</u>
IVGKIEEQIG	AQNEIEKQIE	EKRNQIRVII EETARS

### RESULTS

### Cloning of *BsaXI* Restriction–Modification Genes in *Escherichia coli*

The BsaXI restriction-modification (R-M) system was originally identified from random shotgun sequencing of B. stearothermophilus 25B (Geobacillus sp. 25B) gDNA (data are not shown). A part of the sequences was verified by the primer walking using Sanger sequencing. The R-M system contains two genes, the bsaXIRM gene encoding the restrictionmodification subunit (RM fusion, 911 aa, predicted molecular mass 107.04 kDa) and the bsaXIS gene coding for the specificity subunit (S, 476 aa, molecular mass 54.97 kDa). The RM and S genes are probably regulated in a single transcription unit with RM gene preceding the S gene. The DNA sequence has been deposited in GenBank (accession number OM373208). The endonuclease domain in the N-terminal region of RM subunit contains two potential PD-D/ExK catalytic sites (D-X16-DxK or D-X17-ExK). It is predicted that the second motif may form the catalytic site since it is analogous to the catalytic residues of BcgI: PE-X12-ExK (Kong, 1998). The exact catalytic residues remain to be confirmed experimentally. The methylase domain in the C-terminal region contains typical N6mA (6 mA) methyltransferase (MTase) conserved sequence motifs. The methylase activity is predicted to modify the adenine (A) in the target site 5'  $\downarrow$ N9 AC N5 CTCC N10 $\downarrow$  3', and the adenine in the bottom strand opposite to the T base. The MTase activity of the purified BsaXI RM-S enzyme complex is low and only provided partial modification in vitro (see below). The sequence alignment with the S subunits of other Type I and Type IIB systems by Phyre2 indicated that the S subunit contains protein subdomains TRD1-CR1-TRD2-CR2 in which CR1 and CR2 are predicted to form long *a*-helix coiled-coil and interact with the RM subunits (see below for more detailed analysis of TRDs, Supplementary Figure 1). The BsaXI RM and S gene products are nearly identical to two putative R-M systems listed in REBASE: The 99-100% sequence identity to Gth3921I and Gka8005I (RM fusion, WP\_052369193; S subunit, WP\_052369191; Geobacillus kaustophilus NBRC 102445 DNA

sequence contig, NZ\_BBJV01000045), suggesting that the *Bsa*XI R–M system (or isoschizomers) may have been evolved in other *Geobacillus* strains *via* horizontal gene transfer. The homologs of *Bsa*XI R–M systems are widely distributed in other microbial genomes: More than 60 RM fusion homologs, with 44–100% aa sequence identity, are found in GenBank in a recent BlastP search (data are not shown).

### Expression and Purification of *BsaXI* Restriction Endonuclease

The co-expression of BsaXI RM and S genes in pUC19 appeared to be unstable (data are not shown). The reason for this might be insufficient methylation. Therefore, the two genes were expressed separately in fusion with intein and chitin binding domain (CBD) in pTYB1. The fusion to intein and CBD may also contribute to the reduced toxicity of RM expression. The cell extracts containing RM and S subunits could be mixed in test tube and the RM/S subunits in the complex was co-purified. Figure 1A shows the co-purified BsaXI after chromatography through six columns (Chitin, DEAE, Heparin Sepharose, Source 15 S, Source 15Q, and gel filtration). Based on the protein molecular mass and band intensity (Figure 1A and Supplementary Figure 2), it was estimated that enzyme stoichiometry of RM:S is 2:1 [band intensity (total pixels ratio) = 2:1.2 as measured by gel imaging software, Bio Rad], and the active enzyme complex can be shown as trimeric complex [RM]2 S similar to the first characterized Type IIB enzyme BcgI (Kong, 1998). However, we cannot rule out the possibility of forming dimer of trimer ([RM]2 S)2 in the presence of cognate DNA. Two BsaXI subunits could be purified separately by chromatography through chitin and heparin columns (Figure 1B). To increase the expression level for BsaXI S subunit, we also constructed an expression clone in pET21b with 6xHis tag. Then BsaXI S (6xHis) was purified by chromatography through a nickel agarose column (Supplementary Figure 2). Figure 2A shows the restriction activity of co-purified BsaXI endonuclease. At high enzyme concentrations, the activity was inhibited either due to protein aggregation or to a lack of available target sites as "overcrowded" enzyme non-specific binding. The BsaXI activity can also be reconstituted by mixing the partially purified RM and S (6xHis) subunits to achieve partial digestion (Figure 2B). Increasing BsaXI RM subunit concentration slightly improved the cleavage reaction, but still showing partial digestion (Supplementary Figure 3). The activity of in vitro reconstituted BsaXI activity by mixing the two subunits together in digestion of  $\lambda$ -DNA is lower than the co-purified enzyme complex, probably due to slow subunit unit association, non-optimal subunit ratio or non-productive complex formation.

# Alteration and Rearrangement of *BsaXI* S Subunit

TRD1-CR1 (Deletion of TRD2-CR2), TRD1-CR1-8Aa $\Delta$ , TRD1-CR1-15aa $\Delta$ , TRD1-CR1-21 aa $\Delta$ , TR $\Delta$ 1-CR1-57aa $\Delta$ 

The C-terminal region of the S subunit (TRD2-CR2) was deleted to construct an S variant TRD1-CR1. The synthetic gene



encoding TRD1-CR1 was cloned into pTYB1 and the insert was verified by Sanger sequencing. Following IPTG induction the protein was purified from a chitin column (**Supplementary Figure 2**) and its endonuclease activity was reconstituted with

purified RM subunits. No BsaXI sites are present in pBR322 so any cleavage of the substrate by the S variant/RM should be a new cleavage specificity. The pBR322 was digested into small fragments of less than 1 kb as detected in the agarose gel in Figure 3A (lanes 1-2). The TRD1-CR deletion variants with 8-32 aa deleted also showed partial activity (Figure 3B, lanes 3-8). The TRD1-CR1 with 21-aa deletion is also partially active (data are not shown). Figure 3C shows the virtual restriction digestion of pBR322 by NEBcutter (Vincze et al., 2003) on predicted 5' AC N5 GT 3' sites. The digested DNAs were subjected to run-off sequencing and the cut sites were found N9-12 bases outside of recognition sequence 5' AC N5 GT 3' (Supplementary Figures 4A,B). The cleavages at more than 12 nt distance were also detected in TRD1-CR1/RM digested DNA, but it is not clear whether the wobble cleavage is carried out by the TRD1-CR1/RM complex or by a contaminating exonuclease or cleavage directed from star sites. We also detected a cleavage outside of a star site at 5' TC N5 GT N9 $\uparrow$  3' (Supplementary Figure 4C). It was concluded that TRD1-CR1 in complex with BsaXI RM subunit created a new restriction specificity (AC N5 GT) and cleavage took place outside of its recognition sequence. Part of CR1 in the long α-helix in TRD1-CR1 can be deleted up to 32aa residues, but the deletion variants in CR1 region impaired the endonuclease activity. It was somewhat unexpected that large deletion in the CR1 region did not alter the spacer N5 in AC N5 GT. We also attempted to delete the entire CR1 region, but an expression clone with 42-aa deletion could not be constructed. The deletion of the entire CR1 plus 15-aa residues in the TRD1 domain failed to express the mutant protein, suggesting that the 57-aa deletion affected the protein folding and expression negatively.





# TRD1-CR2 (Deletion of TRD2 and Replacing CR1 by CR2)

We also constructed and purified TRD1-CR2 (CR1 was replaced by CR2). This S variant generated partial restriction activity on pBR322 in complex with the RM subunits (**Figure 3B**, lanes 4–8). It was concluded that the long  $\alpha$ -helix forming CR2 can replace CR1 and TRD1-CR2 can still form active complex with the RM subunits to create limited partial activity. It would be interesting to swap in other related CRs in Type I and IIB S subunits and test their interactions with *Bsa*XI RM subunits. It was concluded that the CR1 sequence is not absolutely required for TRD1 to interact with the RM subunits. A partial deletion of the CR1 segment did not completely eliminate TRD1 interactions with the RM subunits (see in the following).

# TRD1 Duplication: TRD1-CR1-TRD1-CR2 (Deletion of TRD2)

We also attempted expression and purification of 2xTRD1 (TRD1-CR1-TRD1-CR2) in which TRD2 was replaced by TRD1. The S variant was purified from a chitin column and restriction

activity was reconstituted with purified RM subunits. The duplicated TRD1 S variant generated partial restriction activity (data are not shown). The digested DNA was subjected to DNA run-off sequencing and the recognition sequence was confirmed to be 5' AC N5 GT 3', but the cut sites are variable. When a DNA template strand is cut (nicked), the sequencing Taq DNA polymerase adds an extra peak "A" (adenine) to the sequencing read after the broken backbone (i.e., the template independent terminal nucleotide transferase activity), creating doublets such as A/T, A/C, or A/G. If the original base call

is "A," then the extra overlapping "A" will create a high "A." The sequencing peaks after the broken template will significantly drop off, thus termed DNA run-off sequencing. In a partially digested template, some DNA molecules are cut/nicked, while others remain intact and the base calls sometimes continue after the sudden peak drop off. Three examples of cut sites are shown in **Figure 4**. The cleavage took place either upstream or downstream as asymmetric cleavage AC N5 GT N9-12, which differs from coordinated WT cleavages on both 5' and 3' of its recognition sequence 5' AC N5 CTCC 3'. We concluded



peak or a sudden drop in peak height.



from this experiment that 2xTRD1 (TRD1-CR1-TRD1-CR2) can new p interact with the RM subunits and activate cleavage, although cleavage near the symmetric sequence 5' AC N5 GT 3' appeared

# TRD1 Deletion: TRD2-CR2 (No Protein Expression) and CR1-TRD2-CR2 (Nicking)

We attempted cloning and expression of TRD2-CR2 in *E. coli*. The expression plasmid can be constructed, but no protein was detected after IPTG induction and chitin column purification. The reason for TRD2-CR2 poor expression is unknown. However, CR1-TRD2-CR2 protein can be expressed and purified (**Supplementary Figure 2**). This variant generated nicked pBR322 when it was reconstituted with the RM subunits (data are not shown). The nicking sites generated by CR1-TRD2-CR2/RM remain to be determined. The 8-bp site (GGAG N5 CTCC) is not present in pBR322, pBC4 (pUC19 with an adenovirus DNA insert, see NEBcutter V3 at NEB.com) or phage  $\lambda$  DNA. Thus,

new plasmid DNA substrate needs to be constructed to detect sequence-specific nicking or cleavage.

## Circular Permutation of *BsaXI* Target Recognition Domains: TRD2-CR2-TRD1-CR1

It has been observed that TRDs in a Type I RM system can be shuffled with circular permutation (Loenen et al., 2014). Due to the modular organization of TRD1 and TRD2, we constructed and purified the rearranged *Bsa*XI TRDs in circular permutated form TRD2-CR2-TRD1-CR1. The S variant is active in cleavage of M13 dsDNA in complex with the RM subunits. **Figure 5** shows the run-off sequencing of two *Bsa*XI sites 5' GGAG N5 GT 3' in M13 dsDNA, which confirmed the recognition sequence and slightly shifted cleavage distance at N12-13 and N10-13 over the longer staggered cuts. The shorter staggered cuts were not affected at N7 and N9. It is not clear why TRD2-CR2 variant

to be asymmetric.

TABLE 1 | Summary of BsaXI TRD rearrangement and mutagenesis.

Specificity (S) subunit	Activity with WT RM	Expression level	Note
WT: TRD1-CR1-TRD2-CR2	+ ++	+	AC N5 CTCC
TRD1-CR1	+ +	+	AC N5 GT
TRD1-CR2	+ (partial)	+	AC N5 GT
TRD1-CR1*-8aaD	+ (partial)	+	AC N5 GT
TRD1-CR1*-15aaD	+ (partial)	+	AC N5 GT
TRD1-CR1*-21aaD	+ (partial)	+	AC N5 GT
TRD1-CR1*-32aaD	+ (partial)	+	AC N5 GT
TRD1-CR1*-57aaD	NT (not tested)	Poor expression	
TRD1-CR1-TRD1-CR2 (2xTRD1)	+ (partial)	+	AC N5 GT
TRD2-CR2	NT	poor expression	?
CR1-TRD2-CR2	+	+	Nicking activity (Specificity not determined)
TRD2-CR2-TRD1-CR1 (circular permutation)	++	+	GGAG N5 GT (Imprecise cuts)
TRD1-CR1-[CspC0110l TRD]-CR2 (TRD2 replaced by a homolog)	++	+	AC N5 CTCC
TRD1-CR1-[CstTRD]-CR2	NT	poor expression	

Restriction activity was reconstituted by mixing WT RM subunits with the S subunit variants. CR1\*, aa deletion variants in the CR region. CspC0110I TRD, a TRD2 homolog found in the genome of Cyanothece sp. CCY0110. Cst TRD, a distant TRD2 homolog found in the genome of Corynebacterium striatum. Poor expression means that no target protein of the predicted molecular mass was detected in SDS-PAGE analysis.

could not be expressed in *E. coli*, but S variants CR1-TRD2-CR2 and TRD2-CR2-TRD1-CR1 could be expressed and purified.

### Replacing *Bsa*XI TRD2 With a Target Recognition Domain Homolog With 59% Aa Sequence Identity: TRD1-CR1-TRD2\*-CR2

A *Bsa*XI TRD2 homolog CspC0110I TRD (partial sequence) was found in GenBank by a BlastP search. The two TRDs share 59% aa sequence identity (see Materials and Methods section for the aa sequence). It was annotated as a Type I HsdS partial sequence in *Cyanothece* sp. CCY0110 which is similar to EcoEI HsdS. We constructed and purified the chimeric protein consisting of TRD1-CR1-[CspC0110T TRD]-CR2. The restriction activity was reconstituted when the chimeric S subunit is combined with the *Bsa*XI RM subunits and the cleavage pattern is identical to *Bsa*XI (data are not shown).

Another S variant (Chimeric S subunit: TRD1-CR1-[CstTRD]-CR2) (28% aa sequence identity between Cst TRD and *Bsa*XI TRD2) could not be expressed in *E. coli*. We concluded from this experiment that highly homologous TRD (59% sequence identity) can be used to replace *Bsa*XI TRD2 and generate active enzyme with *Bsa*XI RM subunit. TRDs with low sequence identity to *Bsa*XI TRD2 may be problematic possibly due to incompatible TRD with *Bsa*XI CR1 and CR2.

The results of all *Bsa*XI TRD rearrangement and mutagenesis are summarized in **Table 1**.

### **BsaXI Methylase Activity**

The *Bsa*XI is predicted to carry both endonuclease and methylase activity. To detect *Bsa*XI methylase activity of the RM/S complex, phage  $\lambda$ -DNA was first incubated with *Bsa*XI in the presence of methyl donor SAM without Mg<sup>2+</sup> cations, and in the presence of 1 mM EDTA (restriction activity is inhibited by the absence

of divalent cations). After 2 h methylation, MgCl<sub>2</sub> (10 mM) was added to the reaction, and 2 U of *Bsa*XI was added to chase the cleavage reaction. After DNA methylation reaction, the DNA was partially resistant to *Bsa*XI digestion (data not shown). The substrate DNA was only partially modified, in agreement with the *in vivo* expression result that the co-expression of RM and S genes in *E. coli* host is toxic and unstable on the same plasmid due to insufficient methylation. It is not clear whether there is another MTase that modified the same sequence from the native host. We have not extensively analyzed the methylase activities of the newly engineered TRDs (TRD1-CR1, TRD1-CR1-TRD1-CR2, and CR1-TRD2-CR2) in complex with the WT RM subunits.

### **BsaXI TRD1 and TRD2 Homologs:** Standalone Target Recognition Domains From Sequenced Microbial Genomes

Only two homologs of BsaXI standalone TRD1 are found in GenBank (WP\_052564242 = 177 aa and KAA0244900 = 175 aa). Phyre2 search predicted that both TRD1 homologs lack the CR region (data are not shown). Therefore, either sequencing errors resulted in premature termination (missing CR region) or a separate peptide for the CR region or the TRD homologs may function in the absence of CR sequence. In the genome of *Candidatus Brocadia* sinica, two ORFs encoding TRD2 (240 aa) and RM (928 aa) are also present, suggesting they might encode an R-M system (CbrSI). It was noted that Cbr TRD1 and RM putative proteins share high sequence identity to the counterparts in BsaXI (39.3% and 55.9% aa sequence identity to BsaXI TRD2. Most BsaXI TRD1 homologs are present as fusion to other TRDs in the S subunits of putative Type I and IIB restriction systems.

BlastP search also identified more than 36 standalone *BsaXI* TRD2 homologs (196 to 248 aa long). Some homologs are shown in **Supplementary Figure 5**. Most of the *BsaXI* TRD2 homologs

are fused to *Bsa*XI TRD1 homologs or fused to other TRDs with low sequence identity, suggesting *Bsa*XI TRD2 homologs may have partnered with other TRDs to create new specificities.

### DISCUSSION

In this work, we cloned and expressed bsaXIRM and bsaXIS genes in E. coli. BsaXI enzyme was purified by multi-step chromatography from cell extracts containing RM and S subunits and mixed together before purification. We also purified RM and S subunits separately and reconstituted BsaXI activity in vitro by mixing the purified subunits. The reconstituted enzyme activity is lower than the enzyme purified as a natural complex. Similar to BcgI endonuclease complex, BsaXI enzyme also consisted of two RM subunits and one S subunit in the form of [RM]2 S (see Supplementary Figures 1C,D for schematic diagrams). The S subunit is analogous to Type I HsdS in domain organization of TRD1-CR1-TRD2-CR2. By rearrangement of TRDs and CRs, we examined the activity of TRD1-CR1, TRD1-CR2, 2xTRD1 (TRD1-CR1-TRD1-CR1) and C-terminal deletion variants in TRD1-CR1 and created a new cleavage specificity 5' AC N5 GT 3'. As expected, the circular permutation of TRDs in TRD2-CR2-TRD1-CR1 created an active enzyme with the same specificity as the WT when the S variant is complexed with RM subunits. The cleavage was not very precise only in the longer nick on the staggered cuts (5' CTCC N10-13/N7, 5' GT N12-13/N9), and the shorter nicks at N7 and N9 were not affected. The reason for this different effect is unknown. BsaXI TRD2 domain can be substituted by a homologous TRD protein (59% aa sequence identity) found in Cyanothece sp. CCY0110 bacterial genome and the chimeric S subunit can form active complex with BsaXI RM subunits. However, a distant TRD homolog with 28% sequence identity failed to express the chimeric S subunit, which may be due to the incompatible protein folding of CstTRD and BsaXI CR2 (steric conflict). More structure-guided S protein engineering is required to evaluate compatibility between TRDs and CRs. For example, the newly designed TRD-CR protein can be fused to *lacZ*<sup>\alpha</sup> peptide in an *in vivo* protein solubility assay to screen functional TRD-CR protein and its binding activity to cleavage-deficient RM subunit (Chevalier et al., 2002).

Our deletion analysis of TRD1-CR1 in the CR1 long  $\alpha$ -helix conserved region indicated that the full-length CR1 is not absolutely required for activity, although longer deletion (i.e., 32-aa deletion) had the most impaired activity. The *in vitro* activity of CR1 32-aa deletion is similar to that of TRD1-CR2 (the entire CR1 was replaced by CR2) with residual partial activity. It is suspected that the spacing distance N5 in the 5' AC N5 CTCC 3' bipartite sequence is mostly controlled by the horizontal length of the dumbbell shape of two TRDs (O = O, i.e., the distance of two dumbbells plus the handle) and it is not entirely determined by the length of the long  $\alpha$ -helix conserved regions CR1 and CR2.

There are currently 31 Type IIB restriction systems (REBASE) which cleave DNA sites with bipartite sequences.<sup>2</sup> Some of them might be amenable to the TRD deletion strategy described

here to create new specificities. It is also important to examine the functionality of BsaXI TRD2 specificity. In this work, we demonstrated that CR1-TRD2-CR2 can be expressed and purified. If CR1-TRD2-CR2/RM can form active complex and activate cleavage, the TRD variant may encode a rare cutter (GGAG Nx CTCC). An alternative enzyme discovery strategy is to use BsaXI RM subunit to pair with some BsaXI S homologs with low aa sequence identity (20-34% sequence identity) provided that the BsaXI RM subunit can form complex with these homologs to cleave DNA. This screening method may discover new REases with unique but related recognition sequences. In general, Type II REases with more than 35% aa sequence identity encode isoschizomers with the same specificity or with one base off recognition. Enzymes with 22-34% sequence identity sometimes encode REases with altered target sites (Wilson and Murray, 1991; Pingoud et al., 2014). More biochemical and structure studies of BsaXI and Type IIB S subunit are needed before we can engineer new enzyme specificities at will from the large number of TRDs found in Type I and IIB restriction systems.

### DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm. nih.gov/genbank/, OM373208.

## **AUTHOR CONTRIBUTIONS**

S-YX designed the experiments and wrote the manuscript. SG, DH, and S-YX performed the experiments. DH edited the manuscript. All authors contributed to the article and approved the submitted version.

## FUNDING

This work was supported by New England Biolabs, Inc.

## ACKNOWLEDGMENTS

We thank Andy Gardner and Geoff Wilson for helpful discussions and NEB DNA core lab for DNA sequencing. We are grateful to Tom Evans, Richard Roberts, and Jim Ellard for support. SG was supported by NEB M.S. student internship program. The publication cost was paid for by NEB. A part of this work had been presented at a restriction enzyme conference as a poster presentation.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2022. 888435/full#supplementary-material

<sup>&</sup>lt;sup>2</sup>http://hae.neb.com/cgi-bin/doublist

### REFERENCES

- Abadijieva, A., Patel, J., Webb, M., Zinkevich, V., Firman, K., and deletion, A. (1993). Mutant of the type IC restriction endonuclease EcoR124I expressing a novel DNA specificity. *Nucleic Acids Res.* 21, 4435–4443. doi: 10.1093/nar/21. 19.4435
- Chevalier, B. S., Kortemme, T., Chadsey, M. S., Baker, D., Monnat, R. J., and Stoddard, B. L. (2002). Design, activity, and structure of a highly specific artificial endonuclease. *Mol. Cell* 10, 895–905. doi: 10.1016/s1097-2765(02) 00690-1
- Fuller-Pace, F. V., Bullas, L. R., Delius, H., and Murray, N. E. (1984). Genetic recombination can generate altered restriction specificity. *Proc. Natl. Acad Sci.* U.S.A. 81, 6095–6099. doi: 10.1073/pnas.81.19.6095
- Guan, S., Blanchard, A., Zhang, P., and Zhu, Z. (2010). Alteration of sequence specificity of the type IIS restriction endonuclease BtsI. *PLoS One* 5:e11787. doi: 10.1371/journal.pone.0011787
- Gubler, M., and Bickle, T. A. (1991). Increased protein flexibility leads to promiscuous protein-DNA interactions in type IC restriction-modification systems. *EMBO J.* 10, 951–957. doi: 10.1002/j.1460-2075.1991.tb08029.x
- Heitman, J., and Model, P. (1990). Mutants of the EcoRI endonuclease with promiscuous substrate specificity implicate residues involved in substrate recognition. *EMBO J.* 9, 3369–3378. doi: 10.1002/j.1460-2075.1990.tb07538.x
- Jurenaite-Urbanaviciene, S., Serksnaite, J., Kriukiene, E., Giedriene, J., Venclovas, C., and Lubys, A. (2007). Generation of DNA cleavage specificities of type II restriction endonucleases by reassortment of target recognition domains. *Proc. Natl. Acad Sci. U.S.A* 104, 10358–10363. doi: 10.1073/pnas.0610365104
- Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858. doi: 10.1038/nprot.2015.053
- Kennaway, C. K., Taylor, J. E., Song, C. F., Potrzebowski, W., Nicholson, W., White, J. H., et al. (2012). Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Genes Dev.* 26, 92–104. doi: 10.1101/gad.179085.111
- Kong, H. (1998). Analyzing the functional organization of a novel restriction modification system, the BcgI system. J. Mol. Biol. 279, 823–832. doi: 10.1006/ jmbi.1998.1821
- Kong, H., Morgan, R. D., Maunus, R. E., and Schildkraut, I. (1993). A unique restriction endonuclease, bcgi, from bacillus coagulans. *Nucl. Acids Res.* 21, 987–991. doi: 10.1093/nar/21.4.987
- Lanio, T., Jeltsch, A., and Pingoud, A. (1998). Towards the design of rare cutting restriction endonucleases: using directed evolution to generate variants of EcoRV differing in their substrate specificity by two orders of magnitude. *J. Mol. Biol.* 283, 59–69. doi: 10.1006/jmbi.1998.2088
- Liu, Y. P., Tang, Q., Zhang, J. Z., Tian, L. F., Gao, P., and Yan, X. X. (2017). Structural basis underlying complex assembly and conformational transition of the type I R-M system. *Proc. Natl. Acad. Sci. U.S.A.* 114:11151. doi: 10.1073/ pnas.1711754114
- Loenen, W. A., Dryden, D. T., Raleigh, E. A., and Wilson, G. G. (2014). Type I restriction enzymes and their relatives. *Nucleic Acids Res.* 42, 20–44. doi: 10.1093/nar/gkt847
- Lunnen, K. D., Barsomian, J. M., Camp, R. R., Card, C. O., Chen, S. Z., Croft, R., et al. (1988). Cloning Type II restriction and modification genes. *Gene* 74, 25–32.
- Morgan, R. D., and Luyten, Y. A. (2009). Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity. *Nucleic Acids Res.* 37, 5222–5233. doi: 10.1093/nar/gkp535
- Murray, N. E. (2000). Type I restriction systems: sophisticated molecular machines. *Microbiol. Mol. Biol. Rev.* 64, 412–434. doi: 10.1128/MMBR.64.2.412-434.2000
- Nagaraja, V., Shepherd, J. C. W., and Bickle, T. A. (1985). A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* 316, 371–372. doi: 10.1038/316371a0

- Pingoud, A., Wilson, G. G., and Wende, W. (2014). Type II restriction endonucleases-a historical perspective and more. *Nucleic Acids Res.* 42, 7489-
- 7527. doi: 10.1093/nar/gku447 Pingoud, A., Wilson, G. G., and Wende, W. (2016). Type II restriction endonucleases –a historical perspective and more. *Nucleic Acids Res.* 44:8011. doi: 10.1093/nar/gkw513
- Roberts, R. J., Belfort, M., Bestor, T., Bhagwat, A. S., Bickle, T. A., Bitinaite, J., et al. (2003). A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812. doi: 10.1093/nar/gkg274
- Roberts, R. J., Vincze, T., Posfai, J., and Macelis, D. (2015). REBASE–a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* 43, D298–D299.
- Samuelson, J. C., Morgan, R. D., Benner, J. S., Claus, T. E., Packard, S. L., and Xu, S. Y. (2006). Engineering a rare-cutting restriction enzyme: genetic screening and selection of NotI variants. *Nucleic Acids Res.* 34, 796–805. doi: 10.1093/nar/ gkj483
- Samuelson, J. C., and Xu, S. Y. (2002). Directed evolution of restriction endonuclease BstYI to achieve increased substrate specificity. J. Mol. Biol. 319, 673–683. doi: 10.1016/S0022-2836(02)00343-1
- Vasu, K., Nagamalleswari, E., Zahran, M., Imhof, P., Xu, S. Y., Zhu, Z., et al. (2013). Increasing cleavage specificity and activity of restriction endonuclease KpnI. *Nucleic Acids Res.* 41, 9812–9824. doi: 10.1093/nar/gk t734
- Vincze, T., Posfai, J., and Roberts, R. J. (2003). NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res.* 31, 3688–3691. doi: 10.1093/ nar/gkg526
- Wilson, G. G., and Murray, N. E. (1991). Restriction and modification systems. Ann. Rev. Genet. 25, 585-627.
- Xu, M. Q., and Evans, T. C. Jr. (2003). Purification of recombinant proteins from E. coli by engineered inteins. *Methods Mol. Biol.* 205, 43–68. doi: 10.1385/1-59259-301-1:43
- Youell, J., and Firman, K. (2008). EcoR124I: from plasmid-encoded restrictionmodification system to nanodevice. *Microbiol. Mol. Biol. Rev.* 72, 365–377. doi: 10.1128/MMBR.00043-07
- Zhu, Z., Zhou, J., Friedman, A. M., and Xu, S. Y. (2003). Isolation of BsoBI restriction endonuclease variants with altered substrate specificity. *J. Mol. Biol.* 330, 359–372. doi: 10.1016/s0022-2836(03)0 0595-3

**Conflict of Interest:** DH and S-YX are employees of New England Biolabs, Inc., a company that develops restriction enzymes and other reagents for research and diagnostic applications.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Gidwani, Heiter and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.