research papers

Acta Crystallographica Section D Biological Crystallography

ISSN 0907-4449

Hiroshi Itou,^a‡ Min Yao,^{a,b} Nobuhisa Watanabe^{a,b} and Isao Tanaka^{a,b}*

^aFrontier Research Center for Post-genomic Science and Technology, Hokkaido University, Sapporo 001-0021, Japan, and ^bDivision of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060-0810, Japan

Present address: Biomolecular Structure Laboratory, Structural Biology Center, National Institute of Genetics, Mishima 411-8540, Japan.

Correspondence e-mail: tanaka@castor.sci.hokudai.ac.jp

Structure analysis of PH1161 protein, a transcriptional activator TenA homologue from the hyperthermophilic archaeon *Pyrococcus horikoshii*

The crystal structure of the Bacillus subtilis TenA-homologue protein PH1161 from the hyperthermophilic archaebacterium Pyrococcus horikoshii was determined. TenA is known to belong to a new family of activators that stimulate the production of extracellular proteases in B. subtilis. A sequence-similarity search revealed that TenA-homologue proteins are widespread in bacteria and archaea, suggesting that this family of proteins plays an essential role in these organisms. In the present study, the first three-dimensional structure of a member of the TenA family of proteins was determined, unexpectedly revealing that the protein has a fold identical to that of haem oxygenase-1. Analysis has also shown that the protein has a unique ligand-binding pocket. Electron density of a bound ligand molecule was observed in this pocket. These results provide a valuable insight into the functional understanding of the TenA family of proteins.

Received 12 December 2003 Accepted 9 April 2004

PDB Reference: PH1161, 1udd, r1uddsf.

1. Introduction

To adapt to environmental change, organisms possess a wide variety of signal-transduction systems that mediate gene expression. The *Bacillus subtilis deg* system consisting of *degS*, *degU*, *degQ* and *degR* is known to be one such signaltransduction system and regulates the rate of synthesis of degradative enzymes, including an intracellular protease and secreted enzymes (Msadek *et al.*, 1990). The products of the *degS* and *degU* genes, DegS and DegU proteins, have been characterized as being a sensor and a transcriptional regulator, respectively, from their amino-acid sequence similarity (Msadek *et al.*, 1990), and the pair act as a two-component (modulator–effector) pair, similar to the *Escherichia coli* EnvZ–OmpR proteins (Stock *et al.*, 1989).

Another protein factor that affects the expression levels of extracellular protease *via* the *deg* system was identified in *B. subtilis*. This new effector protein, named TenA, is a protein of molecular weight 27.4 kDa that consists of 236 amino acids; it enhances the production of several extracellular proteases by between 11-fold and 55-fold. A functional DegS is required to observe this stimulatory effect, suggesting that TenA acts indirectly to enhance the production of extracellular enzymes (Pang *et al.*, 1991). A BLAST (Altschul *et al.*, 1997) search against all protein-sequence databases revealed that proteins highly homologous to TenA are widely conserved in both the bacteria and archaea, suggesting that these TenA-like proteins play an essential role during gene activation in these organisms. The product of ORF *ph1161* from the hyperthermophilic archaeon *Pyrococcus horikoshii* OT3 (Kawarabayasi *et al.*,

C 2004 International Union of Crystallography Printed in Denmark – all rights reserved

Summary of data collection.

Values in parentheses are for the outermost resolution shell.

		SeMAD		
	Native	Peak	Edge	Remote
Wavelength (Å)	1.5418	0.9776	0.9778	0.9815
Resolution (Å)	40-2.15 (2.27-2.15)	40-2.5 (2.59-2.5	50)	
Space group	P3 ₂	P222 ₁	,	
No. observed reflections	220886 (31384)	_	_	_
Unique reflections	58735 (8595)	37833 (3496)	37819 (3506)	37695 (3324)
Completeness (%)	99.7 (99.7)	98.6 (93.0)	98.6 (93.4)	98.1 (87.9)
Average redundancy	3.8 (3.7)	6.9 (6.5)	6.9 (6.6)	7.4 (6.7)
Average $I/\sigma(I)$	7.4 (3.0)	17.7 (5.1)	17.1 (4.9)	17.0 (4.8)
R_{meas} $(\%)$	8.2 (28.4)	_ ``	_ ` `	_ ``
$R_{\rm merge}$ \ddagger (%)	_ `	6.8 (24.1)	6.3 (23.9)	6.3 (24.4)
R_{λ} § (%)	_	4.9 (10.4)	4.2 (10.3)	_ ` `

† $R_{\text{meas}} = \sum_{h} [m/(m-1)]^{1/2} \sum_{j} |\langle I \rangle_{h} - I_{h,j}| / \sum_{h} \sum_{j} I_{h,j}$, where $\langle I \rangle_{h}$ is the mean intensity of symmetry-equivalent reflections and *m* is the redundancy (Diederichs & Karplus, 1997). This factor is also called $R_{r.i.m.}$ (the redundancy-independent merging *R* factor; Weiss, 2001). ‡ $R_{\text{merge}} = \sum_{h} \sum_{j} |\langle I \rangle_{h} - I_{h,j}| / \sum_{h} \sum_{j} I_{h,j}$, where $\langle I \rangle_{h}$ is the mean intensity of symmetry-equivalent reflections. § $R_{\lambda} = \sum_{h} ||F_{\lambda j}| - |F_{\lambda 0}|| / \sum_{h} |F_{\lambda j}|$, where $F_{\lambda j}$ is the structure factor of the data collected at the remote wavelength.

1998) is a protein that has 26% amino-acid sequence identity (45% positives with 5% gaps) to *B. subtilis* TenA. For structural-based functional analysis of TenA, the threedimensional structure of PH1161 protein, a homologue of TenA, was determined at 2.15 Å resolution. PH1161 is a protein of molecular weight 25.5 kDa that consists of 218 amino acids. The crystal structure shows that the protein has a unique ligand-binding pocket, while exhibiting significant structural similarity to haem oxygenase-1. The structure also showed that an unknown ligand molecule was bound in the pocket. These facts provide valuable information for elucidating the functions of the TenA protein, the detailed biological activities of which are still unknown.

2. Materials and methods

2.1. Sample preparation

The full-length (amino acids 1–218) PH1161 gene was amplified by PCR using the *P. horikoshii* OT3 genome as template DNA and was inserted into a pET-26b vector (Novegen) with *NdeI* and *XhoI* restriction-enzyme sites. The 3'-end primer for this cloning was designed for direct attachment of a $6 \times$ His tag to the C-terminal end of the PH1161 gene *via* a two-amino-acid linker Leu and Glu (pET/PH1161-His). The vector was co-transformed into *E. coli* BL21-Star DE3 (Stratagene), the expression host cell with pT-RIL plasmid (Stratagene).

The cells were grown at 310 K. Overexpression of recombinant PH1161-His (termed PH1161 protein) was induced with 1 m*M* IPTG. After IPTG induction, the medium was incubated at 298 K for 14 h with shaking in order to prevent the formation of inclusion bodies of the target protein. The cells were harvested by centrifugation at 3500g for 20 min at 277 K and were resuspended in 50 m*M* Tris–HCl pH 7.5 containing 0.15 *M* NaCl, 0.1 m*M* PMSF, 1 m*M* DTT, 5 m*M* MgSO₄ and 1 mg DNaseI. The cells were disrupted by a

French press (Aminco Inc.) at 83 MPa and the homogenate was clarified by centrifugation at 14 000g for 0.5 h at 277 K. The supernatant of the cell extract was heat-treated (333 K) for 0.5 h to remove proteins from the E. coli host cell by heat denaturing. After heat treatment, the supernatant was clarified by centrifugation at 14 000g for 0.5 h at 293 K. The supernatant was applied onto an Ni-chelating affinity column (Hi-trap Chelating, Amersham Biosciences) equilibrated with binding buffer (20 mM Tris-HCl pH 7.5 containing 500 mM NaCl and 20 mM imidazole). After washing out the unbound proteins with a sufficient amount of binding buffer, PH1161 protein was eluted with elution buffer (20 mM Tris-HCl pH 7.5 containing

500 m*M* NaCl and 500 m*M* imidazole). The PH1161containing fraction was loaded onto a Hi-load 26/60 Superdex200pg size-exclusion chromatography column (Amersham Biosciences). The PH1161 protein was eluted as a single peak. The PH1161 fraction was dialyzed against crystallization buffer (20 m*M* Tris–HCl pH 7.5 containing 200 m*M* NaCl and 10% glycerol) and concentrated by ultrafiltration using Apollo (Orbital Biosciences) to a final concentration of 5 mg ml⁻¹. In this study, the multiple wavelength anomalous diffraction (MAD) method was used for phase calculation. SeMetsubstituted PH1161 protein for use in the MAD method was also prepared following the same method as for the native protein with slight modifications.

2.2. Crystallization

Crystallization experiments were performed using the sitting-drop vapour-diffusion method in a 96-well plate (Greiner Bio-one) and the hanging-drop vapour-diffusion method in a 24-well VDX plate (Hampton Research) at 293 K. The initial crystallization condition screening (96-well plate) was carried out using 0.1 ml reservoirs of sparse-matrix crystal screening kits from Hampton Research and deCODE Genetics. Each drop contained 1 µl reservoir solution and an equal volume of protein solution. The best native crystal was obtained using condition No. 31 of Cryo 1 (0.1 M citrate pH 5.5, 35% 2-propanol, 5% PEG 1K; deCODE Genetics) and further optimization was carried out using the hanging-drop method. The dimensions of the native crystal were $0.3 \times 0.3 \times$ 0.2 mm and the crystal belonged to space group $P3_2$, with unitcell parameters a = b = 116.8, c = 71.0 Å, $\gamma = 120^{\circ}$; the asymmetric unit contained four molecules of PH1161 protein. The crystal volume per protein weight ($V_{\rm M}$; Matthews, 1968) was 2.7 Å^3 Da⁻¹, with a solvent content of 55.1%.

The initial SeMet PH1161 crystal was obtained using condition No. 9 of Crystal Screen (0.1 M sodium citrate pH 5.6, 0.2 M ammonium acetate, 30% PEG 4K; Hampton

Table 2

Summary of phase calculation for SeMet PH1161 protein.

	Remote	Edge	Peak
Resolution (Å)	20-2.5		
Space group	$P222_{1}$		
MAD phasing			
$R_{\text{Cullis}}^{\dagger}$		0.460	0.503
Phasing power_iso‡		3.586	3.105
Phasing power_ano§	0.668	1.162	1.538
FOM	0.631		
NCS averaging			
CC††	0.886		

† R_{Cullis} is the mean residual lack-of-closure error divided by the dispersive difference. Values are for centric reflections. ‡ Phasing power_iso is the root-mean-square of F_H/E , where F_H is the dispersive difference of F_H and E is the lack-of-closure error. § Phasing power_ano is as for Phasing power_iso, except that F_H is the anomalous difference of F_{H} . ¶ FOM is the mean figure of merit. †† CC is the standard linear correlation coefficient between the ρ values of the electron density related by the NCS operation.

Research). Optimization of the crystallization condition was carried out and the best crystal of SeMet PH1161 was obtained using 0.1 *M* sodium citrate pH 5.2, 0.2 *M* ammonium acetate, 30% PEG 4K and 15% glycerol. The dimensions of the SeMet crystal were $0.4 \times 0.2 \times 0.2$ mm and this crystal belonged to space group *P*222₁, with unit-cell parameters *a* = 154.1, *b* = 77.6, *c* = 89.6 Å; the asymmetric unit again contained four molecules of PH1161. The crystal volume per protein weight (*V*_M) was 2.6 Å³ Da⁻¹ and the solvent content was 52.7%.

2.3. Data collection

Data-collection statistics are given in Table 1. Native data were collected at a wavelength of 1.5418 Å at 100 K by flashcooling on an in-house R-AXIS IV^{++} data-collection system equipped with a Micro-Max 7 X-ray generator (Rigaku MSC). The MAD data set was collected at three different wavelengths (0.9778, 0.9776 and 0.9815 Å) from a single SeMetsubstituted crystal at 100 K on beamline BL38B1, SPring-8 (Hyogo, Japan). The native data were processed using the programs *MOSFLM* (Leslie, 1993) and *SCALA* (Evans, 1997) and the MAD data were processed using the program *DENZO* (Otwinowski & Minor, 1997).

2.4. Phase calculation

The asymmetric unit contains four molecules of PH1161, each of which contains six SeMet residues. 13 of the 24 Se sites were found using *SOLVE* (Terwilliger & Berendzen, 1999). Initial phases were calculated with the 13 sites and three additional sites were located by difference Fourier methods using *SHARP* (de La Fortelle & Bricogne, 1997). The phases were improved by NCS averaging using *DM* (Cowtan & Main, 1998). The phasing statistics are summarized in Table 2.

2.5. Model building and refinement

After the NCS averaging, a partial model (two-thirds of the molecule) was built using the graphics program *O* (Jones *et al.*, 1991). To improve the electron-density map, multi-crystal NCS averaging using the SeMet crystal and the native crystal was performed using *DMMULTI* (Cowtan & Main, 1998). The

Table 3

Final refinement statistics.

Resolution range (Å)	10.0-2.15
No. reflections	58170
Completeness (%)	99.8
Total number of non-H atoms	
Protein	7229
Solvent	155
R factor [†] (%)	21.8
$R_{\rm free}$ factor \ddagger (%)	24.6
R.m.s. deviation from standard values	
Bonds (Å)	0.009
Bond angles (°)	1.196
Dihedral angles (°)	18.12
Average B factor $(Å^2)$	34.6
Ramachandran plot§	
Residues in most favoured regions (%)	97.8
Residues in additional allowed regions (%)	1.8
Residues in generously allowed regions(%)	0.4
Residues in disallowed regions (%)	0

† R factor = $\sum |F_{obs} - F_{calc}| / \sum F_{obs}$, where F_{obs} and F_{calc} are the observed and calculated structure-factor amplitudes. ‡ The R_{free} factor was calculated as for the R factor, using an unrefined subset of reflection data (10%). \$ The Ramachandran plot was calculated using *PROCHECK* (Laskowski *et al.*, 1993).

native crystal was not isomorphous with the SeMet derivative; the orientation and the position of molecules in the native crystal were therefore determined by a molecular-replacement method using AMoRe (Navaza, 1994). From the improved electron-density map after averaging, another part of the partial model was built using O at 2.15 Å resolution. The other three molecules in the asymmetric unit were generated by NCS operations from the model. Structure refinement was carried out on the native data using CNS (Brünger et al., 1998). Manual model fitting was carried out between refinement rounds using O. After some rounds of refinement, the electron density of the unknown ligand molecules that bound to PH1161 were identified from both $2F_{0} - F_{c}$ and $F_{0} - F_{c}$ maps. Water molecules were found using CNS and the final model has an R factor of 21.8% and a free R factor of 24.6% for the data between 10 and 2.15 Å. The final refinement statistics are given in Table 3.

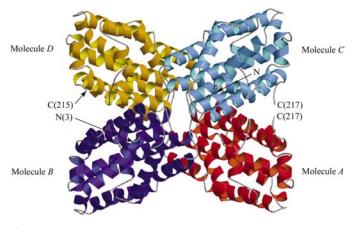


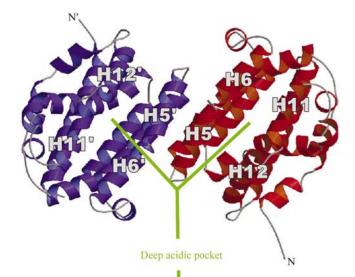
Figure 1

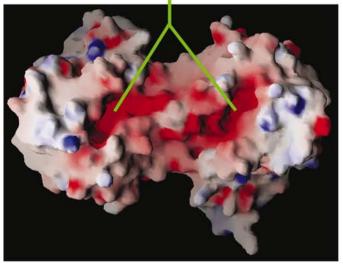
Ribbon representation of the PH1161 protein structure. The protein contains a dimer of dimers in the asymmetric unit, with 222 point-group symmetry. All ribbon representations were generated using *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Structure description: the monomer structure and its oligomer formation

A view of the overall structure of the PH1161 protein is shown in Fig. 1. Atomic models were successfully built from the first residue to Val217 for molecules A and C and from Val3 to Gly215 for molecules B and D. The unresolved parts of the models (C-terminal residues, including the attached His tag, and also the N-terminus for molecules B and D) were a consequence of disorder in the electron-density map of the corresponding regions. PH1161 is an α -helical protein and the molecule consists of 13 α -helices (H1, Ile5–Arg11; H2, Glu14–Phe21; H3, Phe24–Ser31; H4, Leu36–Lys63; H5, Met68–Glu79; H6, Val82–Glu93; H7, Glu98–Lys103; H8, Leu108– Lys123; H9, Ile126–Tyr146; H10, Lys148–Leu151; H11, Lys157–Leu167; H12, Asn169–Asp183; H13, Phe190–





(a)

Arg213). There is a deep pocket in the area surrounded by helices H5–H6 and H11–H12 (Fig. 2*a*). This pocket is highly charged in the acidic environment and the electron density of an unknown bound ligand molecule was observed in the pocket (Fig. 2*b*).

In Fig. 3, the amino-acid sequences of TenA-homologue proteins are aligned based on the structure of the PH1161 protein. These proteins show 26-77% sequence identity (41-87% positives) with each other and, as shown in this alignment, hydrophobic residues are well conserved among these homologues. The present PH1161 structure shows that most of these conserved hydrophobic residues (coloured yellow) are included in an internal hydrophobic core that maintains the tertiary structure of the monomer molecule. Hydrophobic residues corresponding to Phe21, Leu98, Val109, Leu151 and Ile158 in the PH1161 protein with relatively low levels of conservation are located at the edge of the hydrophobic core formed in the interior of the molecule and aminoacid substitutions in these residues are not likely to cause major structural changes in the protein structure. These findings suggest that the present structure of the PH1161 protein represents the three-dimensional structure of the TenAhomologue proteins.

The native and SeMet-substituted crystals of PH1161 protein used in the present structure analysis contained four molecules in the asymmetric unit. In both crystals, molecules were arranged with 222 point-group symmetry and formed a tetrameric oligomer (Fig. 1). Structural investigation showed that molecules A and C (and B and D) interact with each other *via* their H8 and H13 helices. In the interface, hydrophobic

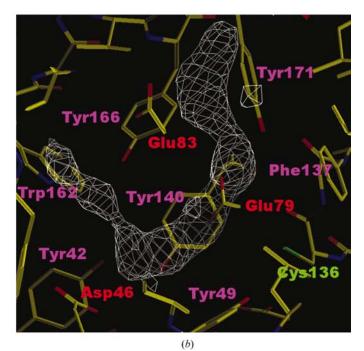


Figure 2

(a) Ribbon (top) and surface electric potential (bottom) representations viewed from the direction of the bottom of Fig. 1. In surface electric potential representation, red indicates acidic and blue represents basic charged areas. Each molecule has a deep pocket between helices H5–H6 and H11–H12 and the pockets are highly charged under acidic conditions. All surface electric potential representations were generated using *GRASP* (Nicholls *et al.*, 1991). (b) $F_0 - F_c$ map for the bound molecule at the acidic pocket of PH1161 protein. The map was contoured at 3σ . Amino acids around the unknown bound ligand molecule are labelled: red, acidic residues; purple, hydrophobic residues; green, conserved Cys residue.

interactions were observed, including two hydrogen-bonding pairs between the residues from each molecule. Each molecule contributes 1359 Å² of buried surface area to the formation of this dimer. Dimerization between molecules A and B (and Cand D) is achieved by an antiparallel interaction between helices H4 and H5 from each molecule. In the interface, 1809 Å^2 of buried surface area from each molecule contributes to the formation of the dimer. The H4 and H5 helices form a relatively flat surface as a dimer interface and extensive hydrophobic interactions and four hydrogen bonds between the residues from these molecules were observed. By this method of dimer formation, the opening of the deep acidic pocket from each molecule is arranged in the same direction and this side of the molecule presents a characteristic functional area of the protein (Fig. 2a). These structural features suggest that dimerization of molecules A and B (and C and D) is functionally more important. On the other hand, on formation of this tetramer of the PH1161 protein, the side chains of Tyr122 stick out from the molecular surfaces and Val59 residues from each molecule are in close contact, forming a hydrophobic core at the centre of the point-group symmetry. In addition, two hydrogen-bonding pairs between the main-chain carbonyl O atoms of the Tyr122 residues from diagonally opposite molecules (molecules A and D, and molecules B and C) via a water molecule were observed.

3.2. Functional implications: structural similarities and the bound ligand in the pocket

A DALI structural similarity search (Holm & Sander, 1995) revealed that the PH1161 protein has significant structural similarity to haem oxygenase-1 (HO-1) proteins (human, PDB code 1n45, Schuller *et al.*, 1999; rat, 1dve, Sugishima *et al.*, 2000; bacteria, 1j77; Schuller *et al.*, 2001). In particular, human HO-1 (hHO-1) is most similar to the PH1161 protein (Z score = 15.9; r.m.s. difference of 3.5 Å for 214 C^{α} atoms; Fig. 4).

fur_ionsis [000180] 1:					H1 -	H2 -	H3 -		H4		H5	H6	H7	
					10	20	30	40	50	60	70 8	90 90	100	
					1	1	1	1	1	1	1	1	1	
fur_issis [981188] 1	P horikoshii [C7105	58] 1	:	MRVMIT	DKLRRDSEQ	IWKKIFE	FVVQLYSGTLPL	EKEKEYVLQ	FNYLVGLTRAL	AVISSKAEYP-LM	AELIELARDE	TV VENYVKLL	KELDLTLEDAIKT	104
$ \begin{array}{c} \bern lim [pr2655] & 1 & \cdots & \bern Presentation by Line [pr VA, YSSS price in YVA, YSSSS price in YVA, YSSSS price in Y$	P abyssi [H75087]	1	;	MIT	ERLRRDADH	IWRKIFEHP	FVVQLYSGNLPM	EKFKFYVLQ	FNYLVGLTRAL	AVIASKADYP-LL	AELIELAREEI	TT MKNYEELL	KELGLTLEDAIKT	101
arcophilum [097289] 1:	P furiosus [Q8U188]	1] 1	:	MIT	ARLREEATE	IWEKIFKHD	FVAQLYEGTLPLI	DKEKFYILQ	FNYLVGLSRAL	AVIASKAEYP-LM	AEILELARAEI	TTEMKNYEDLL	NKLGFTLKDAIKV	101
ickolati [27]9186894] 1:	A pernix [872655]	1	:b	AVPRMGESLLS	DRLKRDNMD	LWSLLPSHP	FVKALYSGSLPLI	DKERF YAVQ	YNYLVGLVRSL	SIAASKSWSFEVA	RLAL SHASFL	STEMANYERLL	GELGLSL SEVLRE	110
solfatarious (29532) 1:	P aerophilum [Q8ZZb													
subtilis [25052] subtilis [25052] 1: HOTSEECERAALSHIGES VERY VOC GOT PETER VICE SYNTER YVICE SYNTER SYNTER SYNTER YVICE SYNTER YVICE SYNTER YVICE SYNTER SYNTER SYNTER YVICE SYNTER YVICE SYNTER YVICE SYNTER SYNTER SYNTE	s tokodaii [2719186	6BEM] 1	:	MLS	DELWGSISD	IYSSILSHP	FIKGLTDGTLEEI	EKEKYYILQ	FLYLREFSRAL	AILSAKAEKQDDS	VLFATHVSDVI	RVERSLHEYFI	SLWNVDVSRVK	100
autharcais [081071] 1:	s solfataricus [A99	9433] 1	:	MMENS	EKLWNFISD	IYSSILRHP	FILELVEGILSR	DKEKYYIIQ	YLYLREFSKAL	ALLAAKAEDEEQT	LIFTTHIQDA	KV KALHKFYI	KEFNLDIEDYE	102
cerease (glinz] 1	subtilis [P25052]	1 1		MKFS	EECRSAAAE	WEGSEVILP	FVQGIGDGTLPI	DREKYYVLQ	SYYLTHFAKVQ	SEGAAYAKDLYTT	GRMASHAQGTY	EA MALHREFA	ELLEISEEERKAF	103
halodurans (c83984) 1:	authracis [Q81UY]	1] 1	;	MKFC	DRLLETVQP	WEMSHNHP	FVVGMGDGTLEK	DKFQYYIIQ	YLYLLDYAKLY	AIGVVKATNPQVM	GKFAEQIDG1	NGEMTIHKQYA	KRLGISIEEIESA	103
iheyensis (2824433ACC) 1:	cereus [Q81HR2]	1	;	MKFC	DRLLETVQP	WEMSHNHP	FVVGMGDGTLEK	DKFQYYIIQ	YLYLLDYAKLY	AIGVVKATNPQVM	GKFAEQIDGI	NGEMTIHKQYA	KRLGISIEE IESA	103
perfringens (280530EER) 1:	halodurans [G8398	84] 1												
glutation [08401] 311:	iheyensis [282443	[33AGC] 1	:	MLFS	DSLREATKE	SWQLSLNHP	FVQKVVSGDLPL	ETEKNYILO	IYYLKHYGKVH	AFAAAHSDDFHIA	AKLAEKAKKT	EALTVHKEHA	EILQITEAEIENF	103
Iongum [086646] 1: HHTPDAARGPTITUMLPPEAGRIPHEANILYPECOYNOPTER LOND THE LON	perfringens [2805		;	MRFT	EKLLKEAMP	IWEKYLDHP	FIKE IGEGT LDK	NKEKE YLVQ	YLYLKEYAKVE	CAGVVKAKTMEEM	KFFYNSTKGT	EDUTAVHIEYL	KE <mark>FGI</mark> SELD <mark>A</mark> EKR	103
japonicus [095001] 1:	glutamicum [Q8NQF													
miliocida [09CLI00] 1:	longum [Q8G646]													
influenzae [P44659] 1:	japonicum [Q89DQ]	1] 1	:	MSFF	ERLKTEASV	EWRAYTE	FINGLADGSLPE	AAFRHYLVOL	YLELIEFARAY	ALAVYKSPRLAD	REAAAGLSAI	DVEMNLHVKLC	ADWGLSPTDLEQT	103
put La [08116] i:	multocida [Q9CLHC	10] 1	:	MSTT	QOLINNSGR	WLDYIHHP	FVQQLADGTLPK	ACTOHYLKO	YLELFOYNRAL	SLGIYKADNFAOM	KAAQDAIGALI	H-LIQLHIQYC	ESWG IDENTLFRT	102
jejuni [F81388] 1:	influenzae [P4465	59] 1	;	MI	EQLIQQAQP	YWQQYIEHE	FVQQLAKGTLPK	ACFORYLKO	YLYLFHYSRAF	ALGVEKAKNEAEM	ETPRKTLEILO	Q-LIQLHLNYC	REWGISEQE IFTT	100
pylori [G64680] 1:	putida [Q88I16]	1		MDIF	ERLKAAATP	WNRYVDHD	FVRQMGAGTLSEI	EAFRT YLVO	YLFLIQFARAW	ALAAYKSRRPADT	RAAQAGLSAT	D-TELHLRLC	ARWGLTQADIEAA	102
H0 H8 H9 H10 H10 H10 H10 H10 H10 H10 H10	jejuni [F81388]	1	:	MMLF	SNLIKENQK	IWNAYLHHD	FVKKLEDKSLKQ	ENFLFYLKO	YIYILNYAKCY	ARLALNSNTAKEL	REAMKEONYI	EGEMELHR-AL	LSLGINADELDIK	102
H8 H9 H11 H12 H13 - 10 120 130 140 150 160 170 180 190 200 210 abyssi 105: EPTLVNS MOPHATAKKAN IEGTLALPCINS ALTAEYHKOK II. ONP-IK THE GKV LSNETLA VGRUK LIDSGHS	pylori [G64680]	1		MQVS	QYLYONVQS	IWGDCISHP	FVQGIGRGTLERI	DKERFYIIQ	YLFILEYAKVE	ALGVVKACDEAVM	REFSNATODI	NNEMS THNHYT	RGLOTTOKELONA	103
<pre>horikoshi lot prives with the set of th</pre>		-				_H1	0							
horikoshi 105 EPT VNS MODEVILTAVGKU LEGULALIP CENS VAE LAE YHKDKID - GNP - IK THE GKV T.SNE T.LN VGR LAK LID SGBSOYDRLRE IF TIGSK ELLA VERAWRGGDVF		-0-	Н8 -	н	19			H	H12		H13	-	-	
abyssi 102: EPT VNSR YMDPMLSTAYKGRVVEGITALLOCHS X/E LAFYHKUKL D-GND-VPETNEE GRVTLSEELST VSRLWE IDSLGGDYEELE IF ITGSK ELAFHRMARGGDVF			and the second se	2.0	12112/J		H11		and the second	190			-	
abysi 102: EPT VNS YMDPMLSTAYKGRVVEGTALLOCINS KE LA EYHKUKLD-GNP VETIME GRV1LSEE LS VSRLER I IDSLGGDYERLER IT ITGSK ELAFMEMAWRGGDVF		110	and the second se	2.0	12112/J		H11		and the second	190			-	
furiosus 102: EPT WKS YMDFHLSTAYRGSVEGTALLP EPKSYRE TAINEEL REPHERKE END VKTWE REEV X.SEEX LKT VKR.RE.TD SQGGEYEKLKRTF TGSKPELAPWDAWRGGUF	horikoshii 105:	1	120 	130 	140 	150 I	H11	170 	180 	1	200 	210 	DVF	
accophilum 103: EPRRVNY SYGAYLKSTCALEGF YQCHAALD CONSINGE TAEHAGGK IR END - HWYKK ASV TISPE YRGLUER IRAUDSSCISAEELMPYKR ASLYELEFWOAAYEGH accophilum 101: -PSPTNL YTSYLLXVYSRPFYEGVSAVDCONTINE VGKELLKKGS - PNPLYKRVIET YGGEP EKGVRAVLDINNSPNLTKEEKEAVKRIPLITSME YMPNDA YRLEKFPFDF		EPTLVNSAY	120 I DFMLATAYKGNI	130 I IIEG <mark>LTALL</mark> PC	140 Fwsyaeiae	150 I YHKDKLR-D	H11 160 I NP-IKIYRENGK	170 I VVLSNEVLNI	180 I VGRLRK 11DSS	I GHSGYDR <mark>L</mark>	200 RR <mark>IF</mark> ITGSK <mark>F</mark> F	210 I LAFWEMAWRGG		
103: EPHRYINY SYMAYLKSTCILLEGFYQCHARLIP CUSYARE TAREHGGKURE FADP-OF UVYKKARSVILSPE YRGLUPERLRAVID SGLSAEELMPYKERSLYELEFWOAN YGCHSAEELMPYKERSLYELEFWOAN YGCHSAEELMPYKERSLYELEFWOAN YGCHSAEELMPYKERSLYELEFWOAN YGCHSAEELMPYKERSLYELEFWOAN YGCH	abyssi 102:	I EPTLVNSAM EPTLVNSAM	120 I DFMLATAYKGNI DFMLSTAYKGN	130 I IIEGLTALLPC VVEGLTALLPC	140 I FwsYae Iae FwsYae Iae	150 I YHKDKLR-D YHKDKLD-G	H11 160 I NP-IKITRENGKY NP-VEITRENGRY	170 I VYLSNEYLNI VYLSEEYLSI	180 VGRLRKIIDSSC VSRLREIIDSLO	 GHSGYDRL GGDYERL	200 RR <mark>IF</mark> ITGSKFF RE <mark>IF</mark> ITGSKFF	210 I ELAFWEMAWRGG ELAFWEMAWRGG	DVF	
solfatricus 103: -MSPTNLAYTSYLLAVANSRPFNEVISAVLPCATLAKVGKELLKQGS-KDKY QK IET KGEEYEKGVRAVLDTWSLKVSEEEFNKUK HERTASIYEYHENDSAYRLERGPFPTEKNKQ	abyssi 102: furiosus 102:	I : EPTLVNSAYM : EPTLVNSAYM : EPTLVNSAYM	120 I DFMLATAYKGN DFMLSTAYKGR DFMLSTAYRGS	130 LIEGLTALLPC VVEGLTALLPC VVEGLTALLPC	140 I PWSYAE IAE PWSYAE IAE FWSYAE IAE	150 I YHKDKLR-D YHKDKLD-G YHREKLE-R	H11 160 I NP-IKIYRENGK NP-VEIYRENGE NP-VKIYRENGE	170 I VYLSNEYLNI VYLSEEYLSI VYLSEEYLKI	180 VGRLRK IIDSS VSRLRE IIDSLA VKRLRE IIDSQ	 GHSGYDR <mark>L</mark> GGDYERL GGEYEKL	200 RR <mark>IF</mark> ITGSKFF RE <mark>IF</mark> ITGSKFF KRIFITGSKFF	210 I LAFWEMAWRGG LAFWEMAWRGG LAFWDMAWRGG	DVF	
solfaaricus 103:-MSPTNLAYTSYLLAVANSREPHEVISAVLECANLEKKUGKS-KDKYYXKIIETKGEEYEKGURAVLDIVNSLKVSEEEFNKUKIHERASITSYMEWOSAYRLEREPEPTEKUKGU	abyssi 102: furiosus 102: pernix 111:	I : EPTLVNSAYM : EPTLVNSAYM : EPTLVNSAYM : EPAPTNEAYV	120 I DFMLATAYKGN DFMLSTAYKGR DFMLSTAYKGR NFMLATCSTGT	130 liegitalipc vvegitalipc vvegitalipc alecmvslipc	140 I PWSYAE IAE PWSYAE IAE FWSYAE IAE YWSYRE IAI	150 I YHKDKLR-D YHKDKLD-G YHREKLE-R ANERLLR-E	H11 160 I NP-IKIYRENGKY NP-VEIYRENGKY NP-VKIYRENGE NS-VDLYRRNAS	170 I VYLSNEVLNI VYLSEEVLSI VYLSEEVLKI VYLSSEYGEI	180 VGRLRK LIDSS VSRLRE LIDSLO VKRLRE LIDSLO VKRLRE LIDSQ VEE YRRAVDRL	I GHSGYDRL GGDYERL GGEYEKL WAEEGGVYSRL	200 RRIFITGSKFF REIFITGSKFF KRIFITGSKFF KRIFRKATRYF	210 I ELAFWEMAWRGG ELAFWEMAWRGG ELAFWEMAWRGG YMFWEMAWRME	DVF DVF KWPV	
nubilis 104 kPSPTATYSTSRMYRSTSGRPETIALLDCHLYVE GEKLLHCD-CHEIXOK HCT KGOW RQOUEQUIREDELAENSTEEVE	abyssi 102: Furiosus 102: pernix 111: merophilum 103:	I : EPTLVNSAYM : EPTLVNSAYM : EPTLVNSAYM : EPAPTNEAYV : EPNRVNVSYM	120 I DFMLSTAYKGN DFMLSTAYKGN DFMLSTAYKGN NFMLSTAYKGN AYLKSTCALEGE	130 IIEGITALLPC VVEGITALLPC VVEGITALLPC ALECMVSLLPC FYQCMAALLPC	140 FWS VAE LAE FWS VAE LAE FWS VAE LAE YWS VAE LAE FWS VAE LAE	150 I YHKDKLR-D YHKDKLD-G YHREKLE-R ANERLLR-E RHGGKLR-E	H11 160 I NP-IKIYRENGK NP-VKIYRENGK NP-VKIYRENGK NP-VKIYRENGK NP-VKIYRENGK NP-VKIYRENGK	170 I VYLSNEYLN VYLSEEYLS VYLSEEYLK VYLSSEYGE VYLSPEYRG	180 J VGRLRK LIDSSG VSRLRE LIDSLG VKRLRE LIDSGG VEE YRRAVDRLD VERLRAVLDSSG	I GHSGYDRL GGDYERL GGEYEKL WAEEGGVYSRL GLSAEEL	200 RRIFITGSKFF REIFITGSKFF KRIFITGSKFF KRIFRKATRYF WPYFKEASLYT	210 I ELAFWEMAWRGG ELAFWEMAWRGG ELAFWEMAWRGG STAFWEMAWRME ELEFWQAAYEGH	DVF KWPV	
authracis 104: KPSÄKHLÄYTNYNESVSONGTLAELIAALLPCMES WEIGKRIN-DIP-GARDHEFTGE I LOGYSSEE YGNÜCINLIDLLNEMAVG-KSEKELDRLEE IF LYSSREYIPHOMSYRKEMAGFEEQEHTTVS	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101:	I : EPTLVNSAY : EPTLVNSAY : EPTLVNSAY : EPAPTNEAY : EPAPTNEAY : EPNRVNVSY : - PSPTNLLY	120 I DFMLATAYKGN DFMLSTAYKGN DFMLSTAYKGN NFMLATCSTGT AYLKSTCALEGE SYLLSVVYSRP	130 I IEGUTALLPC VVEGUTALLPC VVEGUTALLPC ALECHVSLLPC FYQCMAALLPC FYEGVSAVLPC	140 İ FWS YAE IAE FWS YAE IAE FWS YAE IAE FWS YAE IAE FWS YAE IAE YWI YAE YAE	150 I YHKDKLR-D YHKEKLE-R ANERLLR-E RHGGKLR-E ELLKKGS-P	H11 160 1 NP-IKIYRENGK NP-VEIYRENGK NP-VEIYRENGK NP-VKIYRENGK NP-VHYKKIAS	170 I VYLSNE YLN VYLSEE YLS VYLSEE YLK VYLSSE YGE TYGGEE YEKG	180 J VGRLRK LIDSSG VSRLRE LIDSSG VKRLRE LIDSGG VEEYRRAVDRL VEEYRRAVDRL VERLRAVLDSSG	 GHSGYDRL GGDYERL GGEYEKL WAEEGGVYSRL GLSAEEL NLTKEEKEAV	200 I RRIFITGSK IFITGSK KRIFITGSK KRIFRKATRYF WPYFKEASLYF KRIFRITSM	210 I LLAFWEMAWRGG LLAFWEMAWRGG LLAFWEMAWRGG YMFWEMAWRME ELEFWQAAYEGH YMFWEMAWRME	DVF DVF KWPV KFPFDF	
corons 104: PDSAKHL AYTHYNESSONGT LAELAALLPCAESNEE GKRULP OF CARDHEFT GE TIGG YSEE YGNUCTWI IDLAENAVG-KSEKELERIEE TELYSSREY (LAWDESYREEMARGFEGEHTTYS	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103:	I EPTLVNSAY EPTLVNSAY EPTLVNSAY EPAPTNEAY EPAPTNEAY EPNRVNVSY EPNRVNSAY EPNRVNSAY EPNRVNSAY EPNRVNSAY EPNLVNSAY EPNRVNSAY	120 J DFMLATAYKGN DFMLSTAYKGN DFMLSTAYKGS NFMLATCSTGT GAYLKSTCALEGF SYLLSVVYSRP SYLLAVAYSRP	130 I IEGITALLPC VVEGITALLPC VVEGITALLPC ALECMVSILPC FYQCMAALLPC FYEGVSAVLPC FNEVISAVLPC	140 I FWS VAE LAE FWS VAE LAE FWS VAE LAE FWS VAE LAE FWS VAE LAE FWS VAE LAE FWS VAE VGK	150 I YHKDKLR-D YHKCKLR-R ANERLLR-R ANERLLR-E ELLKKGS-P ELLLKQGS-K	H11 160 1 NP-IKIYRENGR NP-VKIYRENGR NS-VDLYRRNGS NP-VHVYKKAS NPIYRRIE 10KYYOK IE	170 I VYLSNE YLN VYLSEE YLN VYLSEE YLK VYLSSE YGE TYGGEE YEKG TYGGEE YEKG	180 VGRLRK I DSSG VSRLRE I DSLG VKRLRE I DSQ VEE VRAVDRID VERLRAVLDSSG VRAV ILI VNSFI VRAV LD IVNSLI	I GHSGYDRL GGDYERL GGEGEVYSRL GLSAEEL NLTKEEKEAV KVSEEEFNKM	200 I RRIFITGSK FITGSK KRIFITGSK KRIFRKATRYF WPYFKEASLYF KRHFRITSM FI	210 ELAFWEMAWRGG ELAFWEMAWRGG SYMFWDMAWRGG ELEFWQAAYRGG SYMFWDMAYRLE SYMFWDMAYRLE EYMFWDSAYRLE	DVF DVF KWPV KFPFDF- RFPFPTEKNKGV	
halodurans 104: EPARTILAYTSYMENYARGSILDLIAAVLPCTHSYE GVKLK-CD-CASDHEP'NCE IKLYASDE SKELADWIDGDEEKKC-LSSKEKKLETTEITTSREENE DWHAYNEG- 104: EPARTISHLYTSYMENYARSESGIAQUVANUPCNENADDGLTYKDAKP-KEKTYONILNYKSDWIQESTQEHIDLANTLAGAGAGAE-KEKLKTOFIIAKEYELAFWENYKG- 104: EYKSTIISYTSHQALAIGDLDE LAINTDCTNESSYGKYISKKYS-DKLQGHTYKP LOEVASDWIQESTQEHIDLANTLAGAGAEAE-KEKLKTOFIIAKEYELAFWENYKG- 104: EYKSTIISYTSHQALAIGDLDE LAINTDCTNESSYGKYISKKYS-DKLQGHTYKP LOEVASDWIQESTQEHIDLANTLAGAGAGAE-KEKLKTOFIIAKEYELAFWENYCFANISKYS-DKLQGHTYKP LOEVASDWIQESTQEHIDLANTLAGAAQAKAALSASVHEREFTDOTTRIGHTMGSS- longum 10: -PSHTMYTDFIIAKTYTEDYVCGVANUPCYNLXAEGILAAEFT-DTLDSNYKYSTER SSKAKKIETTEILSAERKKENTUDTYKASISELDPWNAYDEGEK- japonicum 121: ROSACARYTSYNUDVUVVVVCGVALVAEGILAAEFT-DTLDSNYKSVONTEE MSSKAKIENTIELLSAERKKENTUELVSVONTOVELYBEVTGVONEYBENSVONTUKSS japonicum 104: PPAREMLYTRYUDAGHGGLALKVAU DCVCKVAKGKALAAEFT-DTLDSNYKSVONTUEKE MSSKAKI IEHTEJITTERLSAERKENELVSVOTUGVNEYBENSVONTUKSS japonicum 101: RESACVNYTRYUDAGHGGLALKVAU APCVICKAYA TAKHVESGKSPANNDYAN II TEXGAEPVON AAKARAMEHLADLYATPAREAELTAITKEANDENTERMSSKAN TANDYTKYERVUDAGHGGLALLXVAU APCVICKAYA JAKHVESGK-SPANNDYAN II TEXGAEPVON AAKARAMEHLADLYATPAREAELTAITKENTE LANDGAGLDLA- 101: GESACCINYTRYUDAGHGGLAELYAAVTCCLKAAVITCHYDYLDKAAANTESGK-SPANNDYAN TANGKETUN ASKE COVAAARAHLADLYATPAREAELTAITKENTE ANDGAGLDLA- 101: GESACCINYTRYUDAGHGGLAELYAAVITCHCLKAAVITCHYD-RLAANTING AN TITY ASEE COVACTIVATYDVTUTTATRESAAVAGAUADCUCAAGAUXITCHYDYDRAAANTESALATTINGKENTUNG 111Cuciaa 101: GESACCINYTRYUDAGHGGLAELYAAVITCHCLKAAAVITCHCLKAAVITCHYDP-RLANNPYD TITYSENTIATTSS-VOAKYKKKKESTTPONSKENTUNGK	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103: subtilis 104:	I : EPTLVNSAY : EPTLVNSAY : EPTLVNSAY : EPAPTNEAY : EPNRVNSY : -PSPTNLAY : -MSPTNLAY : KPSPTAYSYI	120 I DFMLATAYKGN DFMLSTAYKGN MFMLATCSTGT MYLKSTCALEGF SYLLSVVYSRP SYLLSVVSRP SRMYRSVISGN	130 I IIEGITALLPO VVEGITALLPO ALECMVSILPO FYEGVSAVLPO FNEVISAVLPO FAETLAALLPO	140 I FWS VAE TAE FWS VAE TAE FWS VAE TAE FWS VAE TAE YWS VAE TAE YWT 24E VGE YWT 24E VGE YWT 24E VGE	150 I YHKDKLR-D YHKDKLD-G YHREKLE-R ANERLLR-E RHGCKLR-E ELLKQGS-K KLLHCDP-G	H11 160 INP-IKIYRENGKI NP-VKIYRENGE' INP-VKIYRENGE' NP-VHYYKKAS' NP-VHYYKKAS' NP-VHYKKIE' IPKYYQKIE' IPKYYQKIE'	170 I VYLSNE YLN VYLSEE YLS VYLSEE YLK VYLSSE YGE YYLSPE YRG TYGGEE YEKG TYGGEE YEKG	180 VGRLRK IIDSSG VSRLRE IIDSSG VEEYRAVDRIA VEEYRAVDRIA VERLRAVLDSSG VRAVLDIVNSLI VEEQINFPDELJ	 GHSGYDRL GGEYEKL WAEEGGVYSRL GLSAEEL NLTKEEKEAV KVSEEEFNKM AENSTEEVR-AKM	200 I RRIFITGSKFF KRIFITGSKFF KRIFRKATRYF WPYFEASLYF KRHFRITSMFF KIHFRTSYFF KENFVISSYFF	210 LAFWEMAWRGG LAFWEMAWRGG LAFWEMAWRGG LAFWEMAWRGE LEFWQAAYEGH YMFWDAAYRLE YMFWDSAYRLE YMFWDSAYRLE YMFWDSAYRLE	DVF DVF	GASRH
iheyensis 104: KPAPTAYNYTSHLYRASISGSLAQIVAAMLPC WALADD IGLTYKDAKP-KEKIYQNHLNTYGSDWTQESTQEHIDLLYILAEQAGEAE-KEKIKTOF IIAKEYELAFWEMSYTFETKLSEKQQI	abyssi 102: furiosus 102: pernix 1111: aerophilum 103: tokodaii 101: solfataricus 103: subtilis 104: authracis 104:	I EPTLVNSAY EPTLVNSAY EPTLVNSAY EPATNEAYU EPNRVNVSY EPNRVNVSY EPNRVNVSY EPNRVNVSY ENSTNLAYI KPSPTLAYI KPSAKNLAYI	120 I DFMLATAYKGN DFMLSTAYKGN DFMLSTAYKGS NFMLATCSTGT AYLKSTCALEGF SYLLSVVYSRP SYLLAVAYSRP SHMYRSVLSGNT NYMSVSQNTT	130 I IIEGLTALLPC VVEGLTALLPC VVEGLTALLPC ALECHVSILPC FYQCMAALLPC FYEGVSAVLPC FYEGVSAVLPC FAEILAALLPC LAELTAALLPC	140 I FWSYAE IAE FWSYAE IAE FWSYAE IAE YWSYAE IAE YW IYAE YAE YW IYAE YAE YWIYAE YGE YWSYWE IGK	150 I YHKDKLR-D YHKDKLD-G YHREKLE-R ANERLLR-E RHGGKLR-E ELLKGS-P ELLKQGS-K KLLHCDP-G RLN-DIP-G	H11 160 1 NP-VEIYREAGE NP-VEI NP-VE	170 I VYLSNEYLM VYLSEEYLS VYLSEYGE VYLSEYGE YYGEEYEKG TYGGEEYEKG TYGGEWERQ GYSSEEYGN	180 VGRLRK IIDSSG VSRLRE IIDSSG VKRLRE IIDSSG VEEYRAVDRLI VEERAAVLDSSG VRAVIDIVNSFJ VRAVIDIVNSFJ VRAVIDIVNSFJ VEEQINRFDELL GIWLIDILNEM	I GHSGYDRL GGDYERL MAEEGGVYSRL GLSAEEL NLTKEEKEAV KVSEEEFNKM AENSTEEVR-AKM AVG-KSEKELDRL	200 I RRIFITGSK FFITGSK KRIFITGSK FFI FTITGSK FFITGSK	210 LAFWEMAWRGG LAFWEMAWRGG LAFWEMAWRGG CLAFWEMAWRME CLEFWQAAYEGH YMFWDHAYRLE YMFWDHAYRLE CYGFWGHAYRKE YMFWDHAYRKE YLFWDHAYRKE	DVF DVF KWPV KFPFDF RFPFDFTEKNKGV- GWSDSAIKEVEEC MWGFEEQEHTTVS	GASRH
Performant Perform	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103: subtliis 104: authracis 104: cereus 104:	I : EPTLVNSAY : EPTLVNSAY : EPTLVNSAY : EPATNEAY : EPATNEAY : -PSPTNLAY : -PSPTNLAY : KPSPTAYSY : KPSAKNLAY : KPSAKNLAY	120 I DFMLATAYKGN DFMLSTAYKGSY NFMLATCSTGT AYLKSTCALEGG SYLLSVVSRP SYLLAVAYSRP SHLYKSVSQNGTI NYMSVSQNGTI	130 I IEGLTALLPC WVEGLTALLPC ALECMVSLIPC ALECMVSLIPC FYGCMAALLPC FYEGVSAVLPC FYEGVSAVLPC FAE ILAALLPC LAELTAALLPC	140 I FWSYAE IAE FWSYAE IAE FWSYAE IAE FWSYAE IAE FWSYAE IAE YW IDAEVGK YWIDAEVGK YWIDAEVGK YWIDAEVGK YWIYE IGK	HINDERSE	H11 160 HP-IKIYRE GR' HP-VEIYRE GR' HNP-VKIYRE GR' HNP-VKIYRE GR' HP-HVYKKAS' HP-HVYKKAS' HPIYKKIE' HPIYKKIE' HPIYKKIE' HPIFGEIG	170 I VYLSNEYLN VYLSEEYLS VYLSEEYLS VYLSEEYLK VYLSPEYRG TYGGEEYRK TYGGEEYRK TYGGEEYRK GYSSEEYGN GYSSEEYGN	180 VGRLRK IIDSSA VSRLRE IIDSSA VEEYRRAVDRIA VEEYRRAVDRIA VERIRAVIDSS VRAVILJVNSFI VRAVILJVNSFI CIWLIDLINEM CIWLIDLINEM	I GHSGYDRI GGPYERI WAEEGGVYSRI GLSAEEI LI.TKEKERV KVSEEEFNKM AENSTEVR-AKM AVG-KSEKELDRI AVG-KSEKELERI	200 I RRIFITGSKF REIFITGSKF KRIFITGSKF KRIFRATRYF KRIFRASLYF KRHFRITSMF KIHFRTASIYF KENFVISSYF EEIFLYSSRF EEIFLYSSRF	210 I LAFWEMAWRGG LLAFWDMAWRGG YMFWDMAWRME LLEFWQAAYEGH YMFWDMAWRLE YMFWDMAYRLE YMFWDMSYRLE YDFWDMSYRKE	DVF DVF KWPV KFPFDF RFPFPTEKNKGV GWSDSAIKEVEEC MWGFEEQEHTTVS MWGFEEQEHTTVS	GASRE
110:-PSHTTMYTDFLIART YTEDYVCGVAAVLPCWLXAE IGIMLAEQNH-DEHP XKOLDT YSGEE IAGTRAA TARLEKALEN-AGAEQRVDAARAFLSASVHEREFDQATRHGWTMVGSS	abyssi 102: furiosus 102: pernix 111: nerophilum 103: cokodaii 101: solfataricus 103: subtilis 104: authracis 104: alodurans 104:	I EPTLVNSAY EPTLVNSAY EPTLVNSAY EPAPTNEAYU EPAPTNEAYU -PSPTNLAYI KPSPTNLAYI KPSAKNLAYI EPAATTLAYI	120 I DFMLSTAYKGR DFMLSTAYKGR DFMLSTAYKGR DFMLSTAYKGR STLLSVYSRP SYLLAVAYSRP SYLLAVAYSRP SYLLAVAYSRP INYMSVSQNGTI SYMLNVARGSI	130 I IEGITALIPO VVEGITALIPO ALECMVSLIPO FYGVSANALIPO FYGVSAVLPO FYEGVSAVLPO FAE ILAALIPO LAELITAALIPO LAELITAALIPO LAELITAALIPO LAELITAALIPO	140 I FWSYAE TAE FWSYAE TAE FWSYAE TAE YWSYRE TAI FWSYAE TAE YWTDAEVGK YWTDAEVGK YWTYYE TGK TWSYYE TGK TWSYYE TGV	150 I YHKDKID-D YHKDKID-G RHCGKIR-E RHGCKIR-E ELLKQCS-K KLLHCDP-G RLM-DIP-G RLM-DIP-G RLM-DIP-G	H11 160 I	170 I VILSNE VLNI VILSEE VLSI VILSEE VLSI VILSEE VLSI VILSEE VLSI VILSEE VLSI VILSEE VLSI VILSEE VLSI VISEE VLSI CISSEE VLSI LASDE KE	180 VGRIRR I DSSA VKRIRE I DSSA VKRIRE I DSSA VEE YRRAVDRIL VRAVIDISSS VRAVIDIVNSFI VRAVIDIVNSFI CIWLIDIINEM CIWLIDIINEM CIWLIDIINEM ADWLIDMIDEEJ	I GHSGYDRI GGDYERI GGEYEKI WAEEGGVYSRI GLKEEKEAV KVSEEEFNKM AENSTEEVR-AKM ANG-KSEKELDRI AVG-KSEKELERI AVG-KSEKELERI AKG-LSSKEKAKI	200 I RRIFITGSKE KRIFITGSKE KRIFRKATRYE MPYPKEASLY KRIFRITSME KRIFRITSME KENFVISSYY EIFLYSSRE EEIFLYSSRE TIFLTTSRE	210 I LAW WEMAWRGG LAF WEMAWRGG STAF WEMAWRGG STAF WOM STREEH STAF WOM STREE STAF WOM STREE STAF WOM STREE STAF WOM STREE STAF WOM STREE STAF WOM STREE	DVF- DVF- KMPV	GASRH
longum 121: RQSAFARAYTSNILSTAKONPLUOVI, VAVLPCARVKADVOQRLAAFT - DTLDSNP KSS WOMYTTEE NSSSAWL IEH IEQLTEH-LSAERKDELVEFVTGUQNEYHENSSSAYDAQYTWKPEWDV	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103: subtilis 104: authracis 104: cereus 104: halodurans 104: inbeyensis 104:	I EPTLVNS AY EPTLVNS AY EPTLVNS AY EPAPTNE AY EPAPTNE AY EPAPTNE AY EPAPTNE AY SKILAY KESPTAYSYI KESPAKILAYI KESAKILAYI EPARTILAYI	120 I I DFMLSTAYKGR DFMLSTAYKGR DFMLSTAYKGR DFMLSTAYKGSV NFMLSTACSTGT SYLLSVVSRP SSHJYRSVLSGN SYLLAVASSP SSHJYRSVSQNGT SYMLNVAQKGS SSHJYRASISGSI	130 ITEGITALLPC VVEGITALLPC VVEGITALLPC EVQCHAALLPC EVQCHAALLPC EVQCHAALLPC EVQCHAALLPC EVGCHAALLPC ENE ILAALLPC LAELTAALLPC LAELTAALLPC LAELTAALLPC LLDLIAAVLPC LLDLIAAVLPC	140 I IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	150 I YHKDKLR-D YHKDKLD-G YHREKLE-R ANERLLR-E RHGGKLR-E ELLKKGS-P ELLKKGS-P ELLKQGS-K KLLHCDP-G RLN-DIP-G KLK-GIP-G KLK-GIP-G	() H11 (60 () ()	170 I VILSNE VLN VILSEE VLS VILSEE VLS VILSEE VEK TVGEE VEK TVGEE VEK TVGEE VEK TVGEE VEK GVSSEE VGN GVSSEE VGN LVASDE KE TVGSDW-QES	180 VGRIRKTIDSS VKRIRETIDSL VKRIRETIDSQ VEEYRAVDRI VEEYRAVDRI VEEQTNEPDLI VEEQTNEPDLI CTWIDDLINEM ADWIIQMIDES	I GHSGYDRL GGEYEKL WAEEGGVYSRL GLSAEEL NLTKEEKEAV KVSEEEFNKM AENSTEEVR-AKG AVG-KSEKELBRL AVG-KSEKELBRL AKG-LSSKEKAKL AKG-LSSKEKAKL	200 I RRIFITGSK FRIFITGSK RRIFITGSK RRIFITGSK RRIFI RATSK RRIFITS REAL FRIFTS REAL FRIFTS REAL	210 IL AU WEMANGG IL AU WEMANGG IL AU WEMANGG IL AU WEMANGG IL AU WOMANGG IL AU WOMANG IL AU WEMANG IL AU WOMANG IL AU WEMANG IL AU WEM	DVF	GASRH
japonicum 104: PPARPHLTYTRYULDAGHRGDLLALKVALAPCVICYAE TATELAARDD-ADARTHARKVI TAE KAGAPYGE VAAKARANGEHLADLYATPAREAELITAITKEATRIEADOWEMARAGQUVQ	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103: subtilis 104: authracis 104: authracis 104: halodurans 104: halodurans 104: perfringens 104:	I EPTLVNS AY EPTLVNS AY EPTLVNS AY EPTLVNS AY EPTLVNS AY EPTLLVNS AY EPTLL SPSPTLLY KPSAKILAY ERSAKILAY ERSAKILAY ERAPTAYAY EVSATISY	120 I DEFULSTAYKGN DEFULSTAYKGN DEFULSTAYKGS NEMIATCSTGT SYLLAVATKSP SIMYRSVISGNGT SYLLAVASRP SIMYRSVISGNGT SYMLAVAQRGSI SIMYRASIGGSI SIMYRASIGGSI SIMYRASIGGSI	130 I IIEGITALLPC VVEGITALLPC VVEGITALLPC ALECAWSLLPC FYQCWAALLPC FYQCWAALLPC FYQCWAAVLPC FAE IIAALLPC LAELIAALLPC LAELIAALLPC LAQIVAAMLPC LAQIVAAMLPC DE IATATLPC	140 I PWSYAEIAE FWSYAEIAE FWSYAEIAE YWSYREIAI FWSYREIAI FWSYREIG WWIYAEVGK YWIDHKVGK YWIDHKVGK YWIDHKVGK TWSYEIGK TWSYEIGW TWIYADIGI TWSYEYGK	150 I YHKDKIR-D YHKDKIE-B ANERLIR-E RHGCKIR-E ELLKKGS-P ELLKKGS-P ELLKQS-K KLH-DIP-G RLM-DIP-G KLK-GDP-G TYKDAKP-K YISKKYS-D	H11 160 Imp-ikitereket ims-vbitereket ims-vbitere	170 I VILSEYLSI VILSEYLSI VILSEYLSI VILSEYGE TYGGEYEKSI TYGGEYEKSI GISSEEYGNI LIASDERSE TYGSDERSE TYGSDERSE TYGSDERSE	180 VGRLRK I IDSK VSRLRE I IDSK VKRLRE I IDSK VKRLRE I IDSK VEE YRRAVDRI VEE YRRAVDSS VRAVLD IVNSL VEE Q INFDEL CIVIT IDLINEM ADWL IQMIDEE: TQEMIDI ANTL IDSKLVYVDKK	I GHSGYDRL GGDYERL GGSAEEL NLTSAEEL NLTKEEKEAV KYSEEFNKM AENSTEEVR-AKM ANG-KSEKELERL AKG-LSSKEKAKL AKG-LSSKEKAKL AEQAGEAE-KEKT SN-LSEEKQKRL	200 I RRIFITGSK F RRIFITGSK F RRIFIRGSK F RRIFIRATRY MYYKRASLY KIHPRTASIYF KIHPRTASIYF KENFVISSYYF EEIFLYSSR F EEIFLYSSR F EEIFLYSSR F EEIFLYSSR F EFIFLTSR F KTOFILAKEY	210 ILDI WEMAWRGG ILDI WEMAWRGG ILDI WEMAWRGE ILDI WEMAWRGE ILDI WEMAWRGE ILDI WEMAYREE INDI WEMAYREE ILDI WEMAYDEE ILDI WEMAYDEE	DVF	GASRH
multocida 103: EESAACVAYTRYVLDAGHTGGLAELYTALAPCATGYAVTAKHIVESGKSPANNPYOANIDTYSGEEFONAAQNATATLDALCAD-RSEAQLAKLQQIFNTATRMESAFNQMGLDLS	abyssi 102; furiosus 102; pernix 111; aerophilum 103; tokodaii 101; solfataricus 103; subtilis 104; authracis 104; authracis 104; halodurans 104; iheyensis 104; gerfringens 104; jultamicum 410; 104;	I EPTLVNSAY EPTLVNSAY EPTLVNSAY EPTLVNSAY EPRRVNSY PSPTNLY EPSPTNLY KBSPTNLY KBSPTAYSY KBSAKILAY EFSAKILAY EFSAKILAY EFSATTLAY EFXSTTISY EFXSTTISY	120 I DEFULATAYGEN DEFULSTAYGEN DEFULSTAYGEN DEFULSTAYGEN SYLLSYVSPE SRIYESVSGHGTI NYMESVSGHGTI NYMESVSGHGTI NYMESVSGHGTI SYLLSYAQAGGSS SYLLAYAQAGGS SYLLYAQAGGS SYLQATATGDI DEFULATETTED	130 I TEGITALLPC WYEGITALLPC WYEGITALLPC UYEGITALLPC TALECHYSLEPC FYGCHARLPC FYEGVSAVLPC FYEGVSAVLPC FYEGVSAVLPC LAELTAALLPC LAELTAALLPC LAELTAALLPC LAELTAALLPC LAELTAALPC LAELTAALPC UDE TAATAVLPC	140 I FWSYAEIAE FWSYAEIAE FWSYAEIAE YWSYREIAI FWSYAEIAE YWIYAEVGE YWIYAEVGE HWSYYEIGK TWSYYEIGK TWSYSIGK TWSYSYIGK YWIYAEIGI	150 1 1 1 1 1 1 1 1 1 1 1 1 1	(H11 160 I I MIP - IX1 YREBGR WIP - VE1 YREBGR WIP - VE1 YREBGR WIP - WE1 YREBGR WIP - WIY YREBGR WIP - UVY YREBGR WIP IVY KIG ARDHEFT CEI IQ =K1Y QKHIET	170 I VILSNE ILM VILSEE ILS VILSEE ILS VILSEE VEK VILSEE VEK TYGGEV FKG GYSSEE VGM LYASDE FKE TYGSDW QES E YASDG FKE TYSGEF ING	180 VGRLRK TIDSK VSRLRE TIDSK VSRLRE TIDSK VKRLRE TIDSK VKRLRE TIDSK VFR VILVSK VFR VILV	I GHSGYDRL GGFYEKJ GGFYEKJ GGSYEE NILTSAEE NILTKEEKEAW KVSEEEFNKG ARG-ISSEKEAKAKJ ANG-KSEKELDRL ANG-KSEKELDRL ANG-KSEKELDRL CSN-LSEEKQKRI CSN-LSEEKQKRI LEN-GAEQRVDA	200 I RRIFITGSKF REIFITGSKF KRIFITGSKF KRIFITGSKF MYPKEASLY KRIFRITSMF KRIFTISSY EEIFLYSSRF EEIFLYSSRF EEIFLYSSRF EIFLYSSRF TIFLTTSRH KTOFITAKEY IDIFKRASLY	210 ILW WE HAWROG ILW WE HAWROG ILW WE HAWROG ILW WE HAWROG ILW WE HAWRON ILW WE HAWRON YOL WORS AYRLE ILW WE HAWRON ILW WE HAWR	DVF	GASRH
influenzae 101: QESAACIAYTRYLLDCGHTGSLAELYAAVTPCALGXQVARYITQHYP-RLP-NNPYQTHIDTYASEEFQQAAQETVDFLFALCKP-LNPSQLAETQQTFTTATRMEIAFNQHGLDLA- putida 103: PENQATAYTRYYLDCGAAGDLELHVALAYCVIGXAE GGRILAERIGC-DLSNNPYREH IGEXAGGGYQCVAAARKHIDELAARSHTEQRFAELAGTUGQASI EADTWQGLDLAT- jojuni 103: DESLAVIYTRYYLDCGHAGDLAGLAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103: subtilis 104: authracis 104: cereus 104: chedurans 104: perfringens 104: perfringens 104: glutamicum 410: longum 121:	I EPTLVNS AY EPTLVNS AY EPTLVNS AY EPTLVNS AY EPTLVNS AY EPTLVNS Y EPSPTNLJYT KPSPTAYSYT KPSPTAYSYT EPATTLAY EPATTLAY EPATTLAY EPATTLAY EPATTLAY EPATTLAY EPATTLAY ESPATTAY F	120 I DEFULSTANCEN DEFULSTANCEN DEFULSTANCEN INFULATCSTGT INVIKSTCALEGI SYLLAVASRP SHLYRSVLSGNI NYMSVSQNGTT SYLLAVASRP SHLYRSSLSGSI SYMQAIALITGD DE LIARTTED DE LIARTTED	130 I EGITALLPC VVEGITALLPC VVEGITALLPC VVEGITALLPC VVEGITALLPC FYQCMAALLPC FYQCMAALPC FYQCMAALPC FAEILAALLPC LAEILAALLPC LAEILAALLPC LDCITAAVLPC LDCITAAVLPC VYCGVAAVLPC	140 I PWSYAEIAE FWSYAEIAE FWSYAEIAE FWSYAEIAE FWSYAEIAE FWSYAEIAE FWSYAEIGK FWSYAEIGK TWSYAEIGK TWSYAEIGK FWSYAEIGK FWSYAEIGK FWSYAEIGK FWSYAEIGK FWSYAEIGK FWSYAEIGK FWSYAEIGA FWSYAEIGA FWSYAEIGA FWSYAEIGA FWSYAEIGA FWSYAEIGA	150 I VHKDKIR-D VHKDKIR-D VHKDKIR-G VHKDKIR-G VHKDKIR-E RHGGKIR-E ELLKGGS-K KLIK-GP-G RLIN-DIP-G KLIK-GP-G TYKDARE-K YISKKYS-D MLAEQNH-D RLAAEFT-D	() H11 (60 () (170 I VILSEE YLS VILSEE YLS VILSEE YLS VILSEE YLS VILSEE YLS VILSEE YKS GISSEE YKS GISSEE YKS GISSEE YKS UASDE KE VASDE KE VASDE KE VASDE AK TYSGEE IAK TYSGEE IAK	180 VGRLRK TIDSS VSRLRE TIDSL VKRLRE TIDSQ VER YRRAVDRL VER YRRAVDRL VER CONFEDEL CIVET DLINEL CIVET DLINEL CIVET DLINEL ADMITCH.DESS TOPMTDLINEL TDEWLVYVDKK TRAALAREKA	GHSGYDRL GGDYERL GGEYERL WAEEGGVYSRL GLSAEEL NLTKEEKEAW AENSTEEVR-ANO ANG-KSEKELERL ANG-KSEKELERL ANG-KSEKELERL CSN-LSEEKQKRL CSN-LSEEKQKRL CSN-LSEEKQKRL CSN-LSEEKQKRL	200 I RRIFITGSK REIFITGSK REIFITGSK REIFITGSK REIFITGSK REIFITGSK REIFITGSK REIFITGSK REIFITSR REIFITS	210 ILU WEMAWRGG LLA WEMAWRGG LLA WEMAWRGG LLA WEMAWRGE ILE PWQAAYRGH YBE WOANYRLE YDE WOMSYRKE YDE WOMSYRKE INE WOANYRES ILD WMSY YTEE ILD FWMSY YTEE ILD FWMSY YTEE ILD FWMSA YTEE ILD FWMSA YTEE ILD FWMSA YTEE	DVF- DVF- KFPFDF- RFPFPTEKNKGV- GWSDSAIKEVEECC MWGFEEQEHTTVS- MWGFEEQEHTTVS- MWGFEEQEHTTVS- MWNFWGFEQEHTTVS- MWNFWGSS- WIMVGSS- YIWKESWDV-	GASRH
put ida 103: pehoatväytryvidcoaacdilelhvalapovicyaeigrilaerigdlsnhpyren igeyagegyocyaaaarkhidelaarsmteorfaelagifgoass <mark>b</mark> eadfnomgidgty	abyssi 102; furiosus 102; pernix 111; aerophilum 103; tokodaii 101; solfataricus 103; subtilis 104; authracis 104; authracis 104; halodurans 104; perfringens 104; japanicum 410; longum 121; japonicum 104;	I EPTLVNS AY E EPTLVNS AY E EPTLVNS AY E EPAPTNE AY E EPAPTNE AY E EPAPTNE AY E EPAPTNE AY KPSPTNLAY KPSPTNLAY KPSPTNLAY E EPAATTLAY E EPAATTLAY E EYKSTTISY E OSAFARATI ROSAFARATI E PPAAEMLAY	120 I DEFLATAYKGN DEFLISTAYKGN DEFLISTAYKGN DEFLISTAYKGN DEFLISTAYKGN SYLLSVVSRP SYLLSVVSRP SYLLSVVSRP SYLLSVANSVSRP SYLLSVANSVS SYLLSVANKSS SYLDVANKSS SY	130 I EGITALLPC WEGITALLPC WEGITALLPC WEGITALLPC EGITALLPC FYEGYAALLPC FYEGYSAVLPC ENEVISAVLPC LACITAALLPC LACITAALLPC LACITAALLPC LACITAALLPC LACITAALLPC LACITAALPC LACITAALPC LACITAALPC LACITAALPC	140 WS VAE TAE WS VE TAE WS VE TGK WS VYE TGK WS	150 150 150 1 150 1 150 1 150 150	160 1 160 1 10 10 10 10 10 10 10 10 10	170 I VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS ISSEET GEGEVEKO TIGGEETEKO GYSSEETGN LIASDEFKE TIGSEETLA CASEFIA TISGEETLA MORTEEWSS	180 VGRLRK TIDSSA VSRLRETIDSJA VKRLRETDSJA VERTRAVDORJ VERTRAVDORJ VERTRAVDDSJ VRAVINI VNSFJ VRAVINI VNSFJ CIMIDDIJNEJ CIMIDDIJNEJ TOEMIJVYOKK STOMIDDIJNEJ TOEMIJVYOKK STAAIRAEKAI SAMIJENEDEJ	I GHSGYDRL GGFYERL GGFYERL JLSAEEL HLTKEEKEAW KVSEEEFHKW ARO-KSEKELDRL ANG-KSEKELDRL ANG-SKEKELDRL ANG-SKEKELAKL CSN-LSEEKKKRL LEN-AGAEQRVDA TEH-LSAERKDEL	200 I RR IF ITGSK B RR IF ITGSK B KR IF ITGSK F KR IF ITGSK F KR IF RTSK F F V SKASLY KR IF RTSK F KR IF	210 IL W HE MANRGG IL DU WE HANRGG IL DU WE HANRGG IL DU WE HANRGE IL DU WE HANRGE IL DU WE HANRE IL DU WE HANRE IL DU WE HAN YEE IL DU WE HAN YEE IL DU WE HAN YEE IL DU WE HAN YEE IRE FF DO AT RIG YEE WE A YE HAN YEE IRE FF DO AT RIG YEE WE A YE HAN YEE IRE FF DO AT RIG YEE WE A YE HAN YEE IN DU WE HANNARG	DVF DVF KEPFDF GWSDSAIKEVEECK MWGFEEQEHTTVS MWGFEEQEHTTVS TWLSEKQQI EK WTMVGSS YTWKPENDV QLVQ	GASRH
jejuni 103: des <mark>tvnla</mark> ysry <mark>m</mark> .svg <mark>e</mark> ngd <mark>fldm.valsacaigyakigaeiinrlknenlkdhpykeniltygsenfoneakefedfvnsytss-vgaokfoklseifhtvtrlevafwehslrmelnl</mark>	abyssi 102; furiosus 102; pernix 111; aerophilum 103; tokodaii 101; solfataricus 103; subtilis 104; authracis 104; halodurans 104; perfringens 104; glutamicum 410; longum 121; japonicum 104; multocida 103;	I EPTLVNS MY EPTLVNS MY EPTLVNS MY EPTLVNS MY EPSPNLD EPNRUNVS Y EPSPNLD MSPTNS KSSKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ENSAKILMY ESTLMY ESTLMY ESTLMY ESTLMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY ESTLMY MICHAELMY MICHAELMY MICHAELMY ESTLMY MICHAELMY	120 I DEFIGATAPKGN DEFIGSTAPKGN DEFIGSTAPKGSV SYLLSVVSRP SYLLSVVSRP SHYRSVSGNGT NYASVSQNGT NYASVSQNGT SYGLAVASSS SYLQAXSTGS SYLQATATGD DE LARTYED SNLSTAPK	130 I EGLTALLPC VVEGITALLPC VVEGITALLPC VVEGITALLPC VVEGITALLPC FYCGVSAVLPC FYCGVSAVLPC FYCGVSAVLPC LAELTAALLPC LAELTAALLPC LDCLTAAVLPC LQCVAAMDC LDCLTAATLPC LVDVLVAVLPC LAELTAALPC	140 I I I I I I I I I I I I I I I I I I I	150 150 YIKDKGP-D YIKDKGP-G YIKEKLE-R RHGCKIR-E ELLKGGS-F ELLKGGS-F ELLKGGS-K KLLHCDP-G KLLN-DIP-G KLM-GIP-G KLM-GIP-G TYGDKFY-D RLARFT-D RLARFT-D RLARFD-A	() H11 (60 () (170 I VISRETLS VISEETLS VISEETLS VISEETLS VISEETLS VISEETCS VISEE VISEETCS VISEETCS VISEETCS VISEETCS VISEE VISE VIS	180 VGRIRK TIDSS VSRIRE TIDSS VSRIRE TIDSS VKRIRE TIDSS VKRIRE TIDSS VRAVINIVSS VRAVINIVSS VECYRRAVDI VECORPOLIANS CIWI TOLINEM ADMITOLOGOEL TOLINIA TOLINEM SAMI TENTATALEXA SAMI TENTATALEXA	GHSGYDRL GGDYERL GGDYERL GGSAEEL NLTSAEEL NLTSAEEL NLTSAEEL NLTSKEKAEL RLG-KSEKELDEL NVG-KSEKELDEL NVG-KSEKELDEL NVG-KSEKELDEL NVG-KSEKELDEL SCN-LSSEKAKL ELN-AGAEQRUDA TEH-LSAERKOEL DULATT AREAEL	200 I RRIFITGSKD RRIFITGSKD RRIFITGSKD RRIFITGSKD RRIFITGSKD RRIFRITGSKD RRIFRITGSKD HYDRIGSKD HYDRIGSKD RRIFRIGSKD HYDRIF HYDRIGS	210 ILDI WEMAWRGG LLDI WEMAWRGG LLDI WEMAWRGG LLDI WEMAWRGG ILDI WOANREE ILDI WOANREE ILDI WOANYEGH INDI WOANYEGH INDI WOANYEGH INDI WOANYEG LLDI WOANYEG LLDI WOANYEG ILDI WAANYEG ILDI WA	DVF- DVF- KFPFDF- RFPFPTEKNKGV- GWSDSAIKEVEEC MWGFEECHTTVS- MWGFEECHTTVS- MWNYNG- TWLSEKQQI- EK- WINVGSS- YTWKPEWDV- QLVQ-	GASRH
	abyssi 102; furiosus 102; pernix 111; aerophilum 103; tokodaii 101; solfataricus 104; authracis 104; authracis 104; authracis 104; perfringens 104; glutamicum 100; longum 121; japonicum 103; influenzae 101;	I EPTEVNS MY EFTEVNS MY EFTEVNS MY EPTEVNS MY EPPRUNS MY EPPRUNS MY EPPRUNS MY EPPRUNS MY EPPRUNS MY EPPRUNS MY EPPRUNS MY KSSAKHAY EFSAKHAY EFSAKHAY EFSAKHAY EPPAEMLAY EPSACHAY	120 I DEFULSTAYGR DEFULSTAYGR DEFULSTAYGR MENTATOSTGT SYLLAVSTCHLEG SYLLSVSUSGT SYLLAVSUSGT SHIYRSVLSGHT SYLLAVSUGGT SYLLAVSUGGT SYLLAVSUGGT SYLLAVSUGGT SYLLAVSUGGT SYLLAVSUGGT SYLLAGTGGT SYLLAGTGGT	130 I EGITALLPC VVEGITALLPC VVEGITALLPC VVEGITALLPC VVEGITALLPC FYGCMAALPC FYGCMAALPC FAEILAALLPC FAEILAALLPC FAEILAALLPC LAELTAALPC LIDLTAAVLPC LIDLTAAVLPC LIDLTAAVLPC LIDLAKVALAPC LAELYAAVTPC	140 1 WS VAE LAE WS VAE LAE WS VAE LAE WS VAE LAE WS VAE LAE WH DREVCK WH DYEVCE WH DYEVCE WH DYEVCE WS WE ICK WH YAE ICH WS YE ICK WH YAE ICH WS YE ICK WH YAE ICH TWS IS YICK WH YAE ICH ACK YAV TAK ACK YAV TAK	150 150 YHIKDKID-G YHIKDKID-G YHIREKLE-R RHIGGKIR-E RHIGGKIR-E RHIGGKIR-E RHIGGKIR-E RHIN-DIP-G RLIN-DIP-G KLIX-GIP-G KLIX-GIP-G KLIX-GIP-G RLIADEPT-D RLIADET-D RLIADET-D RLIADET-D RLIADET-D RLIADET-D RLIADET-D RLIADET-D RLIADET-D RLIA	(170 VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS VILSEETLS SEETS SEETS SEETS VILSEETLS VILS	180 VGRLRK I IDSIA VSRLRE I IDSIA VKRLRE IDSQ VELYRRAVDRI VELYRRAVDRI VEEYRRAVDRI VEEQINFDEL CIVITOLIMEM ADMITOLIMEL TODENLYVVDKK TRAALARLEKA SAMITEH IEQUI AAKARAMEHLI AQUITVDE UTAAA	GHSGYDRL GGDYERL GGEYERL GGEYERL LISAEEL NITKEEKEAW AENSTEEVR-ANG AENSTEEVR-ANG AENSTEEVR-ANG ANG-KSEKELDEL ARG-LSSKEALT, EN-LSEEKQKRL LEN-AGAEQRVDA CAG-RENDA EN-LSEEKQKRL CAD-RESPQLARC CAD-RESPQLARC	200 I RRIFITGSK REIFITGSK REIFITGSK RRIF RRIFR RRI	210 ILD WEMAWRGG ILD WEMAWRGG ILD WEMAWRGG ILD WEMAWRGE ILD WEMAWRGE ILD WEMAWRGE ILD WAMRGE YE WEMAYREE YE WEMAYREE ILD WEMAYNEE ILD WAMRYDEG ILD WAMRAYDEG ILD WAMRAGELDLA SD WOGLDLA	DVF DVF KFPFDF KFPFDF KFPFDFENKGV MWGFEEQEHTTVS MWGFEEQEHTTVS TWLSEKQQI EK WINVRGS YTWKPEWDV QLVQ	GASRH
	abyssi 102: furiosus 102: pernix 111: aerophilum 103: tokodaii 101: solfataricus 103: subtraisi 104: authracis 104: authracis 104: haloduras 104: jutamicum 100: longum 104: japonicum 104: multocida 103: influenzae 101:	I EPTLUNS AM EPTLUNS AM ETTLUNS A	120 I DE HLATA AYGEN DE HLSTA AYGEN DE HLSTA AYGEN HENDER STATHESS SYLLSVA SYRP SRHYRSV SGHET NYHESS SGHET NYHESS SGHET NYHESS SGHET SHL YRAS LSGS SHL YRAS	130 IFEGITALLPE VVEGITALLPE VVEGITALLPC VVEGITALLPC VVEGITALLPC VEGITALLPC FYGVSAVLPC FYGVSAVLPC FYGVSAVLPC FYGVSAVLPC LAELTAALLPC LAELTAALLPC LAELTAALLPC LAELTAALPC LAELYAAVTPC LAELYAAVTPC LAELYAAVTPC LAELYAAVTPC LAELYAAVTPC LAELYAAVTPC	140 1 15 YAE LAE TWS YAE LAE TWS YAE TAE TWS YAE TAE TWS YAE TAE TWS YAE TAE TWS YEE TAE	150 150 YHKDKLD-G YHKDKLD-G YHREKLE-R RHIGCKIR-E ELLKQGS-K KLLHCDP-G ELLKQGS-K KLLHCDP-G ELLKQGS-K KLLHCDP-G TYKDAKP-K YISKKYS-D TYKDAKP-K YISKKYS-D RLAAEP-A-H HIVESGK YITGIMP-R YITGIMP-R	(H11 160 I 1 1 1	170 I LSNETH V LSEE LS V LSEE LS V LSEE ELK V LSEE EK T GGEE BK G SSEE GN L ASDE KC S SSEE GN L ASDE KC T GGEU BK T GGEU	180 VGRLRKTIDSS VSRLREIDSSL VSRLREIDSSL VELYRAVDRL VELYRAVDRL VELYRAVDRL VELYRAVDR VELYRAVDR SVRAVINVSF SVRAVINVSF SVRAVINVSF SVRAVINVSF SVRAVINVSF SVRAVINVSF SVRAVINVSF SVRAVINSF SVRAVI	GHS GYDRL GG GYDRL GG SYERL GG SYERL GL SYERL NILT SKERL NILT SKERL NILT SKERL KUS EEFENR ARG -LSSKEKLDRL NYG SSKELDRL NYG SSKELDR NYG SSKELDR	200 RRIF ITGSK D RRIF ITGSK D RRIF ITGSK D RRIF ITGSK D RRIF ATGSK D RRIF ATGSK D RRIF ATGSK D RRIF ATGSK D RRIF ATGSK D RRIF ATGSK D RRAF LSASVH VE VF VTCVOH RAF LSASVH VE VF VTCVOH ATTFRATRA QQ IF ATARM G IF GASS D	210 LLW WE HANREG LLW WE HANREG LLW WE HANREG LLW WE HANREG LLW WE HANRE WE WE HANRE WE WE HANRE WE WE HANRE LLW WE HANRE LLW WE HANRE LLW WE HANRE LLW WE HANRE LLW WE HANRE LLW WE HANRE MOWEN S DU WE HANRE S DU WE	DVF- DVF- KEPFDF- GWSDSAIKEVEECC MWGFEEQEHTTVS MWGFEEQEHTTVS TWLSEKQQI- EK- TVLSEKQQI- EX- YTWKPEWDV- QLVQ- V-	GASRH

Figure 3

Amino-acid sequence alignment of TenA-homologue proteins. The secondary structure of PH1161 protein is also presented above the alignment. The amino-acid sequence homology search was performed against all protein-sequence databases *via* the BLAST (Altschul *et al.*, 1997) server with the PH1161 amino-acid sequence and the top 20 retrieved proteins from unique organisms are aligned based on the structure of the PH1161 protein. The accession No. of each protein is also indicated in parentheses after the organism name. *Pyrococcus horikoshii*, *P. abyssi* and *P. furiosus* are euryarchaeota from the archaea. *Aerophyrum pernix*, *Pyrobacterium aerophilum*, *Sulfolobus tokodaii* and *S. solfataricus* are crenarchaeota from the archaea. *Bacillus subtilis*, *B. authracis*, *B. cereus*, *B. halodurans*, *Oceanobacillus iheyensis* and *Clostridium perfringens* are firmicutes from the bacteria. *Corynebacterium glutamicum* and *Bifidobacterium longum* are actinobacteria from the bacteria. *Bradynizobium japonicum*, *Pasteurella multocida*, *Haemophilus influenzae*, *Pseudomonas putida*, *Campylobactor jejuni* and *Helicobactor pylori* are proteobacteria from the bacteria. As shown here, TenA-homologue proteins are widely spread in various archaea and bacteria. Amino acids on coloured backgrounds indicate the following: yellow, conserved hydrophobic residues that are involved in the internal hydrophobic core; orange, other conserved hydrophobic residues – most of these residues are involved in intramolecular hydrophobic interactions; blue, conserved His residues; red, acidic residues that construct the ligand-binding pocket; green, conserved Cys residue in the ligand-binding pocket.

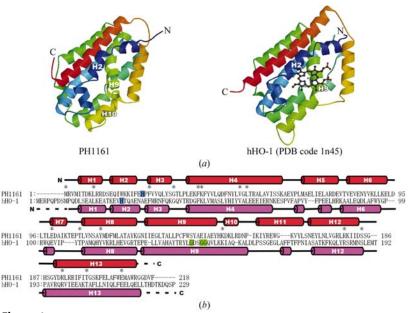


Figure 4

Comparison between PH1161 protein and human HO-1. (a) Tertiary structure. The haem molecule bound to hHO-1 is drawn as a ball-and-stick model. (b) Primary and secondary structures. The amino-acid sequences of the two proteins are aligned based on their secondary structures. His25 of hHO-1 and His23 of PH1161 are highlighted in blue. Glycine residues in the distal helix of hHO-1 are highlighted in green.

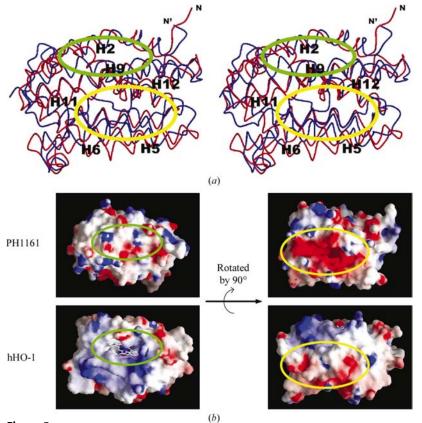


Figure 5

Comparison of the ligand-binding sites of PH1161 protein and hHO-1. The green circle indicates the haem-binding pocket and the yellow circle indicates the acidic ligand-binding pocket. (*a*) Stereoview of the main-chain superposition of PH1161 protein (red) and hHO-1 (blue). Helices constructing their ligand-binding sites are labelled. (*b*) Surface electric charge representation of the PH1161 protein and hHO-1 viewed from same direction. The bound haem molecule is drawn as a ball-and-stick model.

The hHO-1 protein also has a helical structure and its main-chain trace superimposes well onto that of the PH1161 molecule (Fig. 5a), although there is only 11% amino-acid sequence identity between them. This extensive structural similarity shows that the PH1161 protein, a TenAhomologue protein, and HO-1 proteins can be classified into the same superfamily. However, although PH1161 and hHO-1 share an identical overall fold, these two proteins exhibit quite different properties. The most notable differences were observed in their ligand-binding sites. HO-1 catalyzes the oxidative cleavage of protohaem to biliverdin (Schuller et al., 1999). There is a haem-binding pocket between its helices H2 (proximal) and H9 (distal) and a ligand-binding His residue is conserved in the proximal helix (Fig. 4). In the middle of the distal helix, Gly residues are also conserved in HO-1 proteins and these glycines play a significant role in the interaction with the haem. The distal helix is kinked at the glycines, providing a structurally favoured orientation for other distal-side residues to interact with the haem. A structural comparison between holo HO-1 and apo HO-1 showed that entry of the haem molecule results in a conformational change in the proximal helix and the latter half of the distal helix (Sugishima et al., 2002). In PH1161, the C^{α}-atom traces of helices H9-H10 differ from those of the HO-1 distal helix H9 (Fig. 5a) and the region corresponding to the haem-binding pocket of HO-1 is filled with bulky amino-acid side chains such as Trp, Phe and Arg, with no cleft being observed (Fig. 5b). The His residue that is essential for binding to the ferric ion of the haem is also not conserved here. Although the PH1161 protein has a His residue (His23) neighbouring the haem-binding His of HO-1 (Fig. 4), the side chain of His23 is directed toward the opposite side of the helix and the residue no longer has functional importance as a ligand-binding residue, while an amino-acid sequence investigation of the TenAhomologue proteins shows that the histidine is highly conserved among these proteins (Fig. 3). In spite of these differences, the remarkably similar folding of two proteins may suggest these two protein families are derived from a common ancestral protein and that during the evolutionary process these two protein families acquired different functions while retaining similar protein folding and topology.

On the other hand, PH1161 has a distinct deep pocket on the other side of the molecule (Fig. 5b). This pocket is formed between helices H5–H6 and H11–H12 (Fig. 2a). The HO-1 protein has different main-chain trajectory in the

corresponding region and no such pocket is observed (Fig. 5). As shown in Fig. 2(b), this pocket of the PH1161 protein is constructed by the acidic residues Asp and Glu, the hydrophobic residues Phe, Tyr and Trp and a Cys residue. As shown in Figs. 2(a) and 5(b), this pocket is highly charged in an acidic environment. The acidic character of the pocket arises from the acidic and hydrophobic amino-acid residues clustered here. Sequence alignment shows that the acidic residues Asp46 and Glu83 of the PH1161 protein are highly conserved among TenA-homologue proteins (Fig. 3). The alignment also shows that the hydrophobic residues Tyr42, Tyr49, Phe137, Trp162, Tyr166 and Tyr171 are also highly conserved and that Cys136 is perfectly conserved in TenA-homologue proteins. These findings indicate that this deep acidic pocket of the PH1161 protein is also commonly observed in other TenAhomologue proteins and suggests that TenA proteins may recognize and bind their ligand molecule using this characteristic pocket and exhibit their biological functions here. There is very little literature referring to the TenA protein (Ouzounis & Kyrpides, 1997; Pang et al., 1991) and this limited information is insufficient to reveal how the protein acts. In the present structure analysis, we observed a large electron density from a bound ligand molecule in this pocket in both $F_{\rm o} - F_{\rm c}$ and $2F_{\rm o} - F_{\rm c}$ maps. The same electron density was found in all molecules in the crystal. As shown in Fig. 2(b), the 'sea dragon'-shaped electron density of the bound ligand is sandwiched by the residues that construct the pocket. No compound that fits such an electron-density shape was added during protein purification and crystallization experiments, thus indicating that the PH1161 protein bound this ligand molecule under in vivo conditions in the E. coli host cell. As described above, amino acids interacting with this ligand molecule are highly conserved in TenA-family proteins, suggesting that these residues are likely to be conserved as functional residues. We could not identify the ligand molecule from the present electron density alone and other experiments using NMR spectroscopy and MS analysis are now in progress in order to characterize this bound ligand.

B. subtilis TenA is a protein that acts indirectly to enhance the production of extracellular proteases (Pang *et al.*, 1991). The tertiary structure of the PH1161 protein revealed that this protein has a unique deep acidic ligand-binding pocket, while the protein shares a common overall structure with another enzyme, haem oxygenase-1, the function of which is quite different. As discussed above, this unique character of the ligand-binding pocket of PH1161 protein suggests that the pocket is the key region for its functional activities. The present results indicate that the ligand-binding pocket of PH1161 protein binds a ligand molecule during its expression in the host cells. Identification and characterization of the unknown ligand molecule bound to the PH1161 protein is likely to accelerate our functional understanding of TenA proteins. We thank K. Miura of SPring-8 (Hyogo, Japan) for her kind help during data collection, and N. Hirano and K. Hatakeyama of Frontier Research Center for Post-genomic Science and Technology in Hokkaido University for their valuable support during sample preparation and crystallization experiments. We also thank K. Fukuyama of Osaka University for valuable comments on HO-1 protein. This work was supported by a grant-in-aid for the National Project on Protein Structural and Functional Analysis from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* 25, 3389– 3401.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). Acta Cryst. D54, 905–921.
- Cowtan, K. & Main, P. (1998). Acta Cryst. D54, 487-493.
- Diederichs, K. & Karplus, P. A. (1997). Nature Struct. Biol. 4, 269–275.
- Evans, P. R. (1997). Proceedings of the CCP4 Study Weekend. Recent Advances In Phasing, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey, pp. 97–102. Warrington: Daresbury Laboratory.
- Holm, L. & Sander, C. (1995). Trends Biochem. Sci. 20, 478–480.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Acta Cryst. A47, 110–119.
- Kawarabayasi, Y. et al. (1998). DNA Res. 5, 55-76.
- Kraulis, P. J. (1991). J. Appl. Cryst. 24, 946-950.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). J. Appl. Cryst. 26, 283–291.

Leslie, A. G. W. (1993). Proceedings of the CCP4 Study Weekend. Data Collection and Processing, edited by N. Isaacs, L. Sawyer & S. Bailey, pp. 44–51. Warrington: Daresbury Laboratory.

- Matthews, B. W. (1968). J. Mol. Biol. 33, 491-497.
- Merritt, E. A. & Bacon, D. J. (1997). Methods Enzymol. 277, 505–524.
- Msadek, T., Kunst, F., Henner, D., Klier, A., Rapoport, G. & Dedonder, R. (1990). *J. Bacteriol.* **172**, 824–834.
- Navaza, J. (1994). Acta Cryst. A50, 157-163.
- Nicholls, A., Sharp, K. A. & Honig, B. (1991). Proteins Struct. Funct. Genet. 11, 281–296.
- Otwinowski, Z. & Minor, W. (1997). Methods Enzymol. 276, 307-326.
- Ouzounis, C. A. & Kyrpides, N. C. (1997). J. Mol. Evol. 45, 708-711.
- Pang, A. H., Nathoo, S. & Wong, S. (1991). J. Bacteriol. 173, 46-54.
- Schuller, D. J., Wilks, A., Ortiz De Montellano, P. R. & Poulous, T. L. (1999). Nature Struct. Biol. 6, 860–867.
- Schuller, D. J., Zhu, W., Stojilkovic, I., Wilks, A. & Poulous, T. L. (2001). Biochemistry, 40, 11552–11558.
- Stock, J. B., Ninfa, A. J. & Stock, A. M. (1989). Microbiol. Rev. 53, 450–490.
- Sugishima, M., Omata, Y., Kakuta, Y., Sakamoto, H., Noguchi, M. & Fukuyama, K. (2000). *FEBS Lett.* **471**, 61–66.
- Sugishima, M., Sakamoto, H., Kakuta, Y., Omata, Y., Hayashi, S., Noguchi, M. & Fukuyama, K. (2002). *Biochemistry*, 41, 7293–7300.
- Terwilliger, T. C. & Berendzen, J. (1999). Acta Cryst. D55, 849-861.
- Weiss, M. S. (2001). J. Appl. Cryst. 34, 130-135.