SUPPLEMENTARY MATERIALS

for Uemura et al. "Crystal structure of the ligand binding form of nanoRNase from *Bacteroides fragilis*, a member of the DHH/DHHA1 phosphoesterase family of proteins"

Supplementary Methods

S1. Protein preparation

The amplified fragment of the bfNrn gene (ORF ID: BF3670) was engineered into pET-21d and the resulting expression construct was used to transform E. coli RosettaTM2(DE3)pLysS cells (Merck, Darmstadt, Germany). Cells harboring the expression construct were cultured at 37°C in 1.5 1 LB medium containing 50 µg /ml ampicillin to a density of 1 \times 10⁸ cells/ml before addition of isopropyl β -D-1-thiogalactopyranoside (50 µg/ml). The culture was then continued for a further 6 h before harvesting the cells by centrifugation. The cell pellet was stored at -20°C until required. The following procedures were carried out at 4°C except for size exclusion chromatography. Frozen cells (8 g) were thawed, suspended in 80 ml of 50 mM Tris-HCl (pH 8. 0), 500 mM NaCl, 1 mM phenylmethylsulfonyl fluoride and 5 mM imidazole, and disrupted by sonication on ice. The lysate was then clarified by centrifugation (38,000 g for 60 min). The resultant supernatant was loaded onto a His-Bind resin (Merck) column (bed volume, 10 ml) equilibrated with 20 mM Tris-HCl (pH 8. 0), 500 mM NaCl and 5 mM imidazole. Bound proteins were eluted with a linear gradient of 0.05–0.5 M imidazole (total volume, 200 ml). Fractions containing the target protein were pooled and concentrated with a Vivaspin (10,000 molecular weight cutoff) concentrator. The concentrated protein solution was then subjected to size exclusion chromatography using a Superdex 75 HR 10/30 column (GE Healthcare Biosciences) equilibrated with 20 mM Tris-HCl (pH 7.5), 100 mM NaCl. Chromatography was carried out on an ÄKTA explorer system (GE Healthcare Biosciences) at 20°C. Fractions containing the target protein (identified by SDS-PAGE) were concentrated and stored at 4°C.

1

The concentration of the purified protein was determined using the molar absorption coefficient at 280 nm calculated with the Protpram program of Expert Protein Analysis System (ExPASy) proteomics server (http://www.expasy.org/). Approximately 3.5 mg bfNrn was obtained from 8 g of cells.

S2. Exonuclease activity assay

The assay was performed as described previously [1]. Briefly, the reaction mixture (10 μ l) contained 50 mM HEPES-KOH (pH 7.5), 100 mM KCl, 5 mM MnCl₂ (or 5 mM EDTA), 10 nM 5'-³²P-labeled 11-mer ssDNA and 1 μ M bfNrn. The assay was carried out using various substrates of different length i.e., 10 nM 5'-³²P-labeled ssDNA (3, 6, or 11-mer), 10 μ M cold ssDNA (3, 6, or 11-mer) and 500 nM bfNrn in 50 mM HEPES-KOH, pH 7.5, 100 mM KCl and 5 mM MnCl₂. The mixture was incubated at 37°C before stopping the reaction by addition of 1 μ l of 100 mM EDTA and 11 μ l of phenol/chloroform. Analysis of the degradation product by mass spectrometry was performed as described previously [2].

S3. Phosphatase activity assay

The reaction mixture (100 μ l) contained 50 mM Tris-HCl (pH 7.5), 100 mM KCl, 5 mM MnCl₂, 25 μ M adenosine 3',5'-bisphosphate (pGp), adenosine 2',5'-bisphosphate, 3'-AMP, or c-di-GMP and 500 nM bfNrn. Reactions were incubated at 37°C for 6 h and then stopped by addition of 20 μ l of 100 mM EDTA. An equal volume (120 μ l) of buffer I (50 mM Tris-HCl (pH 7.5), 5 mM tetra-*n*-butylammonium hydroxide and 10% methanol) was added, and the protein was removed by ultrafiltration using a membrane filter (10,000 molecular weight cutoff). A 200- μ l aliquot of the filtrate was applied to a reversed-phase column (CAPCELL PAK C18 (Shiseido)) equilibrated with buffer I. Bound material was subsequently eluted using a gradient of 10 to 50% methanol. To compare the phosphatase activity, the inorganic phosphate was measured using a

colorimetric assay [3].

Supplementary Tables

Dataset	bfNrn-Mn ²⁺ -GMP complex				
Wavelength (Å)	1.000	1.892			
Resolution (Å)	50-2.95 (3.00-2.95)	50-3.20 (3.26-3.20)			
Space group	P3 ₂ 21	<i>P</i> 3 ₂ 21			
Cell constants (Å) a, b, c	90.8, 90.8, 107.9	90.9, 90.9, 108.0			
Cell angles (°) α , β , γ	90, 90, 120	90, 90, 120			
Data collection					
Observed reflections	205804	91960			
Unique reflections	11209 (556)	8823 (438)			
Completeness (%)	99.8 (100)	99.5 (100)			
Redundancy	21.9 (22.3)	10.4 (10.9)			
$I/\sigma(I)$	48.6 (12.8)	42.3 (10.4)			
$R_{ m merge} \left(\%\right)^a$	9.2 (30.2)	10.3 (29.9)			
Refinement ^d					
<i>R</i> -factor $(\%)^b$	22.09				
$R_{\rm free}$ (%) ^c	26.08				
Average B-factor ($Å^2$)	61.2				
Stereochemistry r.m.s d ^e					
Bonds lengths (Å)	0.007				
Bond angles (°)	1.313				

Supplementary Table S1. Data collection and refinement statistics.

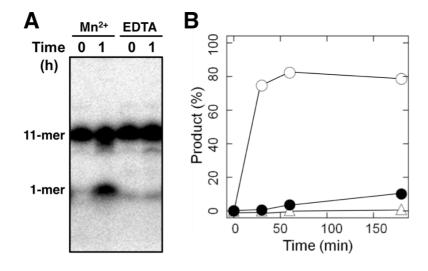
^a Values in parentheses are for the outermost shell.

 ${}^{b}R_{\text{merge}} = \sum_{hkl} \sum_{i} |I_i(hkl) - I(hkl)| / \sum_{hkl} \sum_{i} I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and I(hkl) is the averaged intensity for multiple measurements.

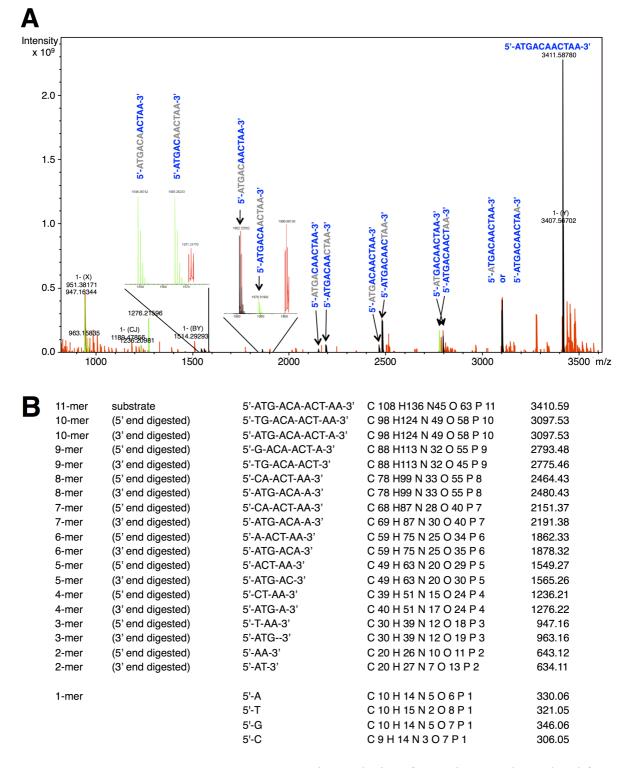
^{*c*} *R*-factor = Σ ||F_{obs}| - |F_{calc}||/ Σ |F_{obs}|

 $^{d}R_{\text{free}}$ is monitored with 5% of the reflection data excluded from refinement.

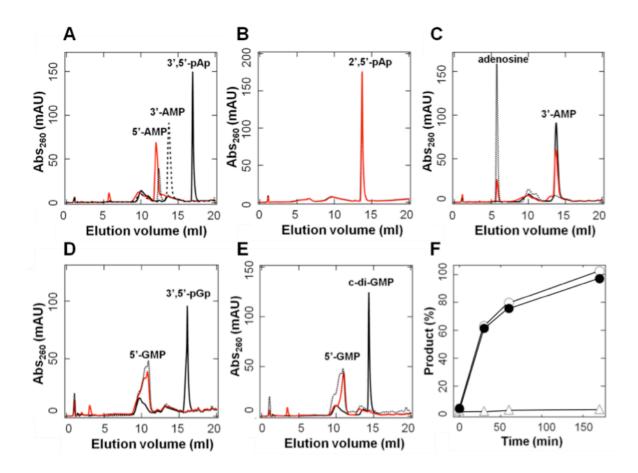
^e Over 90% of main-chain dihedrals fall within the 'most favored regions' of the Ramachandran plot.



Supplementary Fig. S1. Exonuclease activity of bfNrn. (A) Dependence of enzyme activity on Mn^{2+} . An aliquot of 10 nM 5'-³²P-labeled 11-mer oligodeoxyribonucleotide was reacted with 1 μ M bfNrn at 25°C. The reaction mixtures contained 50 mM HEPES (pH 7.5), 100 mM KCl and 5 mM MnCl₂ or 50 mM EDTA. (B) Dependence on substrate length. Reactions were performed as described in (A) except for a lower concentration of bfNrn (0.5 μ M) and the addition of 10 μ M unlabeled oligodeoxyribonucleotide. Circles, 3-mer; filled circles, 6-mer; triangles, 11-mer.



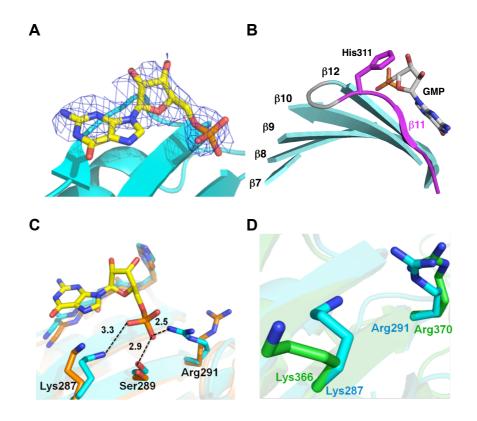
Supplementary Fig. S2. Mass spectrometric analysis of reaction products by bfNrn. Reactions were performed as described in Fig. S1A except for a longer reaction time (11 h). Other procedures were perfomed as described [2]. (A) The deconvoluted spectrum of reacted products of 5'-phosphorylated 11-mer ssDNA. For clarity, information concerning the major species are shown above the corresponding peaks. (B) Theoretical masses of each DNA fragment. These were calculated assuming that bfNrn hydrolyzes a phosphodiester bond at 5'-side or 3'-side of a phosphate.



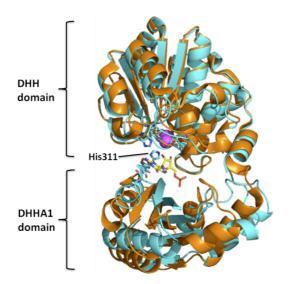
Supplementary Fig. S3. Phosphatase activity of bfNrn. (A–D) HPLC profile of the products of bfNrn phosphatase activity on 3',5'-pAp (A), 2',5'-pAp (B), 3'-AMP (C), 3',5'-pGp (D) and c-di-GMP (E) substrate. Elution profiles of the product of hydrolysis by bfNrn (red lines) and respective control mixtures (black lines) are shown. An aliquot of 25 μ M substrate was reacted with 0.5 μ M bfNrn at 37°C for 6 h. The reaction mixtures contained 50 mM Tris-HCl, pH 7.5, 100 mM KCl and 5 mM MnCl₂. (F) Time course of reactions mediated by bfNrn. Circles, pAp; filled circles, pGp; triangles, 3'-AMP.

		α1							13	α5	R
	10	20		30 4	10 50	pz	60	70	80		90
Nrn	(O)MLTKVIAQAHI			S <mark>H</mark> VSPDGDA I GSS	SLGLYHFLDSQC		VPNAFPDFLK	WMPGSKD11	LLYDRYQEF	4	
Nrn	(0)MDGNAPEPRYWEKMRLV-		GP I Y I A T	F <mark>H</mark> VDPDGDA I GSS	GLGLYRALKAL-	-GKEA-Y	WVADPPRFLF	FLPKEEE-	YSDP	/	-EKL-PPGATLV
NrnA	(0)————MK—		dt <mark>i i lh</mark> f	R <mark>H</mark> VRPDPDAYGSG	CGLYRALKAL-	-GKEA-Y	WVADPPRFLF	FLPKEEE	YSDP	/	-EKL-PPGALVI
Nrn	(0)MVEIF		et <mark>i i i he</mark>	R <mark>H</mark> VRPDPDAYGSC	LGLKLYLERKF	PEKNI-Y	ATGEAEPSLS	FIGDL	DE I)	-DSV-YSDALVI
Nrn	(0)MNSQ	——AEAVSATQAH	DKIVIFF	H <mark>HIRPDGDCLGAG</mark>	QHGLARL I QTNF	PHKQV-F	CVGDPKHNFF	WLEMV	-FTPKEQ	г——	
Nrn	(0)MTTIDPRSELVDGRRRAGA	RVDAVGAAALLSAA	AFVGVVC	C <mark>H</mark> VHPDADT I GAG	GLALALVLDGCO	KRVEVSFA	APATLPESLF	SLPGCHLL'	V F	3	
Nrn	(0)MTAF		dt <mark>i i i h</mark> f	R <mark>H</mark> MKPDPDALGS©	QVGLKEMITSNE	PQKTV-K	VTGYNEPSLS	WLAQM	DDV	3	
RecJ	(48)——PLALLPLKGL	REAAALLEEALRQG	KRIBVHG	G <mark>D</mark> YDADGLTGTA I	LVRGLAALG	-ADVHPF	IPHRLEEGY	VLMER	\	/	
RecJ	(44) VKGMLPWQQLSGV	/EKAVETLYNAFREG	TRIIVVO	G <mark>D</mark> FDADGATSTAL		-CSNIDYL	VPNRFEDGY	LSPEV		/	-DQAHARGAQL I
RecJ	(53)DADFYDPFEMKGM	(KEAADR KQA SQQ	ЕК <mark>ТМТҮ</mark> Ө	G <mark>D</mark> YDADGVTSTSV	/MLHTLQKL-	-SAQVDFY	IPDRFKEGY	PNEQA	F	-	
YybT (304) VKFYGGKTNPMEKRTRVRA	RVISHALKEIVTES	SNVIIM	G <mark>H</mark> KFPDMDS I GAA	IGILKVAQANN	KDGF I V-II	DPNQIGSSVG	RLIGE	— I KKYEEL	WSRF I TPE	EAMEISNDDTLLV
			: :	ĎHH mo	41F I					m	otif İl
					α8			0			10 -
		Bo		- Bo			_	α.9			10
	100 110	120 13	_	140	150		160		170		
Nrn	CLDFNALK <mark>R</mark> IDEMSDIVA#			-ITISHPEISST	ISELVFRL	I CRMG	Y——FS	DI-	—SKEGAE <mark>C</mark> I	I YTGMM-	TC
Nrn	-LDSAEPS <mark>R</mark>	PVEGEVINIDHHGTN	PRFGH-	-LHVVDPSKAAT	FAQMVKDL	IDLLG	۷	EW	-TAETATP	/LTGIL—	TC
NrnA	-CDTANQE <mark>R</mark> I DDQR <mark>Y</mark> PSGA	∖K <mark>LMKIDHHPNE</mark>	DPYGD-	-LLWVDTSASSV	/SEMIYEL	YLEGKI	EHGW	⊢−−−KL	NTKAAELI	IYAGIV—	GC
۱rn	-CDTANAP <mark>B</mark> IDDQR <mark>YLNG</mark> G	}——SLIKIDHHPAT	d <mark>qygd</mark> —	-VNFVNTEASST	ISET I FDF	ISHEN	DLS)II-	—DEHVAR <mark>V</mark> L	_YLGIV—	G[
Nrn	-VDANYKE <mark>R</mark> TECRD_LLDQN	AQFKAVLRIDHHPNE	d <mark>dlnt</mark> —	-THNFVDASYIAA	AEQVVDL	AVQA	KW	⊢−−KL	—SPPAAT <mark>a</mark> l	_YLGIY—	TC
Nrn	TVD1PSVD <mark>B</mark> LGALGPLTD-	SGRELLVIDHHASN	DLFGT—	-ANFIDPSADST	TTMVAE I	——LDA₩G	К	——P1	-DPRVAHCI	I YAGLA	TC
Nrn	-VDTANRP <mark>B</mark> IDDQR <mark>YLNG</mark> N	I−−−F <mark>LIKIDHHPDE</mark>	d <mark>h</mark> ygd—	-LSYVDTKASSA	ASETTOF	ALQN	QL	KL	—SDQAAR <mark>l</mark> l	_YAGIL-	GE
RecJ	TVDCGITN <mark>H</mark> AELRELLE	NGVEV I VTDHHTPG	K <mark>TPPP</mark> —	-GL-VVHPALTPD)LKEKPTGAG	GVAFLLLWAI	LH	ERLGLI	PPPLEYADL/	AAVGTIAD	VAPLWGWNRALVH
RecJ	TVDNG I SS <mark>H</mark> AGVEHARS—	LGIPVIVTDHHLPG	DTLPA—	-AEAT INPNLRDO	NFPSKSLAGVG	GVAFYLMLA	LRTFLRDQGW	FDERNIAII	PNLAELLO	ALGTVAD	VVPLDANNRILT
RecJ	TVDTGI AA <mark>V</mark> HEAKVAKE-	LGLDVI I TDHHEPG	PFI PD-	-VRATVHPKOPGC	TYPEKELAGVO	GVAFKLAHAI	LLGEL			AAIGTIAD	LVPLHDENRLIA
YybT	-VDTHKPS <mark>L</mark> VMEER_VN	KIEH I VV IDHHRRG				LEYQPI	K	KI	NMTEATAL	LAGII-	VC
ҮуЬТ	-VDTHKPS <mark>L</mark> VMEER_VN	: ***				LEYQPI	KRL	KI		*	VD
YybT	VDTHKPS <mark>L</mark> VMEER_VN *	KIEHUVIDHHRRG				LEYQPI	K	KI		*	tif IV
ҮуЬТ	* α11	: *** motif III α12		PLLVYMEPYASST α13	FAELVTEL	14 <mark>- β8</mark>	-	KI	—NMTEAT <mark>AL</mark> α15	*	tif IV
ҮуЬТ	*	: ***		PLLVYMEPYASST	TAELVTEL	14 <mark>- β8</mark>	K	KI-		*	tif IV
YybT Nrn	* α11	i **** motif III α12 200 210		PLLVYMEPYASST α13 220	TAELVTEL 	<mark>14 = β8</mark> : DYNSA	-			*	tif IV 250 26
-	* α11 18090	i *** motif III α12 200 210 ELLSKGIDKDDIYRK	UYNTYSE	PLLVYMEPYASST α13 220 SRLRLMGYVLS-	FAELVTEL - β7α1 230	<mark>14 = β8</mark> : DYNSA	240			*	tif IV 250 26 —GKFDY I KGDSE
Nrn Nrn	× α11 180 190 GGFTYNSN-NRETYFTISE	intering to the second	VYNTYSE	220 SRLPLMGYVLS- SYFRLMGQVLS-	TAELVTEL 	14 <mark>- β8</mark> : DYNSA GGLLV	240 —LISLTKEE	Q		*	tif IV 250 2e —GKFDY I KGDSE —EDAGAEEDSI
Nrn Nrn NrnA	* α11 180 190 GGFTYNSN-NPELYFTISE GNFPFANT-TPEVLRVAAE	i **** motif III 200 210 2LLSKGIDKDDIYRK 2LGYGVKLAELTDR 2LIQYPFSSSELFNQ	VYNTYSE LQFR-PF LYET-KL	220 220 SRLRLMGYVLS- SYFRLMGQVLS- NVVKLNGFTFQ-	TAELVTEL 	14 - β8 DYNSA GGLLV VG-AA	240 —LISLTKEE —TAHLP—	Q		*	tif IV 250 26 —GKFDY I KGDSE —EDAGAEEDSI —EKFGTTASEAS
Vrn Vrn VrnA Vrn	* 180 190 GGFTYNSN-NPELYFTISE GNFRFANT-TPEVLRVAAE GRFLFPNT-TEKTLKYAGE	i **** motif III 200 210 2LLSKG IDKDD I YRK 2LGYGVKLAEL TDR 2LIQYPFSSSELFNQ XLLAYPFNHAELNK	VYNTYSE LQFR-PF LYET-KL MSEK-DF	α13 220 SRLRLMGYVLS- SYFRLMGQVLS- NVVKLNGFTFQ- KLMPFQGYVLQ-	TAELVTEL 230 MKVYKE TVAFHFG NVSLSEN	14 <mark>β8</mark> DYNSA GGLLV VG-AA SHEYC	240 —LISLTKEE —TAHLP— —SVFIKKDT	Q L		*	tif IV 250 26 —GKFDY I KGDSE —EDAGAEEDSE —EKFGTTASEAS —KQFD I QPNEAS
Irn Irn IrnA Irn	* 180 190 GGFTYNSN-NPELYFTISE GNFRFANT-TPEVLRVAAE GRFLFPNT-TEKTLKYAGE GRFLFSNT-SPHTMEVASC	in the second s	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL	α13 220 SRLPLMGYVLS- SYFRLMGYVLS- NVKLNGFTFQ- KLMPFQGYVLQ- KDTQFKQYVFK-	TAELVTEL 230 NMKVYKE TVAFHFC NVSLSEN NFELSDS	14 - β8 DYNSA	240 —LISLTKEE —TAHLP— —SVFIKKDT —QIKITNDV	Q L L Q		*	tif IV 250 26 —GKFDY I KGDSE —EDAGAEEDSE —EKFGTTASEAS —KQFD I QPNEAS —KKLKVTPLEC
Irn Irn IrnA Irn Irn	* 180 190 GGFTYNSN-NPELYFTISE GNFPFANT-TPEVLRVAAE GRFLFSNT-SPHTMEVASC NRFLSNT-SWRTLYLGSN	in the second s	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF	220 SRLPLMGYVLS- SRLPLMGYVLS- NVVKLNGF IFQ- XLMPFQGYVLQ- LKD IQFKQYVFK- TVLPLLSRVLG-	TAELVTEL 230 NMKVYKC TVAFHFC NVSLSEN NFELSDS NFQTFQN	14 - β8 DYNSA GGLLV NG-AA SHEYC SHEYC NV I Y EAVGGRG	240 —LISLTKEE —TAHLP —SVFIKKDT —QIKITNDV —FVADKKF	Q L Q R		*	tif IV 250 26 —GKFDY I KGDSI —EDAGAEEDSI —EKFGTTASEAS —KQFD I QPNEAS —KQFD I QPNEAS —KULKVTPLEC —EWVAARSEEVE
Nrn NrnA NrnA Nrn Nrn Nrn Nrn	* GGFTYNSN-NPETYFTTSE GNFPFANT-TPEVLFVAAE GRFLFPNT-TEKTLKYAGE GRFLFSNT-SPHTMEVASG NRFLYSNT-SWRTLYLGSN GSFFWASYRGYRLAAF	motif III 200 210 2LLSKG IDKDD I YRK 2LGYGYKLAELTRR 2LQYPFNHNAELNK MUYRAQANIAK IHDE 8LVE I GVDNATVSRT 2LLKYDFDFAALARQ	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-P1	220 SRLPLMGYVLS- SRLPLMGYVLS- NVVKLNGF IFQ- XLMPFQGYVLQ- LKD IQFKQYVFK- TVLPLLSRVLG-	AELVTEL 230 NMKVYKU VAFHFG NVSLSEN NFQTFQN SAQLVSE NLEIDKN	14 - β8 DYNSA GGLLV NG-AA SHEYC SHEYC NV I Y EAVGGRG NG-AA	240 —LISLTKEE —TAHLP— —SVFIKKDT —QIKITNDV —FVADKKF —LVYVVVDN —RIILSQKI	Q L Q R L	α15	* mo	tif IV 250 22 EDAGAEEDSI EKFGTTASEAS KOFD10PNEAS KKLKVTPLEC EWVAARSEEVE KKLNTPLEC
Nrn Nrn NrnA Nrn Nrn Nrn RecJ	* 180 190 GGF TYNSN-NRE IYF I ISE GNF FANT-TPEVLRVAAE GRF LFPNT-TEKTLKVAGE GRF LSNT-SWRTLVLGSN GSFRWASV-RGYRLAAF GRFLYPAT-TSKTF I IASE	motif III 200 210 2LLSKGIDKDD I YRK 2LGYGVKLAELTDR 2LQYPFSSSELFNQ XLAYPFNHAELNK MLYRAQAN IAK I HDE 8LVE I GVDNATVSRT 2LLKYDFDFAALARQ AVGY — TGKAVEV	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFR1-/	2LLVYMEPYASST 220 SSRLPLMGYVLS- SYFFILMGQVLS- NVVKLNGFTFQ- KLMPFGGYVLQ- KD1QFKQVVFK- TWLPLLSRVLG- KLAKLQAYVFE- PRTINAASRLGEA	IAELVTEL 230 NMKVYKC NVSLSEN NFCTFGN SAQLVSE NLEIDKN NELDKN NELLIDKN	14 - 68 CYNSA	240 —LISLTKEE —TAHLP— —SVFIKKDT —QIKITNDV —FVADKKF —LVYVVVDN —RIILSQKI GELHFLNARF	Q Q R QTLEEAMLI	α 15	* mo	tif IV 250 21 —GKFDY I KGDSI —EKFGTTASEA: —KQFD I QPNEA: —KKL KVTPLEC —EWVAARSEEVI —KKFNL TDAET: AKA I VLLDPEGHI
Nrn Nrn NrnA Nrn Nrn Nrn RecJ RecJ	* 180 190 GGF TYNSN-NFE I YF I I SE GNFF LFNT-TEKTLKYAGE GRF LFNT-TEKTLKYAGE GRF LFNT-SWRTLYGSN GSFRWASV-RGYFLAAF GRF LYPAT-TSKTF I I ASE GLAR I PAS-SWYGLRLLAE GLAR I PAS-SWYGLRLLAE	motif III 200 210 2LSKG IDKDD I YRK 2LGYGVKLAELTDR 2LQYPFSSSELFNQ 2LLAYPFNHAELNK 4LYRAQANIAK I HOE 2LVE I GVDNATVSRT 2LLKYDFDFAALARQ 2AVGY — TGKAVEV 2VANRDAQKLAASDL	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFR1-/ GFAL-(¢13 220 SRLPLINGYVLS- 2STFILINGQVLS- 2STFILINGQVLS- 2STFILINGQTVC- KDIQFKQYVFK- KUMPFQGYVLG- KLIQFKQYVFK- TWLPLLSRVLG- KIAKLQAYVFE- 1PRINAASRLGEA 3PRLNAAGRLDDM	TAELVTEL 230 NMKVYKE TVAFHE NFELSDS NFELS	14 8 88 SYNSA GGLLV NG-AA SHEYC VI Y EAVGGRG NG-AA VG-AA VI GEARVLAI DAAEAQALVU NI GEARVLAI	240 —LISLTKEE —TAHLP —SVFIKKDT —QIKITNDV —FVADKKF —LVYVVVDN —RIILSQKI GELIFILNARF NELDALNQTF	Q L Q R QTLEEAMLI KE I EQGMQ	α15 RKLLP	* mo	tif IV 250 21 —GKFDY I KGDSI —EKFGTTASEA: —KQFD I QPNEA: —KKLKVTPLEC: —EWVAARSEEVI —KKFNL TDAET: AKA I VLLDPEGHI PGGLAMYHPEWHI
Irn IrnA IrnA Irn Irn Irn Irn RecJ RecJ RecJ	* 180 190 GGF TYNSN-NFE IYF I ISE GNF FFANT-TPEVLFNAAE GRF LFSNT-SPHTMEVASG NFF LYSNT-SWRTT VLGSN GSF WASVRGYFLAAF GRF LYPAT-TSKTF I IASE GLAR IPAS-SWYGLRLLAE GMSR IRAGKCPPG IKALLE	motif III 200 210 21.5KG IDKDD I YRK 21.GYRVLAELTDR 21.GYPFSSELFNQ 21.LYPFOHNAELNK UYAQANIAK IHDE 2.VE IGVDNATVSRT 21.LKYDFDFAALARQ 24.VGY — TGKAVEV VANRDAQKLAASDL 3.SGGD I GEANEETV	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFR1-/ GFAL-(GFQL-/	α13 220 SRLPLINGYVLS- SYFFLINGQVLS- NVKLINGF IFQ- KLIMPFQGYVLS- KLIMPFQGYVLS- TWILDLSRVLG- TWILPLLSRVLG- YKI NASRLGEA SPRLNAAGRLGEA SPRLNAAGRLDDM	TAELVTEL 230 NMKVYKE TVAFHE NFELSDS NFELS	14 - β8 GGLL V VG-AA SHEYC VI Y VI Y VG-AA VI GARVLAI DAAEAQAL VI VI GEARVLAI DSFEAEELA	240 —LISLTKEE —TAHLP —SVFIKKDT —QIKITNDV —FVADKKF —LVYVVVDN —RIILSQKI GELIFILNARF NELDALNQTF	Q L Q R L QTLEEAMLI KE I EQGMQ XKMVSKMTI	α15 RKLLP	* mo	tif IV 250 24 GKFDY I KGDSI EDAGAEEDSI EKGFD I QPNEA: KKFN TDAET3 AKA I VLLDPEGHI PGGLAMYHPEWHG QTA I VVAKAGWNI
Vrn VrnA VrnA Vrn Vrn Vrn RecJ RecJ RecJ	* at 12 at 25 at 25	motif III 200 210 21.5KG IDKDD I YRK 21.GYRVLAELTDR 21.GYPFSSELFNQ 21.LYPFOHNAELNK UYAQANIAK IHDE 2.VE IGVDNATVSRT 21.LKYDFDFAALARQ 24.VGY — TGKAVEV VANRDAQKLAASDL 3.SGGD I GEANEETV	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFR1-/ GFAL-(GFQL-/	α13 220 SRLPLINGYVLS- SYFFLINGQVLS- NVKLINGF IFQ- KLIMPFQGYVLS- KUMPFQGYVLS- TWILDLSRVLG- TWILPLLSRVLG- YKI NASRLGEA SPRLNAAGRLGEA SPRLNAAGRLDDM	TAELVTEL - β7 α.1 230 - NMKVYKL - TVAFHFC - NVSLSEM - NFELSDS - NFEL	14 - β8 GGLL V VG-AA SHEYC VI Y VI Y VG-AA VI GARVLAI DAAEAQAL VI VI GEARVLAI DSFEAEELA	240 — LISLTKEE — TAHLP— — SVFIKKDT — QIKITNDV — RIILSQKI GELIFILNARF NELDALNQTF AE IDQLNKEF	Q L Q R L QTLEEAMLI KE I EQGMQ XKMVSKMTI	α15 RKLLP	* mo	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Irn IrnA IrnA Irn Irn Irn Irn RecJ RecJ RecJ	* 180 190 GGF TYNSN-NFE IYF I ISE GNF RFANT-TPEVLRVAAE GRF LFSNT-SPHTMEVASG NFF LYSNT-SWRTLYLGSN GSF WASY-RSYRLAAF GRF LYPAT-TSKTF I IASE GLAR I PAS-SWYGLRLLAE GLEFLRRT-NFLGLKEL I M	motif III 200 210 21.5KG IDKDD I YRK 21.GYRVLAELTDR 21.GYPFSSELFNQ 21.LYPFOHNAELNK UYAQANIAK IHDE 2.VE IGVDNATVSRT 21.LKYDFDFAALARQ 24.VGY — TGKAVEV VANRDAQKLAASDL 3.SGGD I GEANEETV	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFR1-2 GFAL-4 GFAL-4 GFAL-4	α13 220 SRLPLINGYVLS- SYFFLINGQVLS- NVKLINGF IFQ- KLIMPFQGYVLS- KUMPFQGYVLS- TWILDLSRVLG- TWILPLLSRVLG- YKI NASRLGEA SPRLNAAGRLGEA SPRLNAAGRLDDM	TAELVTEL - β7 α.1 230 - NMKVYKL - TVAFHFC - NVSLSEM - NFELSDS - NFEL	14 - β8 GGLL V VG-AA SHEYC VI Y VI Y VG-AA VI GARVLAI DAAEAQAL VI VI GEARVLAI DSFEAEELA	240 — LISLTKEE — TAHLP— — SVFIKKDT — QIKITNDV — RIILSQKI GELIFILNARF NELDALNQTF AE IDQLNKEF	Q Q R QTLEEAMLI KE I EQGMQ QKMVSKMTI	α15 RKLLP	* mo	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Nrn NrnA NrnA Nrn Nrn Nrn RecJ RecJ RecJ	* at 12 at 25 at 25	CONTRACTOR CONTRACTON CONTRACTON CONTRACTON CONTRACTON CONTRACTON CONT	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSH-PY AFR1-/ GFAL(GFQL-/ LKET-VI	α13 220 SRLRLMGYVLS- SSTLRLMGVLS- SYTRLMGQVLS- NVKLNGF IFQ- KLMPQGYVLQ- LKD IQFKQYVFK- TWLPLLSRVLG- SPRLNAAGRLDDM APRLNAVGR IEQA SYTIKRA-KL IQ-	TAELVTEL - β7 α.1 230 - NMKVYKL - TVAFHFC - NVSLSEM - NFELSDS - NFEL	14 β8 DYNSA	240 — L ISL TKEE — TAHLP — SVF IKKOT — Q IK I TNOV — FVADKKF — LYYVVVDN — RI ILSQKI GELHFNLNFF REL DALNQTF AE IDQLNKEF — ASLPENE —	Q Q R QTLEEAMLI KE I EQGMQ XXKWVSKMTI E	α15 RKLLP	* mo	tif IV 250 26 GKFDY I KGDSE EKGGTASEAS - KQFD I QPNEAS - KKLKVTPLEC- EWVAARSEEVE - KKFNL TDAETS AKA I VLLDPEGHF PGGLAMYHPEWHG QTA I VVAKAGWNF
Nrn NrnA Nrn Nrn Nrn Nrn RecJ RecJ YybT	* attackk attackk attackk attackk attackk attackk attackk a	200 210 200 200 200	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSF-PF MDSF-PY AFR1-/ GFAL-(GFAL-(GFAL-(LKET-VI)	α13 220 SRLPUMGVLS- SSTFRLMGVLS- SYFRLMGVLS- NVVKLNGFTFQ- KLINFEGKVLG- KLINFEGKVLG- KLINFLGLSRVLG- TWLPLLSRVLG- SPRLNAAGRLDDW NPRLNAAGRLDDW SYLKRA-KLIQA SYLKRA-KLIQA 300	TAELVTEL 230 NMKVYKU TVAFHFG NVSLSE NFQTFQN SAQLVSE NFQTFQN SAQLVSE NLEIDKN NEKALRLLIDE NDPAVHLLMSEE HTVLYKE 611 310	14 β8 DYNSA	240 — L ISL TKEE — TAHLP — SVF IKKOT — Q IK I TNOV — FVADKKF — LYYVVVDN — RI ILSQKI GELHFNLNFF REL DALNQTF AE IDQLNKEF — ASLPENE —	Q L Q R L L L KE I EQGMQ XQKMVSKMTI E XQLEEAMLI KE I EQGMQ XQKMVSKMTI E XQLEEAMLI XQLEEAMLI XQLEEAMLI XQLEEAMLI XQLEEAMLI XQ XQ XQ XQ XQ XQ XQ XQ XQ XQ XQ XQ XQ	α15 RKLLP FKLLP IEALTLCEKI DEA IEMVEQO 0 340	* mo	tif IV 250 26 GKFDY I KGDSE EKGGTASEAS - KQFD I QPNEAS - KKLKVTPLEC- EWVAARSEEVE - KKFNL TDAETS AKA I VLLDPEGHF PGGLAMYHPEWHG QTA I VVAKAGWNF
Vrn VrnA VrnA Vrn Vrn RecJ RecJ RecJ YybT	* 180 190 GGF TYNSN-NPE IYF I ISE GNF FANT-TPEVLRVAAE GRF LFPNT-TEKTLKYAGE GRF LFSNT-SWRTLYLGSN GSF WASV-RGYRLAAF GRF LYPAT-TSKTF I IASE GLAR IPAS-SWGLRILAE GLEPLRRT-NPLGLKELIK KSEFSLRTG-SRTFDAASY 270	200 210 200 200 200	VYNTYSE LQFR-PF LQFR-PF LNHT-SL LMDSHPF MDSF-PY AFRI-/ GFAL-C GFQL-/ LKET-VE	α13 220 SRLPURGVLS- 250 SSRLRGVLS- NVVKLNGF IFQ- %LIMPEGGYLQ- KD IQFKQYVFK- TWLPLLSRVLG- YRLNAAGRLDDW APRLNAAGRLDDW SYHLNAAGRLDDW SYHLNAAGRLDDW PRLNAVGR IEQA SSY IKRA-KLIQ QC17 300 PCNRLAAEFFN-	AELVTEL 230 NMKVYKU 230 NVSLSEN NFQTFQN SAQLVSE NECIDKN SAQLVSE NLEIDKN SAQLVSE NLEIDKN SAQLVSE NLEIDKN SAQLVSE MLEIDKN SAQLVSE MLEIDKN SAQLVSE MLEIDKN SAQLVSE MLEIDKN SAQLVSE MLEIDKN SAQLVSE MLEIDKN SAQLVSE SAQLSE SAQ	14. β8 DYNSA	240 —LISLTKEE —TAHLP —SVFIKKOT —QIKITNOV —FVADKKF —LYYVVVDN —RIILSQKI GELIFILNAFF NELDALNGT AE IDQLNKEF —ASLPENE —	20 1 20 1 20 1 20 1 20 20 20 20 20 20 20 20 20 20	α15 FKLLP FKLLP FALTLCEKI DEAIEMVEQG 340 YKPLLKE	* mo	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Vrn Vrn Vrn Vrn Vrn Vrn NecJ NecJ NecJ VybT Vrn Vrn	* 180 190 GGFTYNSN-NPE IYF I 15E GNFFFANT-TPEVLRVAAE GRFLFPNT-TEKTLKYAGE GRFLFSNT-SWRTTLYGSN GSFRWASV-RGYRLAAF GRFLYPAT-TSKTF I 1ASE GLARIPAS-SWGLRLLAE GMSRIRAGKCPPGIKALLE GLEPLRRT-NPLGLKELIK KSEFSLRTG-SRTFDAASY a16 B a16 B C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C C	200 210 200 200 200	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFRI-/ GFAL-C GFQL-/ LKET-VE 0 LKET-VE 0 LKET-VE 0 LKESVGKF	α13 220 SRLPURGVLS- 251FRLMGQVLS- 25YFRLMGQVLS- 25YFRLMGQVLS- 25WLMPFGGYVLG- KLMPFGGYVLG- 7WLPLLSRVLG- 7WLPLLSRVLG- 201 202 203 204 205 205 206 207 208 209 200 200 200 200 200 200 200 201	AELVTEL - β7 α1 230 - NMKVYKU - TVAFHFG - NFELSDS - NFQTFGN - SAQLVSE - NLEIDKN - SAQLVSE - NLEIDKN - NLEIDKN	14. β8 DYNSA	240 —LISLTKEE —TAHLP —QIKITKDV —QIKITKDV —FVADKKF —LVYVVVDA —RIILSQKI GELIFILNAFF MELDALNQTF AEIDQLNKEF —ASLPENE 20 3 TYGTMEEAVK KGLDLQ	Q Q Q Q R L Q TLEEAMLI KE I EQGMQ NKE I EQGMQ NFEQALEK AYERVLEA [*]	α15 FKLLP FKLLP FALTLCEKI DEAIEMVEQG 340 YKPLLKE	* mo —QADPE _ERSRDTL QGL — D	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Vrn Vrn Vrn Vrn Vrn RecJ RecJ Vrn Vrn Vrn Vrn	* 180 190 GGF TYNSN-NFE IYF I 15 GNFF FANT-TFEVLFVAAE GRF LFPNT-SHITIEVASC GRF LYNT-SHITIEVASC GRF LYNT-SWITIEVASC GRF LYPAT-TSKTF I IASE GLAR IPAS-SWYGLRLLAE GLAR IPAS-SWYGLRLLAE GLAR IPAS-SWYGLRLLAE GLAR IPAS-SWYGLRLLAE GLAR IPAS-SWYGLRLAE GLAR IPAS-SWYGLRLA	Construction of the second sec	VYNTYSE LQFR-PF LYET-KL MSEK-DF MSEK-DF MDSF-PY AFRI-4 GFAL-(GFQL-4 LKET-VE LKET-VE 0 LRSVGKF IBSRGGV F <mark>B</mark> SKG-	α13 220 SRLPLAGYVLS- 2STFRLMGQVLS- SYFFRLMGQVLS- WKLMPFQGYVLQ- KLMPFQGYVLQ- KLMPFQGYVLG- TWLPLLSRVLG- YPRLNAAGRLDDW APPLNAAGRLDDW QSY IKRA-KLIQ- 300 PONPLAAEFFN- -SQNIALKLG-	AELVTEL - β7 α1 230 - NMKVYKL - TVAFHFC - NVSLSEN - NFELSDS - NFE	14. β8 DYNSA	240 —LISLTKEE —TAHLP —SVFIKIDT —QIKITKDT —EVADKKF —EVADKKF —LVYVVVDI —RIILSQKI GELIFILNARF NELDALNQTF AE IDQLNKEF —ASLPENE — 20 3 PGTMEEAVK KGLDLDQ — YSWDEAD	Q	α15 RKLLP	* mo —QADPE _ERSRDTL QGL — D	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Vrn Vrn Vrn Vrn Vrn Nrn RecJ RecJ Vrn Vrn Vrn Vrn Vrn	* 180 190 GGF TYNSN-NFE I YF I I SE GNFF FANT-TFEVLFVAAE GRF LFSNT-SPHTMEVASC GRF LFSNT-SWFIT VLGSN GSF WASV-RGYFLAAF GRF LYPAT-TSKTF I I ASE GLAR I PAS-SWGLRLLAE GLAR I PAS-SWGLRLAE GLAR I PAS-SWGLR GLAR I PAS-SWGLR		VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFRI-Z GFQL-4 LKET-VE Q GFQL-4 LKET-VE Q SAL-6 GFQL-4 LKET-VE LBSVGKF IBSRGGN FBSKG-	CA13 CONTRACTOR CONTR	AELVTEL - β7 α1 230 - NMKVYKC - TVAFHFC - NFELSDS - NFE	14. β8 SGLLV SGLA SGLA SENEYC WIY SENEYC WIY SENEYC WAG AA VAGAA SENEYC VAGAA SENEYC VAGAA SENEYC VAGAA SENEYC SEE SENEYC SE SENEYC SE SENEYC SE SE SE SE SE SE	240 —LISLTKEE —TAHLP —QIKITNDV —QIKITNDV —RIILSQKI GELIFILNARF NELDALNQTF AEIDOLNKEF —ASLPENE 20 3 GGTMEEAVK KGLDLDQ UYSWDEAD —YSWDEFE —	Q	α15 RKLLP IEALTLCEKI DEALENTEQUE 340 YKPLLKE VREELTRAG	* mo —QADPE _ERSRDTL QGL — D	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Vrn Vrn Vrn Vrn Vrn Nrn RecJ RecJ RecJ Vrn RecJ Vrn Vrn Vrn Vrn Vrn Vrn	* at 1 at 2 at 3 at 2 at 3 at 3		VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSH-PY MDSF-PY AFRI-2 GFQL-4 GFQL-4 GFQL-4 LKET-VE LBSVGKF IBSKGGA FBSKG- FBSKG- FBSKG-	α13 220 SRLMGVLS- SSTFLMGVLS- SYFFLMGVLS- VKLNGFIFQ- KUMPGQYLS- INVKLNGFIFQ- KUMPGQYLS- INVKLNGFIFQ- KUMPGQYLS- SYFFLMGQVLS- INVKLNGFIFQ- KUMPGQYLS- INVKLNGFIFQ- KUPLSVLG YPLNAGRE SYIKBA-KLIQ- SYIKBA-KLIQ- 300 PCNRLAAEFFN- -SQNIALKLG- YINGLARKYN- ITINDVANQFG- INNREVALKYG-	AELVTEL - β7 α1 230 - NMKVYKL - TVAFHEO - NFELSDS - NF	14. β8 SQLV SQLV VG-AA SHEYC SHEYC VI Y AVGGRG SHEYC VI Y SALACA DAAEAQALVI VI QAAEAQALVI DAAEAQALVI SEVERALA DAAEAQALVI SEVERALA DAAEAQALVI SEVERALA DAAEAQALVI SEVERALA SHEYCALA SEVERALA DI IAI SEVERALA SHEYAAGATI SEVERALA SHEYAAGATI SEVERALAS SHPLASGAS SHPLASGAS SHIQASGAVI SHIQASGAVI	240 — LISLTKEE — TAHLP — SVFIKRD — GIKITNDV — FVAKKF — LVYVVVD — RIILSQKI GELHFLNAFF — ASLPENE — ASLPENE 20 3 RYGTMEEAVK KGLDLQ — YSWDEFE — VYSWDEFE — VYSWDEFE — VSWDEFE — VSWDEFE — VSWDIII—	Q L Q R L Q Q R L C M C C C C C C C C C C C C C	α15 RKLLP FKLLP IEALTLCEKI DEA IEMVEQO 340 YKPLLKE VKELTRAG LETLCKEHE LRQKLLE CQKQIAV	* mo —QADPE _ERSRDTL QGL — D	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Nrn Nrn Nrn Nrn Nrn RecJ RecJ YybT Nrn Nrn Nrn Nrn Nrn Nrn Nrn	* at 1 at 2 at 2	motif III α12 200 210 LLSKG IDKDD I YRK LLGYGYKLAELTDR LQYPFSSELEDFNQ LQYPFSSELEDFNQ LQYPFNHAELNK LYPAQAN I AK I HDE LQYPFNHAELNK LYPAQAN I AK I HDE LQYPADATYSRT LLKYDEDFAALARQ CAVGY — TGKAVEV VANRDAQKLAASDL LSGGD I GEANEETV LLAKGADTVLVQKF 280 29 Z80 29 Z80 EKKWIK IS FLEEDFKKMIK IS FLEEDFKKMIK IS FFVEE—DDQ I RPR FF I EEG-KNHYRVE WFKEVE—PHRWSVS	VYNTYSE LQFR-PF LYET-KL MSEK-DFL LMDSHPF MDSF-PY AFRI-Z GFAL-C G GFAL-C G GFAL-C G G G G G G G G G G G G G G G G G G G	α13 220 SRLPLINGYVLS- SSLPLINGYVLS- SYFFLINGQVLS- NYKLNGF IFQ- KLMPFQGYVLS- YKLMPFQGYVLS- NYKLNGF IFQ- KLMPFQGYVLS- SYFRLSRVLG- TWIPLLSRVLG- YKLMPFQGYVLS- YKLMPFQGYVLS- YKLMPFQGYVLS- YKLMPFQGYULS- YKLMPFQGYULS- YKLMPFQGYULS- YKLMPFQGYULS- YKLAAGRLDM YKRA-KLIQ- YKRA-KLQAYVFE- YKRA-KLQAYVFE- YKRA-KLQAYKSGFG-	AELVTEL - β7 α1 230 - NMKVYKL - TVAFHFC - NFELSDS - NF	14. β8 SQLV SQLV SGLV SHEYC VG-AA SHEYC VV 1Y SAVGRG VG-AA STEYC VG-AA STEYC VG-AA STEYC VG-AA STEAEELA DSFEAEELA STEAEELA DNIAI 37 SHUNASGGEI SHPVAGGGEI SHVPAAGATI SSEAEELA D15. SSEAELAS SHPLASGGS SHUASGGEI SHPLASGAS SHPNASGYS SHIQASGAVI SHEAGAY	240 — LISLTKEE — TAHLP — SVFIKRD — GIKITNDV — FVAKKF — LYYVVVD — RIILSQKI GELHFLNARF NELDALNQTF AEIDQLNKEF — ASLPENE 20 3 TYGTMEEAVK KGDLDQ YSWDEAD — YSWDEFE — KSKFDII — TGSIDDAVA	Q L Q R L Q R L Q Q L E 20 VFEQALEK AYERVLEA — RILADI — RILADI — RILADI — RIVEQA SLRAALG —	α15 RKLLP IEALTLCEKI DEALTLCEKI DEALTLCEKI DEALTLCEKI DEALTLCEKI DEALTLCEKI DEALTLCEKI COKULA COKUL	* mo —QADPE _ERSRDTL QGL — D	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Vrn VrnA VrnA Vrn Vrn AecJ RecJ RecJ YybT Vrn Vrn Vrn Vrn Vrn Vrn Vrn	* at 1 at 2 at 2	motif III α12 200 210 LLSKG IDKDD I YRK LQYGYKLAELTDR LQYGYKLAELTDR LQYGPKI-KAELTDR LQYGPKI-KAELTDR LQYGPKI-KAELNK LYPAQANIAK IHDE LYPAQANIAK IHDE LVYDIPDFAALARQ LQYOHOTAVSRT LLYYDHOFAALARQ LQYOHOTAVSRT LLYYDHOFAALARQ LQYALAYANIAK IHDE LQYADYANIAK IHDE	VYNTYSE LQFR-PF LYET-KL MSEK-DF LNHT-SL LMDSHPF MDSF-PY AFR1-/ GFAL-C GF	α13 220 SRLPLMGYVLS- 221 SRLPLMGYVLS- SYFFLMGQVLS- NVKLNGF IFQ- KLMPFQGYVLS- YKLMPFQYVK- TWIPLLSRVLG- YKLMPFQYVFK- TWIPLLSRVLG- YKLMPFQSLGA YRLAAGRLDDW YRLAAGRLDDW YRLAAGRLDDW PONBLAAGRLDDW YKLAAGRLDDW YKLAAGRLDDW YRLAAGRLDDW YKLAAKGR YKLAAKGGG- YNEVALKYG- YNEVALKGG- YNEVAKGGG- YNEVAKGFA-	ALLVTEL - P7 a1 230 - NMKVYKC - TVAFHFC - NVSLSE - NFELSDS - NFELSDS	14. β8 SQLV SQLV SQLV SQLV VG-AA SHEYC W1Y SAGEAGA VG-AA SHEYC VG-AA SHEYC VG-AA SHEYC VG-AA SHEYC VG-AA SGRAGALVINI SGEAEALA SGRAGALVINI SGHUAAGAGATI SGHUAAGGAS SHPLASGGS SGHUAAGGAS SHILAAGYT SHPLASGAN	240 —LISLTKEE —TAHLP —SVFIKADT —QIKITNDV —FVADKKF —RIILSOKI GELHPLNARF NELDALNQTF AEIDQLNKEF —ASLPENE — YGTMEEAVK UKGLDLDQ YSWDAD — YSWDAD — YSWDAD — YSWDAD	Q L Q R L Q R L Q Q R Q Q Q Q Q Q Q Q Q Q Q Q Q	α15 RKLLP IEALTLCEKL DEALTLCEKL DEALTLCEKL DEALTLCEKL DEALTLCKEHE LETLCKEHE COKQIAV COKQIAV LKNLLK	* mo	tif IV 250 24 GKFDY I KGDSI EDAGAEEDSI EKGFD I QPNEA: KKFN TDAET3 AKA I VLLDPEGHI PGGLAMYHPEWHG QTA I VVAKAGWNI
Vrn VrnA VrnA Vrn Nrn AecJ AecJ AecJ YybT Vrn Vrn Vrn Vrn Vrn Vrn Vrn Vrn Vrn	* attack	motif III α12 200 210 LLSKG IDKDD I YRK LQYGYKLAELTDR LQYGFNITNAELNK LQYDFNITNAELNK LQYDFNITNAELNK LQYDANIXKI HDE LUYDFDFAALARQ LQYDANIXKI HDE LUYDFDFAALARQ LQYDANIXKI HDE LLYVDFDFAALARQ LQYDANIXKI HDE LQYDANIXKI HDE LLYVDFDFAALARQ LQYDANIXKI HDE LLYVDFDFAALARQ LQYDANIXKI HDE	VYNTYSE LQFR-PF LQFR-PF LNHT-SL LMDSHPF MDSF-PY MDSF-PY AFR1-/2 GFAL-C GFQL-/2 LKET-VC 0 LBSVGKF IBSNG- FBSNG- MBAKT- LBSKS- VPSLAP1	α13 220 SRLPLMGYLS- SSTFRLMGYLS- SYFRLMGYLS- SYFRLMGYLS- NVMLNGFIFQ- KLPLGYVLG- KLPLSRVLG- KLPLLSRVLG- KLPLLSRVLG- SPRLNAAGRLDDW SYIKRA-KLIGEA SYIKRA-KLIGEA SYIKRA-KLIGEA SYIKRA-KLIGEA SO0 PCNRLAAGEFEN- -YINGLAKKGN- INTINDANGFG- INVEXULKYG- VINLAAVASGFG- TVINEVALKYG- SVIKAARBA-	ALLYTEL 230 NIKVYKC TVAFHFC NVSLSE NFQTFQN SAQLVSE NFQTFQN SAQLVSE NLEIDKN NEXTFQN NLEIDKN NEXTFQN SAQLVSE NLEIDKN SAQLVSE	14 β8 SQL V SQL V SQL V SQL V VG-AA SHEYC VV I Y SQL V SHEYC VI Y SAEAGAG SHEYC VG-AA STEAEELA. DAAEAQAL VI SGEAELA. DAAEAQAL VI SGEAELA. DNIAI SGEAELA. DOI AI SGHUAAGGAS SHPLASGGE SHPLASGGAS SHPLASGAS SHPLASGAS SHPLASGAAK SHENASGAS SHPLASGAAK SHPLASGAAK	240 —LISLTKEE —TAHLP —SVFIKADT —QIKITNDV —FVADKKF —LYYVVON —FVADKKF —RIILSOKI GELHFLNARF NELDALNQTF AEIDQLNKEF —ASLPENE — YSWDET— USWDET— USSKFDII— TGSIDDAV/ SYSLAENE— NDE—ALFF	Q L Q R L XTLEEAMLI KE IEQGMQ XFEQALEK AYERVLEA —RILADI —RILADI —RILADI —RILADI —RIVQO SLRAALG —DIYQEI AFKARVEA	α15 RKLLP IEALTLCEKL DEALTLCEKL DEALTLCEKL DEALTLCEKL DEALTLCKE VREELTRAG' LETLCKDE LETLCKDE CQKQ1AV CQKQ1AV LKNLLK YAAREPDPVF	* mo	tif IV 250 26 GKFDY I KGDSI EDAGAEEDSI EKGFDI QPNEAS KKLKVTPLEC EWVAARSEVI KKKN TDAETS AKA I VLLDPEGHI PGGLAMYHPEWKK QTA I VVAKAGWNI
Nrn Nrn Nrn Nrn Nrn RecJ RecJ RecJ YybT Nrn Nrn Nrn Nrn Nrn Nrn Nrn RecJ RecJ RecJ RecJ	* attimustory of the second		VYNTYSE LQFR-PF LQFR-PF LNHT-SL LMDSHPF MDSF-PN AFRI-/ GFALC GFQL-/ LKET-VC 0 LKET-VC 0 LKST-VC 0 LRSVGKF IBSRGGV FRSNG FRSNG- FRSNG- FRSNG- BSSLG- URSKS- VRSLAPI GRSLQGL	α13 220 SRLPLMGYVLS- SSTFRLMGYVLS- SYFRLMGYVLS- YVVKLNGF IFQ- YKLPLLSRVLG- KLPLGKQYVFK- TWLPLLSRVLG- TWLPLLSRVLG- YRLNAVGR IEQA SSY IKRA-KLIQA SO0 PCNRLAAGRLDDW YCNRLAAGRLDDW YRLNAVGR IEQA SSY IKRA-KLIQA YRLNAVGR IEQA SO0 PCNRLAAGRLDDW YLNAVGR IEQA YNRLAVSGG YNRLAVSGGG TINEVAKRHA- SAVEALSKAEL -AVEALSKAEL SAVEALRSAEL	ALLEYTEL 230 NKVYKL TVAFHFG NVSLSE NFQTFQN SAQLVSE NLEIDKN NEALLIDKN NEALLIDKN NEKALLLIDE HTVLYKE 600 600 600 600 600 600 600 60	14. B8 DYNSA DYNSA VG-AA SHEYC SHEYC WI Y GAA WI Y GAA VY I Y GACAS VY I Y SHEYC JAAEAQAL VI DAEAQAL VI JSFEAEALA DNI A I B12 JO SHPLASGGE SHPASGSS SHI QASGAV SHI QASGAV SHPLASGAN SHPLASGAN SHPLASGAN SHPLASGAN SHEAGT <td>240 —LISLTKEE —TAHLP —SVFIKKT —QIKITNDV —FVADKKF —LVYVVVDM —RIILSQKI GELHELNAFF AEIDQLNKEF —ASLPENE 20 3 YGTMEEAVK LKGLDQA VYSWDEFE —LSKFRUI TGSIDDAVA SYSLAENE —NDE—ALFF LEE—DKFK</td> <td>Q</td> <td>α15 RKLLP IEALTLCEKI DEALTLCEKI DEALENVEQ 340 YKPLLKE LFTLCKEHE LRQKLLE CQKQIAV LANLLK YAARFPDPVFL</td> <td>* mo QADPE QADPE D D /L </td> <td>tif IV 250 2E GKFDY I KGDSE EKFGTTASEAS - KQFD I QPNEAS - KQFD I QPNEAS - KKLKVTPLEC - EWVAARSEEVE - KKFNL TDAETS AKA I VLLDPEGHF PGGLAMYHPEWHG QTA I VVAKAGWNF</td>	240 —LISLTKEE —TAHLP —SVFIKKT —QIKITNDV —FVADKKF —LVYVVVDM —RIILSQKI GELHELNAFF AEIDQLNKEF —ASLPENE 20 3 YGTMEEAVK LKGLDQA VYSWDEFE —LSKFRUI TGSIDDAVA SYSLAENE —NDE—ALFF LEE—DKFK	Q	α15 RKLLP IEALTLCEKI DEALTLCEKI DEALENVEQ 340 YKPLLKE LFTLCKEHE LRQKLLE CQKQIAV LANLLK YAARFPDPVFL	* mo QADPE QADPE D D /L 	tif IV 250 2E GKFDY I KGDSE EKFGTTASEAS - KQFD I QPNEAS - KQFD I QPNEAS - KKLKVTPLEC - EWVAARSEEVE - KKFNL TDAETS AKA I VLLDPEGHF PGGLAMYHPEWHG QTA I VVAKAGWNF
Vrn Vrn Vrn Vrn Vrn Vrn Vrn Vrn Vrn Vrn	* atti		VYNTYSE LQFR-PF LQFR-PF MSEK-DF MDSF-PY AFRI-/ GFAL-C GFQL-/ LKET-VE U BSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV C BSSRGG C BSSRG C BSSRG C BSSRG C BSSRG C BSS C C BSS C C C C C C C C C C C C C	α13 220 SRLPLMGYVLS- SSTEPELMGQVLS- NVVKLNGF IFQ- %LIMPFGGYVLG- KI JQFKQYVFK- TWLPLLSRVLG- TWLPLLSRVLG- YRLNAVGR IEQA SYFLMAGRLDDW PPLNAVGR IEQA SO0 PPLNAVGR IEQA 300 PCNRLAAEFFN- -SAQNI ALKLG- PVINGLARKYN- I TINDVANOFG- IVNPEVALKYG- TVINEVAKRHA- -SAVEALRSAEC -HYRDALERLD1	ГАЕLVTEL 	14. β8 DYNSA G GGLLV VG-AA SHEYC VI EAVGGRG VVIY GALAC SHEYC VVIY GALAC SHEYC SHEYC VVIY GACAC SACAGAC SHEYC VG-AA DAAEAQALVU NIGEARVLAI DSFEAEELA. DNIAI B12 D 33 SHLNASGGEI SHVPAAGATI SHUASGGAV SHPLASGAS SHPLASGAS SHPLASGAS SHRLAAGYT SHRLAAGFAI SHAMAAGLSI SHAMAAGLSI	240 —LISLTKEE —TAHLP —QIKITNO —QIKITNO —QIKITNO —QIKITNO —IVANKK —LVYVVVD —RIILSQKI GELIFILNAFF MELDALNQTF AEIDQLNKEF —ASLPENE — 20 3 TYGTMEEAP VYSWDEAD VYSWDEAD VYSWDEAD — VYSWDEAD — VSWDEFF — USKRDII — USKRDII — NELDALNQTF — ASLPENE — MELDALNQTF — ASLPENE — MELDALNQTF — MELDALNT	Q	α15 FKLLP FKLLP FKLLP FKLLP FKLLP FKLLF FKLF FKL	* mo	tif IV
Irn Irn Irn Irn Irn Irn Irn Irn Irn Irn	* attimustory of the second		VYNTYSE LQFR-PF LQFR-PF MSEK-DF MDSF-PY AFRI-/ GFAL-C GFQL-/ LKET-VE U BSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV FBSRGGV C BSSRGG C BSSRG C BSSRG C BSSRG C BSSRG C BSS C C BSS C C C C C C C C C C C C C	α13 220 SRLPLMGYVLS- SSTEPELMGQVLS- NVVKLNGF IFQ- %LIMPFGGYVLG- KI JQFKQYVFK- TWLPLLSRVLG- TWLPLLSRVLG- YRLNAVGR IEQA SYFLMAGRLDDW PPLNAVGR IEQA SO0 PPLNAVGR IEQA 300 PCNRLAAEFFN- -SAQNI ALKLG- PVINGLARKYN- I TINDVANOFG- IVNPEVALKYG- TVINEVAKRHA- -SAVEALRSAEC -HYRDALERLD1	ГАЕLVTEL 	14. β8 DYNSA	240 —LISLTKEE —TAHLP —SVFIKKT —QIKITNDV —FVADKKF —LVYVVVDM —RIILSQKI GELHELNAFF AEIDQLNKEF —ASLPENE 20 3 YGTMEEAVK LKGLDQA VYSWDEFE —LSKFRUI TGSIDDAVA SYSLAENE —NDE—ALFF LEE—DKFK	Q	α15 RKLLP IEALTLCEKI DEALTLCEKI DEALENVEQ 340 YKPLLKE LFTLCKEHE LRQKLLE CQKQIAV LANLLK YAARFPDPVFL	* mo	tif IV 250 26 GKFDY I KGDSE EKGGTASEAS - KQFD I QPNEAS - KKLKVTPLEC- EWVAARSEEVE - KKFNL TDAETS AKA I VLLDPEGHF PGGLAMYHPEWHG QTA I VVAKAGWNF

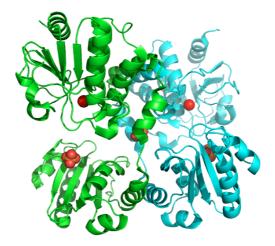
Supplementary Fig. S4. Sequence alignments of bfNrn, its homologs and other DHH/DHHA1 family proteins. The secondary structure of bfNrn is shown above the sequences; the horizontal bars indicate α -helices and arrows indicate β -strands. The five motifs proposed by Aravind and Koonin [4] are highlighted in red boxes. Residues thought to be involved in binding to a ligand are highlighted by an orange background. Species name are as follows: bf, *Bacteroides fragilis*; tt, *Thermus thermophilus*; bs, *Bacillus subtilis*; sh, *Staphylococcus haemolyticus*; mp, *Mycoplasma pneumoniae*; mt, *Mycobacterium tuberculosis*; sm, *Streptococcus mutans*; and ec, *Escherichia coli*. The sequence alignment was prepared using Clustal omega [5].



Supplementary Fig. S5. GMP-binding site. (A) The Fo-Fc electron density map (blue mesh) for the GMP is contoured at the 1.0 σ level. (B) Conformation of β 11 strand formed by the DHH1A motif. (C) Comparison between Mn²⁺-GMP complex (cyan sticks) and ligand-free form of bfNrn (orange sticks; PDB ID: 3DMA). (D) Superimposition of Lys287 and Arg291 of bfNrn (cyan sticks) with the corresponding residues of ttRecJ (green sticks; PDB ID: 2ZXP). The dashed lines indicate the distances (Å).



Supplementary Fig. S6. Putative domain closure of bfNrn to adopt a catalytically active conformation. A model structure after domain closure (cyan) was constructed by moving the DHHA1 domain to the DHH domain using Pymol. The ligand-free form of bfNrn (orange, PDB ID: 3DMA) is superimposed.



Supplementary Fig. S7. Position of sulfate ions in the structure of shNrn dimer (PDB ID: 3DEV). Each subunit is drawn in a different color. The position of each Mg²⁺ ion is shown as a red sphere. Three sulfate ions are shown as space-filling models colored by element type.

Supplementary References

- [1] Wakamatsu, T., Kim, K., Uemura, Y., Nakagawa, N., Kuramitsu, S. and Masui, R.
 (2011) Role of RecJ-like protein with 5'-3' exonuclease activity in oligo(deoxy)nucleotide degradation. J. Biol. Chem. 286, 2807-2816.
- [2] Fukui, K., Takahata, Y., Nakagawa, N., Kuramitsu, S. and Masui, R. (2007) Analysis of a nuclease activity of catalytic domain of *Thermus thermophilus* MutS2 by high-accuracy mass spectrometry. Nucleic Acids Res. 35, e100.
- [3] Wakamatsu, T., Nakagawa, N., Kuramitsu, S. and Masui, R. (2008) Structural basis for different substrate specificities of two ADP-ribose pyrophosphatases from *Thermus thermophilus* HB8. J. Bacteriol. 190, 1108-1117.
- [4] Aravind, L. and Koonin, E.V. (1998) A novel family of predicted phosphoesterases includes Drosophila prune protein and bacterial RecJ exonuclease. Trends Biochem. Sci. 23, 17-19.
- [5] Sievers, F. et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. 7, 539.