Journal of Cell and Molecular Research (2020) 11 (2), 66-81 DOI: 10.22067/jcmr.v11i2.84965

Research Article

In silico Analysis of Determinant Factors in Microbial Protease Thermostability

Parisa Farrokh [1,2*], Fatemeh Salimi[1]

[1]*Department of Cellular and Molecular Biology, School of Biology, Damghan University, Damghan, Iran*
[2] *Institute of Biological Sciences, Damghan University, Damghan, Iran*

Received 7 January 2020 Accepted 16 February 2020

Abstract

Thermostable proteases are one of the pivotal enzymatic groups which play fundamental roles in biotechnology-related industries. The identification of bacterial thermostable enzymes through screening programs is a time and cost consuming process. So, extensive bioinformatics and experimental studies have been conducted to reveal thermo stabilizing factors. The current study was aimed to evaluate distinctive indicators among 33 thermostable and 10 mesostable proteolytic enzymes. The frequency of individual amino acids, aliphatic indexes, melting temperatures, isoelectric points, as well as, the frequency of AXXXA and GXXXG motifs were determined and compared among these enzymes. In addition, types of proteolytic enzymes and their active sites were assigned. Moreover, the frequency of alpha helixes, polar surface regions, and packing volumes of these enzymes with the known structures were characterized. Results showed that the frequency of Ala and AXXXA motifs were significantly higher in thermostable proteolytic enzymes, while they possess lower contents of Met, His, Lys and Leu in comparison to mesostable enzymes (P<0.05). According to statistical analysis, thermostable proteolytic enzymes indicated meaningful lower packing volumes than mesostable enzymes (P<0.05). Findings of the current study in addition to more detailed investigations on the thermostability mechanisms of various protein families are essential for designing more efficient industrial enzymes with functional properties at high temperatures.

***Keywords*:** Bioinformatics analysis, Protein engineering, Proteolytic enzyme, Thermostability

Introduction

The application of biocatalysts in various industries is safer than using chemical compounds and has environmental advantages (Razzaq et al., 2019). Proteases, as one of the main industrial enzymes, are responsible for around 60% of the world enzyme market (Raveendran et al., 2018). Proteolytic enzymes constitute a very large and complex group of hydrolases. Despite the high diversity of proteases' functions and structures, they were simple classified to exo- or endo-proteases according to their site of cleavage (Souza et al., 2015). In addition, exo-peptidases can be further categorized into amino- and carboxy-peptidases (Souza et al., 2015). Proteolytic enzymes also can be grouped based on residues in their catalytic active sites into serine proteases (Patel, 2017), aspartic proteases, asparagine proteases, cysteine proteases (Dadshahi et al., 2016), metalloproteases (Abebe et al., 2014), glutamic proteases, threonine proteases or proteases with mixed or unknown catalytic mechanisms (Rawlings et al., 2007; Rawlings et al., 2017).

Proteases are widely produced by all organisms, including plants, animals, fungi, bacteria, and archaea. Microbial-derived proteases have been applied for commercial purposes due to easier large-scale production (Wang et al., 2008; Haddar et al., 2010). Microbial proteases are deemed vital elements in a wide range of processes including nutritional, pharmaceutical, environmental, detergent, textile, leather, and livestock industries (Homaei et al., 2010; Homaei and Etemadipour, 2015; Barzkar et al., 2018). However, the primary limitation to the application of microbial proteases is their instability under high temperature and pH conditions (Iqbalsyah, et al., 2019). Hence, thermostable and/or thermophilic proteases which possess high abilities to preserve their activities under harsh conditions of industrial processes are more applicable in the field of biotechnology (Wakarchuk et al., 1994).

Thermostable and thermophilic enzymes are normally derived from thermophilic and mesophilic organisms. Most of the thermostable and/or thermophilic proteases are not only stable at high temperatures but also preserve their catalytic activities in the presence of detergents and other

* Corresponding author's e-mail address: p.farrokh@du.ac.ir

Licensed as Open Access CC-BY

denaturing chemicals, such as urea, guanidine-HCl, Dithiothreitol, 2-mercaptoethanol, and chaotropic agents. The application of thermostable and/or thermophilic proteases in the industrial processes have some advantages, including higher reaction rates, enhanced substrate solubility, and decreased solution viscosities. In addition, their large scale production is less susceptible to contamination since there are fewer microorganisms which can grow at high temperatures. However, screening of thermophilic microorganisms for finding thermophilic proteases is a tedious, costly, and time-consuming strategy. In this regard, engineering mesostable enzymes to develop thermostable enzymes is considered as a valuable strategy (Kumar, 2002; Li and Li, 2009: Tavano et al., 2018).

To determine factors which are responsible for thermal stability of proteins several approaches can be applied. Among which, we mention *i)* bioinformatic comparison of the protein structure for thermophilic proteins versus their mesophilic homologues, *ii)* computational studies on a dataset of thermophilic and mesophilic proteins to compare their various features, *iii)* mutational-based studies, and *iv)* comparing whole genome sequences of thermophilic species with their mesophilic homologues (Sadeghi et al., 2006).

In point of structural view, thermostable proteins have higher numbers of ionic interactions, salt bridges, disulfide, and hydrogen bonds (Sadeghi et al., 2006). They have increased packing density (Sadeghi et al., 2006), higher contents of helical structures (Sadeghi et al., 2006), less short surface loops (Sadeghi et al., 2006), decreased surface area to volume ratios (Das and Gerstein, 2000; Tekaia et al., 2002), and decreased internal cavities (Pellegrini et al., 1999). At the level of amino acid sequences, thermophilic proteins have a higher frequency of hydrophobic (Razvi and Scholtz, 2006) and charged amino acids, higher relative content of Arg, Glu (Pack et al., 2013), lower occurrence of bulky polar residues, decreased contents of uncharged polar residues (Ser, Thr, Asn, and Gln) (Haney et al., 1999), and increased contents of aromatic residues (Chakravorty et al., 2011).

Therefore, further investigations about the thermostability encountered mechanisms are essential for theoretical description of protein folding and stability and also, for designing efficient thermostable industrial enzymes. Since thermostable microbial proteases have a key contributory role in the market of industrial enzyme, and whereas there are no shared particular sequences or structural patterns among heat-stable proteases, focusing on this group of enzymes is worthwhile. Here, we selected 43 thermostable/thermophilic and mesophilic proteases which their characteristics had been previously determined in experimental studies. This study, innovatively investigated the consistency of well-known thermostability parameters among thermostable proteolytic enzymes.

Materials and Methods

Sequence collection of microbial proteases

Thirty-three bacterial thermostable proteolytic enzymes, which their optimum temperature and/or thermal stability had been reported in experimental studies were selected from UniProt proteomic server (Consortium 2018). Moreover, ten well-identified mesostable proteolytic enzymes' sequences were also collected from UniProt to compare with thermostable ones. In this regard, amino acid sequences of these enzymes were retrieved in FASTA format.

Investigation the thermostability properties of different proteases

The amino acid composition, aliphatic index, and isoelectric point of each proteolytic enzyme were determined through the ProtParam tool (web.expasy.org/protparam/). Melting temperatures (Tm) and net charges of the enzymes at pH 7 were calculated using Tm Predictor (tm.life.nthu.edu.tw/) and PepCalc (pepcalc.com), respectively. Frequency of AXXXA and GXXXG motifs in the sequences were assigned through manual searching and counting.

Determination of the types of proteolytic enzymes and their active sites

The types of the proteases were determined based on their conserved domain composition using CD-search tool (Marchler-Bauer et al., 2016). The active sites of proteolytic enzymes with protease activity were identified through the literature review (Yamagata and Ichishima, 1995; Wu et al., 2004; Pombejra et al., 2018). To determine the active sites of enzymes with peptidase activity, the MEROPS database (Rawlings et al., 2017) in addition to literature review (Medrano et al., 1998; Goldstein et al., 2005; Bjelke et al., 2006; Ohara-Nemoto et al., 2014; Reddi et al., 2014) were applied.

Multiple sequence alignments

In order to determine the possible differences of conserved motifs and residues in the sequences of mesostable and thermostable proteases, multiple

sequence alignments were performed in IBIVU server using PRALINE software (Simossis and Heringa, 2005). Since the investigated amino acid sequences have significantly different lengths (285-1364 amino acid residues), matrix PAM 250 with gap opening 10 and gap extension 1 were used for alignments.

Structural analyses

Among 43 selected proteolytic enzymes, nine enzymes had a tertiary structure in the Protein Data Bank (PDB) (Bank, 2000) which four and five of them are mesostable and thermostable/thermophilic proteolytic enzymes, respectively. Structural analysis of these enzymes was performed using VADAR (Volume, Area, Dihedral Angle Reporter) software version 1.8 which can be accessed at http://redpoll. pharmacy.ualberta.ca/vadar/ (Willard et al., 2003). Furthermore, frequency of alpha helixes, polar surface regions, and packing ratios in these structures were determined. One of these structures (PDB ID: 5J44, which was related to Q8VSL2 sequence) with more than 2000 residues, was not investigable using VADAR software.

Statistical analyses

Homogeneity of variance and normal distribution of data including frequency of each twenty amino acids, AXXXA and GXXXG motifs, the mean of pI (isoelectric point), aliphatic indexes, Tm, net charges in pH 7, melting temperatures in both thermophilic/thermostable and mesostable enzymes groups were evaluated using Levene and Shapiro Wilk tests in SPSS version 23, respectively. To compare the mean of each variable in two investigated subsets (thermophilic/thermostable and mesostable proteolytic enzymes), independent *t*-test and Mann-Whitney test were applied for data with or without normal distribution, respectively. Finally, chi-square test was carried out to evaluate the possible differences in melting temperatures of proteolytic enzymes in these two groups.

Results

Source and properties of investigated sequences

All selected enzymes have bacterial origin except Q2QC89, which has been derived from an archaeon (*Thermococcus* sp.).

Proteolytic enzymes in the current study with an optimum temperature of 50°C or higher and/or thermal stability in the mentioned temperatures were considered as thermostable enzymes, and the others were placed in the mesostable group. Some properties of the proteolytic enzymes which were obtained from literature are summarized in table 1.

Distinctive features of thermophilic/thermostable proteolytic enzymes

Some characteristics of the enzymes which are related to the thermostability were summarized in table 2. The Tm Predictor categorizes proteins based on their melting temperatures in three ranges; >65°C, <55°C and 55-65°C. Statistical analysis showed that there were no significant differences in melting temperatures of thermophilic/thermostable and mesophilic proteolytic enzymes ($P>0.05$).

In addition, statistical analysis confirmed normal distributions of data related to frequency of AXXXA motifs, and amino acids except Ala, Arg, Cys, Gln, His, Leu, Phe ($P>0.05$), while isoelectric points, net charges, aliphatic indexes, and GXXXG motifs did not show normal distributions ($P<0.05$). In case of pI, aliphatic index and net charge parameters, thermostable and mesostable enzymes were similar; however, the frequency of AXXXA motif was significantly higher in thermophilic/thermostable proteolytic enzymes when compared to mesostable ones ($P<0.05$).

The percentages of amino acids present in proteolytic enzymes are presented in Figure 1. According to the *t*-test and Mann Whitney results, there was a significant difference in frequencies of Met, Ala, Leu, Lys and His between thermostable proteolytic enzymes and mesostable ones. Although, thermophilic/thermostable enzymes had higher percentages of Ala (9.2% *vs.* 8.59% in the other group) and lower percentages of Met (2.01%), Leu (7.16%), Lys (5.06%) and His (2.32%) in comparison to mesostable proteins. The frequency of the above-mentioned residues in the mesostable enzymes were as follows: Met (2.68%), Leu (8.22%), Lys (5.84%), and His (2.61%) (Figure 1). Moreover, to compare each of the twenty amino acids, the analysis of the hydrophobic, aromatic, polar charged, and polar uncharged amino acid groups was performed; but, the subsets of the residues did not show any significant differences between thermostable and mesostable enzymes.

Types of proteolytic enzymes and their active sites

Whereas the types of some proteolytic enzymes were not precisely determined in the previous studies, the retrieved sequences were classified according to their conserved domains. It was confirmed that 25 and 18 of investigated enzymes could be grouped as proteases and peptidases,

respectively (Table 2). Two types of catalytic triads within conserved sequences were observed among the proteases, which are including Asp, His, and Ser or His, His, and Glu. However, peptidases due to their variability had different active sites, and one of them (UniProt ID: G5DCB7) belongs to the peptidase family with unknown catalytic mechanism based on the MEROPS database (Table 2).

Table 1. Investigated microbial proteolytic enzymes and some of their properties obtained from literature.

UniProt ID	Source of microorganisms	Optimum temperature	Thermal stability	Optimum pH	pH stability range	References
Thermostable/thermophilic enzymes						
P06874	*Bacillus stearothermophilus*	*ND*	65ºC	7	*ND*	(Fujii et al. 1983)
P43133	*Bacillus stearothermophilus*	*ND*	65ºC	*ND*	6.5-7.5	(Kubo and Imanaka 1988)
P23341	*Thermus aquaticus* YT-1	75-80ºC	*ND*	*ND*	*ND*	(Motoshima et al. 1990)
P39899	*Bacillus subtilis*	ND	65ºC	6.6	*ND*	(Tran et al. 1991)
P23384	*Bacillus caldolyticus*	77ºC	*ND*	7	*ND*	(Van den Burg et al. 1991)
P42663	*Thermus aquaticus* YT-l	80ºC	*ND*	8	*ND*	(S.-H. Lee et al. 1992)
P41363	*Bacillus* sp. no. AH-101	*ND*	30-70ºC	ND	12-13	(Takami et al. 1992)
P0CH29	*Bacillus megaterium* ATCC 1458 1	58ºC	*ND*	6.4-7.2	*ND*	(KÜHN and FORTNAGEL 1993)
Q45670	*Bacillus* sp. strain AK.1	75ºC	*ND*	8.5	*ND*	(Maciver et al. 1994)
Q45621	*Bacillus* sp. NKS-21	50ºC	*ND*	*ND*	6-11	(Yamagata and Ichishima 1995)
P80146	*Thermus* sp. strain Rt41A	*ND*	70ºC	8	*ND*	(Munro et al. 1995)
P04189	*Bacillus subtilis*	*ND*	50ºC	8	*ND*	(Kamal et al. 1995)
Q43880	*Bacillus sp.*	82ºC	*ND*	*ND*	*ND*	(Vecerek and Kyslik 1995)
Q99405	*Bacillus* sp. KSM-K16	55ºC	*ND*	10	*ND*	(Kobayashi et al. 1995)
Q56365	*Thermoactinomyces* sp. E79	85ºC	*ND*	11	5-12	(J.-K. Lee et al. 1996)
Q59223	*Bacillus* sp. strain EAI	*ND*	85-95ºC	6.5	6-7.5	(Saul et al. 1996)
P74937	*Thermoactinomyces* sp. HS682	65ºC	*ND*	11	6-12	(Tsuchiya et al. 1997)
O33599	*Staphylococcus aureus*	*ND*	100ºC	*ND*	5-8	(Ramadurai et al. 1999)

Q93JY4	*Prevotella albensis* M384	*ND*	60ºC	*ND*	7-8	(Walker et al. 2003)
Q6W4N2	*Bacillus* sp. WF146	58ºC	*ND*	8	*ND*	(Wu et al. 2004)
Q84FM9	*Fervidobacterium islandicum*	80ºC	*ND*	8	*ND*	(Gödde et al. 2005)
Q2QC89	*Thermococcus* sp.NA1	*ND*	70-80ºC	6.5	*ND*	(H. S. Lee et al. 2006)
Q8G6Z9	*Bifidobacterium longum*	50ºC	40-60ºC	8	4-8	(Seo et al. 2007)
E0XH65	*Bacillus* sp. B001	60ºC	20-90ºC	10	5-12	(Deng et al. 2010)
G8HV17	*Bacillus circulans* MTCC 7906	60ºC	*ND*	9	*ND*	(Kaur et al. 2012)
G5DCB7	*Geobacillus thermoleovorans* DSM 15325	50ºC	*ND*	7.4	*ND*	(Jasilionis et al. 2012)
J9XWB6	*Serratia* sp. ZF03	*ND*	50-55ºC	8	8-10	(Salarizadeh et al. 2014)
H2BKX5	*Myroides profundi* D25	60ºC	*ND*	8.5	*ND*	(Ran et al. 2014)
W5RWH8	*Geobacillus thermoleovorans* DSM 15325	40ºC	50-60ºC	7.3	5-8	(Jasilionis and Kuisiene 2015)
A0A0C4XY83	*Streptomyces* sp.M30	75ºC	*ND*	9	6-11	(Xin et al. 2015)
H6WCS0	*Dichelobacter nodosus*	35ºC	15–65ºC	7	4-10	(Wani et al. 2016)
Q3HTI0	*Bacillus cereus* PMW8	60ºC	40-70ºC	9	*ND*	(Esakkiraj et al. 2016)
Q45300	*Bacillus licheniformis*	60ºC	*ND*	10	*ND*	(Ramakrishnan et al. 2017)
Mesostable enzymes						
P46544	*Lactobacillus delbrueckii* subsp. *bulgaricus* CNRZ 397	37ºC	*ND*	*ND*	*ND*	(Atlan et al. 1994)
P94870	*Lactobacillus helveticus* CNRZ32	*ND*	32-37ºC	4.5	*ND*	(Fenster et al. 1997)
O07121	*Lactococcus lactis* MG1363	37ºC	*ND*	*ND*	*ND*	(Hellendoorn et al. 1997)
Q8VSL2	*Shigella flexneri*	37ºC	*ND*	7.5	*ND*	(Benjelloun-Touimi et al. 1998)
Q7MUW6	*Porphyromonas gingivalis*	*ND*	25-37ºC	ND	6-8	(Banbula et al. 1999)
Q9L4G1	*Lactobacillus helveticus*	*ND*	25-37ºC	7.5	6-8	(Savijoki and Palva 2000)
O82882	*Escherichia coli* O157:H7	37-42ºC	*ND*	6.5-7	*ND*	(Grys et al. 2006)
Q29ZA8	*Bacillus intermedius*	37ºC	*ND*	8	*ND*	(Sharipova et al. 2008)

P9WK19	*Mycobacterium tuberculosis* H37Rv	37ºC	*ND*	7.5	*ND*	(Zhang et al. 2009)
B2RIT0	*Porphyromonas gingivalis*	37ºC	*ND*	6	*ND*	(Ohara-Nemoto et al. 2014)

ND: Not Determined

Table 2. *In silico* characterization of investigated proteolytic enzymes.

UniProt ID	Types of proteases	Active sites	pI	Aliphatic Index	Tm	Net charge in pH	Number of AXXXA motifs	Number of GXXXG motifs
P06874	Zn-dependent metalloprotease	H374, H378, E398	5.68	76.72	<55	-8.9	8	2
P43133	Zn-dependent metalloprotease	H377, H381, E401	5.82	76.46	55-65	-6.9	6	3
P23341	Aminopeptidase T	Y352	5.31	89.49	>65	-13.9	7	2
P39899	Zn-dependent metalloprotease	H369, H373, E393	5.43	73.10	55-65	-17.3	6	4
P23384	Zn-dependent metalloprotease	H370, H374, E394	5.64	75.33	<55	-8.2	8	4
P42663	Carboxypeptidase	E277	5.53	83.76	55-65	-13.6	6	6
P41363	Serine protease	D124, H154, S307	6.56	90.80	>65	-1	4	1
P0CH29	Zn-dependent metalloprotease	H388, H392, E412	8.39	70.64	55-65	4	6	5
Q45670	Thermitase-like proteins	D160, H193, S347	4.68	82.97	>65	-17.5	7	1
Q45621	Serine protease	D49, H86, S250	4.83	90.50	55-65	-17.5	3	1
P80146	Proteinase K-like proteins	D171, H204, S356	6.16	90.98	>65	-2.6	5	2
P04189	Serine protease	D138, H170, S327	9.04	81.23	>65	5.8	5	3

Q43880	Zn-dependent metalloprotease	H372, H376, E396	5.47	73.97	<55	-10.1	7	3
Q99405	Serine protease	D143, H173, S326	4.67	91.18	>65	-19.2	5	1
Q56365	Serine protease	D143, H176, S330	6.04	78.57	>65	-3.3	8	1
Q59223	Thermitase-like proteins	H372, H376, E396	5.39	74.51	55-65	-11.1	7	3
P74937	Zn-dependent metalloprotease	D49, H86, S249	4.92	87.13	>65	-15.6	3	0
O33599	Metallopeptidase	H291	6.16	38.10	55-65	-4.6	2	2
Q93JY4	Dipeptidyl peptidase IV	Y511, S600, D674, H706	7.56	70.77	55-65	2.1	5	2
Q6W4N2	Serine protease	D149, H185, S363	5.28	81.41	>65	-8.5	7	3
Q84FM9	Serine protease	D210, H248, S424	5.45	93.84	>65	-9.6	3	4
Q2QC89	Zn-dependent carboxypeptidase	E297	5.61	84.97	>65	-11.7	2	2
Q8G6Z9	Dipeptidase	C3	4.46	71.07	55-65	-47	2	0
E0XH65	Serine protease	D138, H168, S321	4.0	84.08	>65	-40.5	3	1
G8HV17	Serine protease	D155, H187, S384	5.14	79.41	55-65	-14.7	3	2
G5DCB7	Collagenase-like protease	Unknown	5.39	88.06	>65	-13.4	4	1
J9XWB6	Serralysin	E175	4.63	64.18	>65	-23.8	3	4
H2BKX5	Kp43 proteases	D129, H177, S378	5.85	79.68	55-65	-8	7	2
W5RWH8	Oligo peptidase F	E401	8.90	87.28	>65	9.2	9	3
A0A0C4XY83	Proteinase K-like proteins	D156, H187, S339	4.25	80.33	>65	-21.4	7	3
H6WCS0	Serine protease	D156, H220, S392	6.25	78.92	55-65	-4	5	3

Q3HTI0	Zn-dependent metalloprotease	H143, H147, E167	5.18	65.65	55-65	-8.3	6	2
Q45300	Serine protease	D137, H168, S325	8.94	84.70	>65	4.8	4	2
P46544	Proline iminopeptidase	S107, D246, D273	5.09	87.69	>65	-12.4	3	2
P94870	Aminopeptidases	Q64, C70, H362, N383	5.21	75.46	>65	-13.2	1	2
O07121	Dipeptidase	H92	4.73	79.87	>65	-33.7	4	3
Q8VSL2	Serine peptidase	H134, D162, S267	6.10	74.02	>65	-8.9	7	14
Q7MUW6	Dipeptidyl peptidase IV	Y518, S603, D678, H710	6.16	74.44	55-65	-8.2	4	6
Q9L4G1	Tripeptidase	D84, E178	4.77	73.90	>65	-29.8	1	2
O82882	Metallopeptidases	E447	6.39	74.59	55-65	-5.8	5	8
Q29ZA8	Serine protease	D138, H170, S327	8.96	80.73	>65	4.8	5	1
P9WK19	Methionine aminopeptidase	C105	5/07	87.89	55-65	-11.5	3	2
B2RIT0	Dipeptidyl aminopeptidase	S542, D627, H659	5.75	68.52	55-65	-10.7	4	4
Significant value			*NS*	*NS*	*NS*	*NS*	**	*NS*

NS, * and ** represent no significant differences, significant differences at 90% and 95% confidence intervals, respectively.

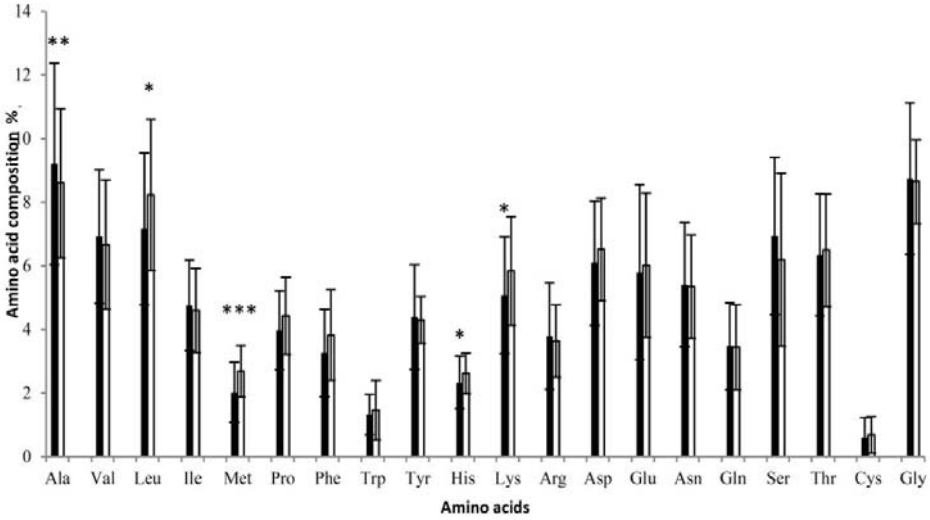


Figure 1. Frequency of different amino acids is reported for thermostable (black columns) and mesostable (white columns) groups. *, ** and *** represent significant differences at 90%, 95%, and 99% confidence intervals, respectively.

Sequence alignments

Results obtained through multiple sequence alignments indicated somewhat conserved sequences (from amino acid residues 1000 to1300) which are shown in supplementary file 1.

The structural features of thermostable/ proteolytic and mesostable enzymes

The numbers of alpha helixes, polar surface regions, and total volumes of available structures were predicted (Table 3). The statistical analyses showed that among structural characteristics, just total volumes of thermophilic/thermostable proteolytic enzymes were significantly lower than that of mesostable ones (P<0.05).

Discussion

In recent years, considerable efforts have been made to understand influential factors involved in thermal stability of thermostable or thermophilic proteins derived from mesophilic, thermophilic or

Table 3. Structural analyses of thermophilic/thermostable proteolytic enzymes and mesostable ones using VADAR software.

UniProt ID/ PDB ID	Type of enzymes	Alpha helix (%)	Polar surface Exposed)regions (polar ASA[1] (A^o) 10^{-2}	Total volume (Packing) 10^{-3} (A^o)
P43133/5WR3	**Thermophile/Thermostable**	37	4374.7	40568.7
Q45670/1DBI		28	3188.9	33870.1
P04189/1SCJ		96	3713.3	42181.7
Q99405/1MPT		78	3197.2	31663.4
O33599/1QWY		23	3810.9	29772.0
Q7MUW6/2D5L	**Mesostable**	96	6132.5	92421.4
O82882/3UJZ		56	6581.3	82474.0
P9WK19/1Y1N		71	2086.1	36599.7
Significant value		*NS*	*NS*	**

NS, and ** represent no significant difference and significant difference at 95% confidence interval, respectively.
[1] Accessible surface area.

hyper-thermophilic organisms. Thermophilic proteins preserve their native structures and consequently their activities under harsh environmental conditions, while, their mesostable homologs already denature. Comprehensive knowledge about factors involved in thermostability can be applied as a promising approach in developing thermostable enzymes through protein engineering. Most of the studies which have been worked on the thermal stability of proteins did not generally focus on a family or a limited group of proteins (Kumar et al., 2000; Liang et al., 2005; Sadeghi et al., 2006; Gromiha and Suresh, 2008). It seems that it is a reason for inconsistent results.

One of the fundamental features which profoundly affected structural, functional and biological properties of proteins is their amino acid composition. Analysis of amino acid composition and frequency can provide beneficial data about the importance of each amino acid in thermal stability of proteins. This feature can be different from one protein to another one in each organism or it can be taxon-specific (Zhou et al., 2008).

The current study showed that thermophilic or thermostable enzymes possess a higher frequency of Ala compared to mesostable ones. This residue is an appropriate helix former (Chakravorty et al., 2011) and through hydrophobic interactions provides conformational stability in the inner parts of proteins (Creighton, 1993). This feature leads to better packing, higher rigidity, hydrophobicity, and consequently more thermostability of proteins (Chakravarty and Varadarajan, 2000). This finding is analogous with previously published studies (Argos et al., 1979; Chakravarty and Varadarajan, 2000; Pack and Yoo, 2004). The percentage of Leu, another nonpolar residue, was lower in thermostable proteases than the mesostable ones. In case of Leu reported frequencies are not similar in various studies. Taylor and Vaisman detected higher percentages of Leu in thermophilic proteins (Taylor and Vaisman, 2010); while, Chakravarty and Varadaragan observed an opposite trend (Chakravarty and Varadarajan, 2000).

The present study indicated that thermophilic or thermostable proteases possess lower frequencies of thermolabile Met residue. This finding is in consistent with a previous finding (Kumar et al., 2000; Xu et al., 2003). The frequency of the other thermolabile amino acids, including Asn, Gln, and free Cys, in spite of their deamination or oxidation at high temperatures (Tomazic and Klibanov, 1988; Russell et al., 1994; Catanzano et al., 1997; Kumar et al., 2000; Xu et al., 2003), did not show indictable differences among the sequences. In the research of Kumar et al., among these temperature sensitive residues, only Cys was significantly lower in thermophilic proteins (Kumar et al., 2000).

Polar charged amino acids, including Arg, Lys, His, Asp, and Glu, contribute to the electrostatic interactions and enhancement of thermostability (Dill, 1990; Creighton, 1993; Ladbury et al., 1995; Vogt et al., 1997; Kumar et al., 2000). Although, in the current study, no significant differences were observed in the frequencies of Arg, Asp, and Glu between thermostable and mesostable proteases. Furthermore, lesser contents of Lys and His observed in thermostable proteins compared to mesostable ones. There are inconsistent reports about Lys frequency. Taylor and Vaisman found the lower percentages of Lys in thermophilic proteins (Taylor and Vaisman, 2010), while Cambillau and Claverie reported a higher percentages of Lys in hyperthermophiles (Cambillau and Claverie, 2000). Lower frequency of His in thermophilic proteins also reported in previous studies (Chakravarty and Varadarajan, 2000; Pack and Yoo, 2004; Sadeghi et al., 2006).

The aliphatic index, which is calculated based on the presence of amino acids with aliphatic side chains (Zhou et al., 2008), and melting temperature (Kumar et al., 2000) are two parameters which have positive effects on protein thermostability. However, in the present comparison, no significant differences were found between mesostable and thermostable groups for these criteria. Panja et al. showed that thermophilic proteins had a negative net charge at neutral pH and a slight acidic pI (Panja et al., 2015); while, thermostable and thermophilic proteolytic enzymes did not show any differences at the significant level of 0.05 in their study. Therefore, it seems that these factors are not determining agents for thermostability of proteases.

The AXXXA motif through helical interactions creates more stability in protein structures (Kleiger et al., 2002; Chakravorty et al., 2011). Higher frequency of AXXXA motif and poly Ala residues were also revealed through sequence analysis of the thermostable *Bacillus* lipases (Chakravorty et al., 2011). However, thermostable sequences in this study in spite of having significant higher Ala residues and AXXXA motifs did not have a higher percentages of alpha helixes in comparison with mesostable ones. If more PDB structures were available for the studied proteases, these contradictory findings might not occur. Since, limited structures of selected proteases and peptidases were available, the comprehensive

structural analyses were not possible. The exposed polar surfaces have been proposed as an effective factor for thermo stabilization of lipases (Chakravorty et al., 2011); however, our results were opposite to this expectation. Only packing volumes of the structures were significantly higher in thermostable enzymes than mesostable ones.

Investigating the thermostable and mesostable sequences through multiple sequence alignments is beneficial for finding short consensus sequences to design degenerate primers (Morya et al., 2012). Furthermore, these alignments will provide better insights for substitution of amino acids in protein engineering studies. Here, no considerable distinction was observed among the active sites or consensus motifs of thermostable and meostable enzymes.

Conclusion

In conclusion, according to the present study, thermostability of proteolytic enzymes not only can be attributed to the higher percentages of Ala, and fewer frequencies of His, Lys and Met, but also can be due to the presence of higher contents of AXXXA motifs and more packing structures in comparison to mesostable proteases. Without a doubt, the results of comparative studies between thermophilic proteins and their mesostable counterparts will revealed some thermo stabilizing factors. However, these results could not be accepted as general modes governing protein thermostability, because they completely depend on studied protein families. In addition, it is so difficult to relate any single factor as the main reason which is responsible for enhancement of thermal stability (Szilágyi and Závodszky, 2000). Therefore, extensive bioinformatic-based studies should be carried out on various thermostable and mesostable protein groups. Furthermore, the contradictory results should be taken into account to reach a reliable data and to enhance the thermal stability of mesostable proteins with desirable properties during protein designing or re-engineering processes.

Conflicts of interest

The authors declare that they have no conflict of interest.

References

Abebe B., Abrham S., Genet A., Hiwot G., Paulos K. and Melese A. (2014) Isolation, optimization and characterization of protease producing bacteria from soil and water in Gondar town, North West Ethiopia. International Journal of Bacteriology, Virology and Immunology 1:20-24.

Argos P., Rossmann M. G., Grau U. M., Zuber H., Frank G. and Tratschin J. D. (1979) Thermal stability and protein structure. Biochemistry 18:25, 5698-5703.

Atlan D., Gilbert C., Blanc B. and Portalier R. (1994) Cloning, sequencing and characterization of the pepIP gene encoding a proline iminopeptidase from *Lactobacillus delbrueckii* subsp. *bulgaricus* CNRZ 397. Microbiology 140:3, 527-535.

Banbula A., Mak P., Bugno M., Silberring J., Dubin A., Nelson D., et al. (1999) Prolyl tripeptidyl peptidase from *Porphyromonas gingivalis*. A novel enzyme with possible pathological implications for the development of periodontitis. The Journal of Biological Chemistry 274:14, 9246-9252.

Berman H. M., Westbrook j., Feng Z., Gilliland G., Bhat T. N., Weissig H., et al. (2000) The protein data bank. Nucleic Acids Research 28:235-242.

Barzkar N., Homaei A., Hemmati R. and Patel S. (2018) Thermostable marine microbial proteases for industrial applications: scopes and risks. Extremophiles 22:3, 335-346.

Benjelloun-Touimi Z., Tahar M. S., Montecucco C., Sansonetti P. J. and Parsot C. (1998) SepA, the 110 kDa protein secreted by *Shigella flexneri*: two-domain structure and proteolytic activity. Microbiology 144:7, 1815-1822.

Bjelke J. R., Christensen J., Nielsen P. F., Branner S., Kanstrup A. B., Wagtmann N., et al. (2006) Dipeptidyl peptidases 8 and 9: specificity and molecular characterization compared with dipeptidyl peptidase IV. Biochemical Journal 396:2, 391-399.

Cambillau C. and Claverie J. M. (2000) Structural and genomic correlates of hyperthermostability. Journal of Biological Chemistry 275:42, 32383-32386.

Catanzano F., Barone G., Graziano G. and Capasso S. (1997) Thermodynamic analysis of the effect of

selective monodeamidation at asparagine 67 in ribonuclease A. Protein Science 6:8, 1682-1693.

Chakravarty S. and Varadarajan R. (2000) Elucidation of determinants of protein stability through genome sequence analysis. FEBS Letters 470:1, 65-69.

Chakravorty D., Parameswaran S., Dubey V. K. and Patra S. (2011) In silico characterization of thermostable lipases. Extremophiles 15:1, 89-103.

Consortium U. (2018) UniProt: a worldwide hub of protein knowledge. Nucleic Acids Research 47:D1, D506-D515.

Creighton T. E. (1993) Proteins: structures and molecular properties. Second ed. W. H. Freeman and Co. New York, 80-100 pp.

Dadshahi Z., Homaei A., Zeinali F., Sajedi R. H. and Khajeh K. (2016) Extraction and purification of a highly thermostable alkaline caseinolytic protease from wastes *Penaeus vannamei* suitable for food and detergent industries. Food Chemistry 202: 110-115.

Das R. and Gerstein M. (2000) The stability of thermophilic proteins: a study based on comprehensive genome comparison. Functional and Integrative Genomics 1:1, 76-88.

Deng A., Wu J., Zhang Y., Zhang G. and Wen T. (2010) Purification and characterization of a surfactant-stable high-alkaline protease from *Bacillus* sp. B001. Bioresource Technology 101:18, 7100-7106.

Dill K. A. (1990) Dominant forces in protein folding. Biochemistry 29:31, 7133-7155.

Esakkiraj P., Meleppat B., Lakra A. K., Ayyanna R. and Arul V. (2016) Cloning, expression, characterization and application of protease produced by *Bacillus cereus* PMW8. RSC Advances 6:45, 38611-38616.

Fenster K. M., Parkin K. L. and Steele J. L. (1997) Characterization of a thiol-dependent endopeptidase from *Lactobacillus helveticus* CNRZ32. Journal of Bacteriology 179:8, 2529-2533.

Fujii M., Takagi M., Imanaka T. and Aiba S. (1983) Molecular cloning of a thermostable neutral protease gene from *Bacillus stearothermophilus* in a vector plasmid and its expression in *Bacillus stearothermophilus* and *Bacillus subtilis*. Journal of Bacteriology 154:2, 831-837.

Gödde C., Sahm K., Brouns S. J., Kluskens L. D., van der Oost J., de Vos W. M., et al. (2005) Cloning and expression of islandisin, a new thermostable subtilisin from *Fervidobacterium islandicum*, in *Escherichia coli*. Applied and Environmental Microbiology 71:7, 3951-3958.

Goldstein J., Kordula T., Moon J., Mayo J. and Travis J. (2005) Characterization of an extracellular dipeptidase from *Streptococcus gordonii* FSS2. Infection and Immunity 73:2, 1256-1259.

Gromiha M. M. and Suresh M. X. (2008) Discrimination of mesophilic and thermophilic proteins using machine learning algorithms. Proteins 70:4, 1274-1279.

Grys T. E., Walters L. L. and Welch R. A. (2006) Characterization of the StcE protease activity of *Escherichia coli* O157: H7. Journal of Bacteriology 188:13, 4646-4653.

Haddar A., Fakhfakh-Zouari N., Hmidet N., Frikha F., Nasri M. and Kamoun A. S. (2010) Low-cost fermentation medium for alkaline protease production by *Bacillus mojavensis* A21 using hulled grain of wheat and sardinella peptone. Journal of Bioscience and Bioengineering 110:3, 288-294.

Haney P. J., Badger J. H., Buldak G. L., Reich C. I., Woese C. R. and Olsen G. J. (1999) Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. Proceedings of the National Academy of Sciences 96:7, 3578-3583.

Hellendoorn M. A., Franke-Fayard B. M., Mierau I., Venema G. and Kok J. (1997) Cloning and analysis of the pepV dipeptidase gene of *Lactococcus lactis* MG1363. Journal of Bacteriology 179:11, 3410-3415.

Homaei A. and Etemadipour R. (2015) Improving the activity and stability of actinidin by immobilization on gold nanorods. International Journal of Biological Macromolecules 72:1176-1181.

Homaei A. A., Sajedi R. H., Sariri R., Seyfzadeh S. and Stevanato R. (2010) Cysteine enhances activity and stability of immobilized papain. Amino Acids 38:3, 937-942.

Iqbalsyah T. M., Malahayati, Atikah, and Febriani. (2019) Purification and partial characterization of a thermo-halostable protease produced by *Geobacillus* sp. strain PLS A isolated from undersea fumaroles. Journal of Taibah University for Science 13:1, 850-857.

Jasilionis A., Kaupinis A., Ger M., Valius M., Chitavichius D. and Kuisiene N. (2012) Gene expression and activity analysis of the first thermophilic U32 peptidase. Central European Journal of Biology *7*:4, 587-595.

Jasilionis A. and Kuisiene N. (2015) Characterization of a novel thermostable oligopeptidase from *Geobacillus thermoleovorans* DSM 15325. Journal of Microbiology and Biotechnology 25:7, 1070-1083.

Kamal M., Höög J. O., Kaiser R., Shafqat J., Razzaki T., Zaidi Z. H., et al. (1995) Isolation, characterization and structure of subtilisin from a thermostable *Bacillus subtilis* isolate. FEBS Letters 374:3, 363-366.

Kaur I., Kocher G. S. and Gupta V. (2012) Molecular cloning and nucleotide sequence of the gene for an alkaline protease from *Bacillus circulans* MTCC 7906. Indian Journal of Microbiology 52:4, 630-637.

Kleiger G., Grothe R., Mallick P. and Eisenberg D. (2002) GXXXG and AXXXA: common α-helical interaction motifs in proteins, particularly in extremophiles. Biochemistry 41:19, 5990-5997.

Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T., Koike K., et al. (1995) Purification and properties of an alkaline protease from alkalophilic *Bacillus* sp. KSM-K16. Applied Microbiology and Biotechnology 43:3, 473-481.

Kubo M. and Imanaka T. (1988) Cloning and nucleotide sequence of the highly thermostable neutral protease gene from *Bacillus stearothermophilus*. Microbiology 134:7, 1883-1892.

Kühn S. and Fortnagel P. (1993) Molecular cloning and nucleotide sequence of the gene encoding a calcium-dependent exoproteinase from *Bacillus megaterium* ATCC 14581. Microbiology 139:1, 39-47.

Kumar C. (2002) Purification and characterization of a thermostable alkaline protease from alkalophilic *Bacillus pumilus*. Letters in Applied Microbiology. 34:1, 13-17.

Kumar S., Tsai C. J. and Nussinov R. (2000) Factors enhancing protein thermostability. Protein Engineering 13:3, 179-191.

Ladbury J. E., Wynn R., Thomson J. A. and Sturtevant J. M. (1995) Substitution of charged residues into the hydrophobic core of *Escherichia coli* thioredoxin results in a change in heat capacity of the native protein. Biochemistry 34:7, 2148-2152.

Lee H. S., Kim Y. J., Bae S. S., Jeon J. H., Lim J. K., Kang S. G., et al. (2006) Overexpression and characterization of a carboxypeptidase from the hyperthermophilic archaeon *Thermococcus* sp. NA1. Bioscience, Biotechnology, and Biochemistry 70:5, 1140-1147.

Lee J. K., Kim Y. O., Kim H. K., Park Y. S. and Oh T. K. (1996) Purification and characterization of a thermostable alkaline protease from *Thermoactinomyces* sp. E79 and the DNA sequence of the encoding gene. Bioscience, Biotechnology, and Biochemistry 60:5, 840-846.

Lee S. H., Minagawa E., Taguchi H., Matsuzawa H., Ohta T., Kaminogawa S., et al. (1992) Purification and characterization of a thermostable carboxypeptidase (carboxypeptidase Taq) from *Thermus aquaticus* YT-1. Bioscience, Biotechnology, and Biochemistry 56:11, 1839-1844.

Li A. N. and Li D. C. (2009) Cloning, expression and characterization of the serine protease gene from *Chaetomium thermophilum*. Journal of Applied Microbiology 106:2, 369-380.

Liang H. K., Huang C. M., Ko M. T. and Hwang J. K. (2005) Amino acid coupling patterns in thermophilic proteins. Proteins 59:1, 58-63.

Maciver B., McHale R. H., Saul D. J. and Bergquist P. L. (1994) Cloning and sequencing of a serine

proteinase gene from a thermophilic *Bacillus* species and its expression in *Escherichia coli*. Applied and Environmental Microbiology 60:11, 3981-3988.

Marchler-Bauer A., Bo Y., Han L., He J., Lanczycki C. J., Lu S., et al. (2016) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Research 45:D1, D200-D203.

Medrano F., Alonso J., Garcia J., Romero A., Bode W. and Gomis-Ruth F. (1998) Structure of proline iminopeptidase from *Xanthomonas campestris* pv. citri: a prototype for the prolyl oligopeptidase family. The EMBO Journal 17:1, 1-9.

Morya V. K., Yadav S., Kim E. K. and Yadav D. (2012) In silico characterization of alkaline proteases from different species of *Aspergillus*. Applied Biochemistry and Biotechnology 166:1, 243-257.

Motoshima H., Azuma N., Kaminogawa S., Ono M., Minagawa E., Matsuzawa H., et al. (1990) Molecular cloning and nucleotide sequence of the aminopeptidase T gene of *Thermus aquaticus* YT-1 and its high-level expression in *Escherichia coli*. Agricultural and biological chemistry 54:9, 2385-2392.

Munro G. K., McHale R. H., Saul D. J., Reeves R. A. and Bergquist P. L. (1995) A gene encoding a thermophilic alkaline serine proteinase from *Thermus* sp. strain Rt41A and its expression in *Escherichia coli*. Microbiology 141:7, 1731-1738.

Ohara-Nemoto Y., Rouf S. M., Naito M., Yanase A., Tetsuo F., Ono T., et al. (2014) Identification and characterization of prokaryotic dipeptidyl-peptidase 5 from *Porphyromonas gingivalis*. The Journal of Biological Chemistry 289:9, 5436-5448.

Pack S. P., Kang T. J. and Yoo Y. J. (2013) Protein thermostabilizing factors: high relative occurrence of amino acids, residual properties, and secondary structure type in different residual state. Applied Biochemistry and Biotechnology 171:5, 1212-1226.

Pack S. P. and Yoo Y. J. (2004) Protein thermostability: structure-based difference of amino acid between thermophilic and mesophilic proteins. Journal of Biotechnology 111:3, 269-277.

Panja A. S., Bandopadhyay, B. and Maiti, S. (2015) Protein thermostability is owing to their preferences to non-polar smaller volume amino acids, variations in residual physico-chemical properties and more salt-bridges. PLoS ONE 10:7, e0131495.

Patel S. (2017) A critical review on serine protease: key immune manipulator and pathology mediator. Allergologia et Immunopathologia 45:6, 579-591.

Pellegrini M., Marcotte E. M., Thompson M. J., Eisenberg D. and Yeates T. O. (1999) Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. Proceedings of the National Academy of Sciences 96:8, 4285-4288.

Pombejra S. N., Jamklang M., Uhrig J. P., Vu K. and Gelli A. (2018) The structure-function analysis of the Mpr1 metalloprotease determinants of activity during migration of fungal cells across the blood-brain barrier. PLoS ONE 13:8, e0203020.

Ramadurai L., Lockwood K. J., Lockwood J., Nadakavukaren M. J. and Jayaswal R. K. (1999) Characterization of a chromosomally encoded glycylglycine endopeptidase of *Staphylococcus aureus*. Microbiology 145:4, 801-808.

Ramakrishnan V., Thambidurai Y., Rajasekharan S. K. and Mohanvel S. K. (2017) Partial characterization and cloning of protease from *Bacillus*. Asian Journal of Pharmaceutical and Clinical Research 10:10, 187-191.

Ran L. Y., Su H. N., Zhou M. Y., Wang L., Chen X. L., Xie B. B., et al. (2014) Characterization of a novel subtilisin-like protease myroicolsin from deep sea bacterium *Myroides profundi* D25 and molecular insight into its collagenolytic mechanism. The Journal of Biological Chemistry 289:9, 6041-6053.

Raveendran S., Parameswaran B., Beevi Ummalyma S., Abraham A., Kuruvilla Mathew A., Madhavan A., et al. (2018) Applications of microbial enzymes in food industry. Food Technology and Biotechnology 56:1, 16-30.

Rawlings N. D., Barrett A. J., Thomas P. D., Huang X., Bateman A. and Finn R.D. (2017) The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Research 46:D1, D624-D632.

Rawlings N. D., Morton F. R., Kok C. Y., Kong J. and Barrett A. J. (2007) MEROPS: the peptidase database. Nucleic Acids Research 36:suppl_1, D320-D325.

Razvi A. and Scholtz J. M. (2006) Lessons in stability from thermophilic proteins. Protein Science 15:7, 1569-1578.

Razzaq A., Shamsi S., Ali A., Ali Q., Sajjad M., Malik A. and Ashraf M. (2019) Microbial proteases applications. Frontiers in Bioengineering and Biotechnology 7:110.

Reddi R., Arya T., Kishor C., Gumpena R., Ganji R. J., Bhukya S., et al. (2014) Selective targeting of the conserved active site cysteine of *Mycobacterium tuberculosis* methionine aminopeptidase with electrophilic reagents. The FEBS Journal 281:18, 4240-4248.

Russell R. J., Hough D. W., Danson M. J. and Taylor G. L. (1994) The crystal structure of citrate synthase from the thermophilic archaeon, *Thermoplasma acidophilum*. Structure 2:12, 1157-1167.

Sadeghi M., Naderi-Manesh H., Zarrabi M. and Ranjbar B. (2006) Effective factors in thermostability of thermophilic proteins. Biophysical Chemistry 119:3, 256-270.

Salarizadeh N., Hasannia S., Akbari Noghabi K. and Hassan Sajedi R. (2014) Purification and characterization of 50 kDa extracellular metalloprotease from *Serratia* sp. ZF03. Iranian Journal of Biotechnology12:3, 18-27.

Saul D. J., Williams L. C., Toogood H. S., Daniel R. M. and Bergquist P. L. (1996) Sequence of the gene encoding a highly thermostable neutral proteinase from *Bacillus* sp. strain EA1: expression in *Escherichia coli* and characterisation. Biochimica et Biophysica Acta 1308:1, 74-80.

Savijoki K. and Palva A. (2000) Purification and molecular characterization of a tripeptidase (PepT) from *Lactobacillus helveticus*. Applied and Environmental Microbiology 66:2, 794-800.

Seo J. M., Ji G. E., Cho S. H., Park M. S. and Lee H. J. (2007) Characterization of a *Bifidobacterium longum* BORI dipeptidase belonging to the U34 family. Applied and Environmental Microbiology 73:17, 5598-5606.

Sharipova M., Balaban N., Kayumov A., Kirillova Y., Mardanova A., Gabdrakhmanova L., et al. (2008) The expression of the serine proteinase gene of *Bacillus intermedius* in *Bacillus subtilis*. Microbiological Research 163:1, 39-50.

Simossis V. A. and Heringa J. (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. Nucleic Acids Research 33:suppl_2, W289-W294.

Souza P. M. D., Bittencourt, M. L. D. A., Caprara, C. C., Freitas, M. D., Almeida, R. P. C. D., Silveira, D., et al. (2015) A biotechnology perspective of fungal proteases. Brazilian Journal of Microbiology 46:2, 337-346.

Szilágyi A. and Závodszky P. (2000) Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. Structure 8:5, 493-504.

Takami H., Kobayashi T., Aono R. and Horikoshi K. (1992) Molecular cloning, nucleotide sequence and expression of the structural gene for a thermostable alkaline protease from *Bacillus* sp. no. AH-101. Applied Microbiology and Biotechnology 38:1, 101-108.

Tavano O. L., Berenguer-Murcia A., Secundo F. and Fernandez-Lafuente R. (2018) Biotechnological applications of proteases in food technology. Comprehensive Reviews in Food Science and Food Safety 17:2, 412-436.

Taylor T. J. and Vaisman I. I. (2010) Discrimination of thermophilic and mesophilic proteins. BMC Structural Biology10:1, S5.

Tekaia F., Yeramian E. and Dujon B. (2002) Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: a global picture with correspondence analysis. Gene 297:1-2, 51-60.

Tomazic S. J. and Klibanov A. M. (1988) Mechanisms of irreversible thermal inactivation of *Bacillus* alpha-amylases. Journal of Biological Chemistry 263:7, 3086-3091.

Tran L., Wu X. and Wong S. (1991) Cloning and expression of a novel protease gene encoding an extracellular neutral protease from *Bacillus subtilis*. Journal of Bacteriology 173:20, 6364-6372.

Tsuchiya K., Ikeda I., Tsuchiya T. and Kimura T. (1997) Cloning and expression of an intracellular alkaline protease gene from alkalophilic *Thermoactinomyces* sp. HS682. Bioscience, Biotechnology, and Biochemistry 61:2, 298-303.

Van den Burg B., Enequist H., Van der Haar M., Eijsink V., Stulp B. and Venema G. (1991) A highly thermostable neutral protease from *Bacillus caldolyticus*: cloning and expression of the gene in *Bacillus subtilis* and characterization of the gene product. Journal of Bacteriology 173:13, 4107-4115.

Vecerek B. and Kyslik P. (1995) Cloning and sequencing of the neutral protease-encoding gene from a thermophilic strain of *Bacillus* sp. Gene 158:1, 147-148.

Vogt G., Woell S. and Argos P. (1997) Protein thermal stability, hydrogen bonds, and ion pairs. Journal of Molecular Biology 269:4, 631-643.

Wakarchuk W. W., Sung W. L., Campbell R. L., Cunningham A., Watson D. C. and Yaguchi M. (1994) Thermostabilization of the *Bacillus circulans* xylanase by the introduction of disulfide bonds. Protein Engineering, Design and Selection 7:11, 1379-1386.

Walker N. D., McEwan N. R. and Wallace R. J. (2003) Cloning and functional expression of dipeptidyl peptidase IV from the ruminal bacterium *Prevotella albensis* M384T. Microbiology 149:8, 2227-2234.

Wang S. L., Wang C. Y. and Huang T. Y. (2008) Microbial reclamation of squid pen for the production of a novel extracellular serine protease by *Lactobacillus paracasei* subsp *paracasei* TKU012. Bioresource Technology 99:9, 3411-3417.

Wani A. H., Sharma M., Salwan R., Singh G., Chahota R. and Verma S. (2016) Cloning, expression, and functional characterization of serine protease Aprv2 from virulent isolate *Dichelobacter nodosus* of Indian origin. Applied Biochemistry and Biotechnology 180:3, 576-587.

Willard L., Ranjan A., Zhang H., Monzavi H., Boyko R. F., Sykes B. D., et al. (2003) VADAR: a web server for quantitative evaluation of protein structure quality. Nucleic Acids Research 31:13, 3316-3319.

Wu J., Bian Y., Tang B., Chen X., Shen P. and Peng Z. (2004) Cloning and analysis of WF146 protease, a novel thermophilic subtilisin-like protease with four inserted surface loops. FEMS Microbiology Letters 230:2, 251-258.

Xin Y., Sun Z., Chen Q., Wang J., Wang Y., Luogong L., et al. (2015) Purification and characterization of a novel extracellular thermostable alkaline protease from *Streptomyces* sp. M30. Journal of Microbiology and Biotechnology 25:11, 1944-1953.

Xu Z., Liu Y., Yang Y., Jiang W., Arnold E. and Ding J. (2003) Crystal structure of D-hydantoinase from *Burkholderia pickettii* at a resolution of 2.7 Angstroms: insights into the molecular basis of enzyme thermostability. Journal of Bacteriology 185:14, 4038-4049.

Yamagata Y. and Ichishima E. (1995) A new alkaline serine protease from alkalophilic *Bacillus* sp.: cloning, sequencing, and characterization of an intracellular protease. Current Microbiology 30:6, 357-366.

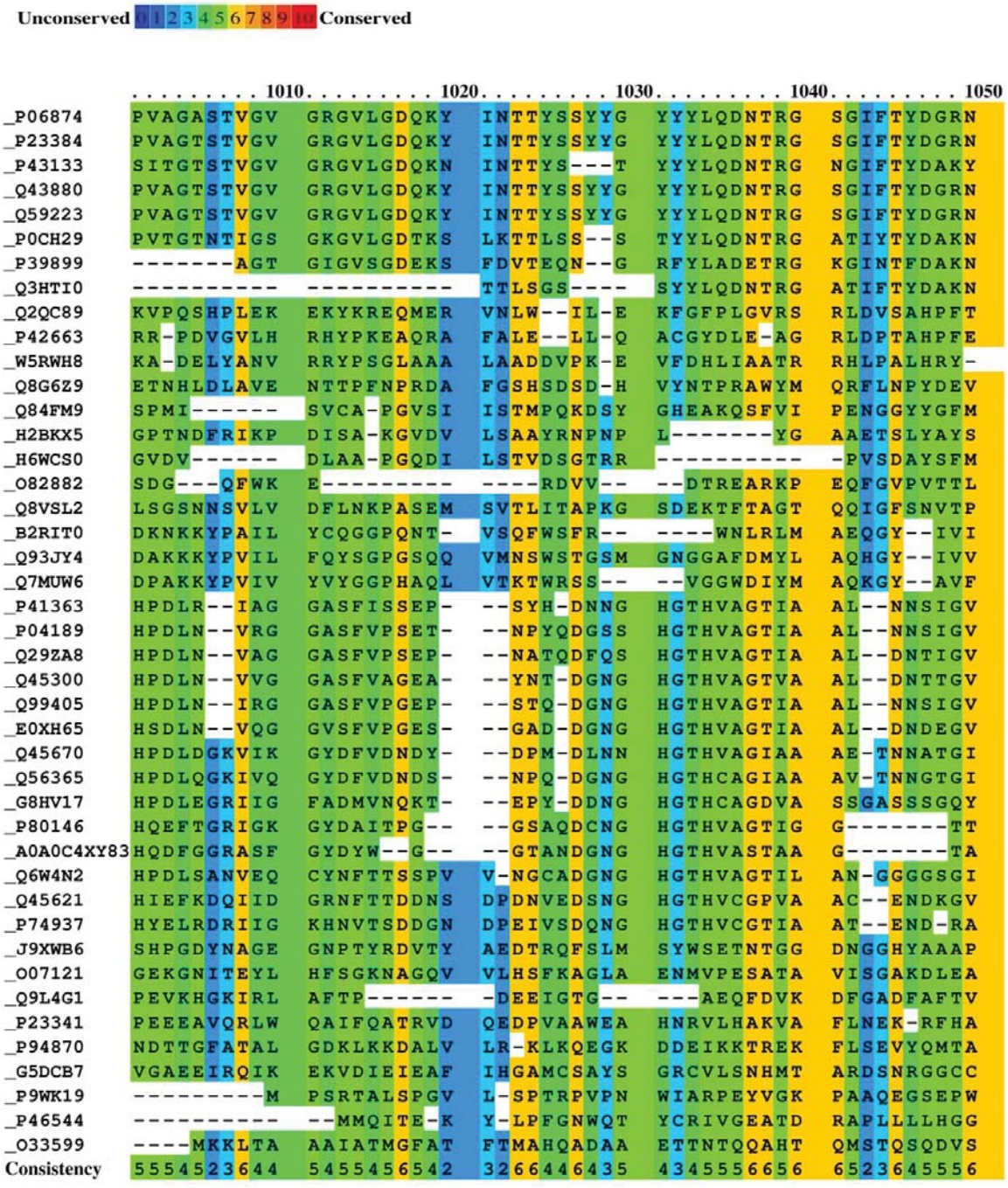
Zhang X., Chen S., Hu Z., Zhang L. and Wang H. (2009) Expression and characterization of two functional methionine aminopeptidases from *Mycobacterium tuberculosis* H37Rv. Current Microbiology 59:5, 520-525.

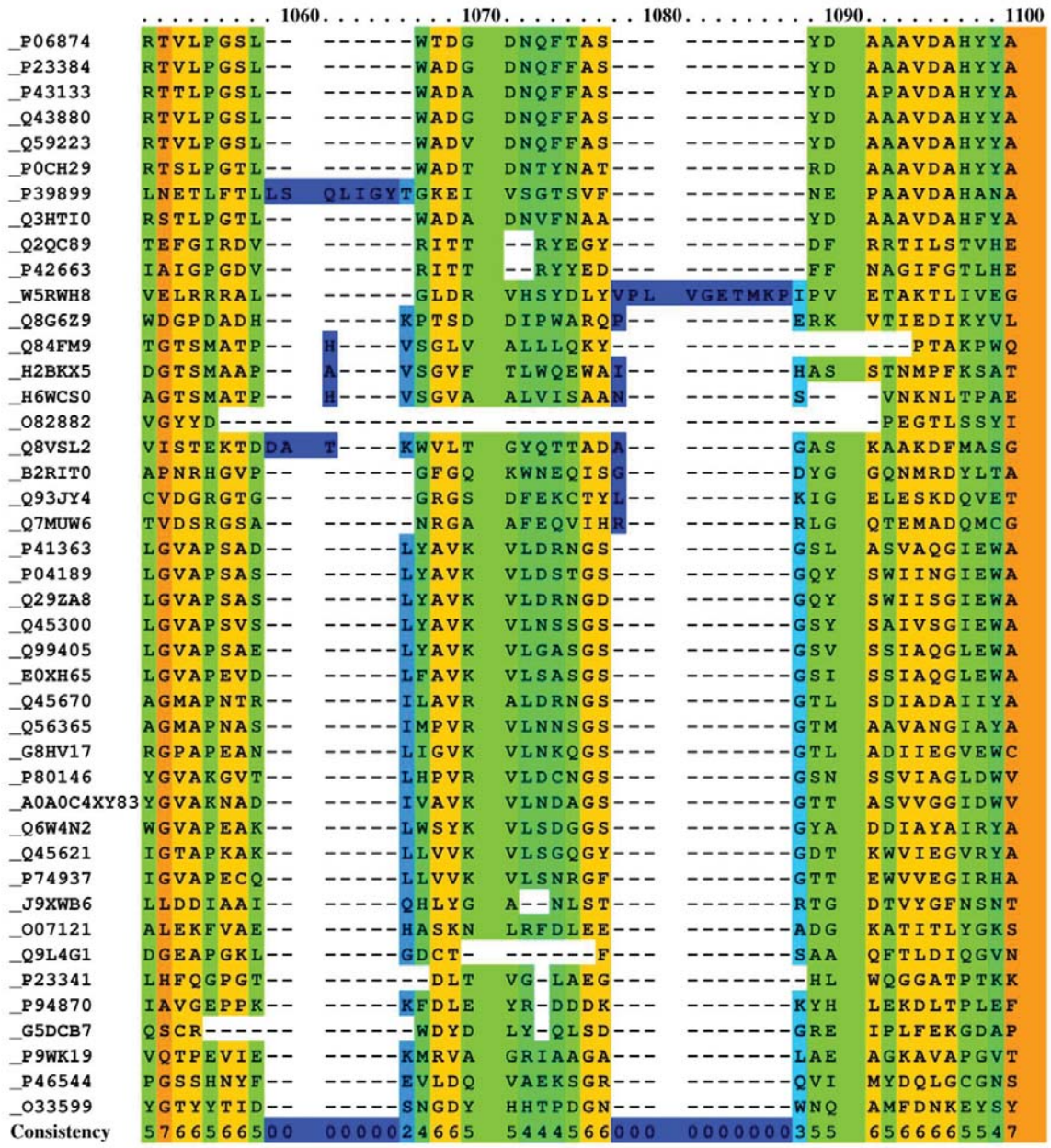
Zhou X. X., Wang Y. B., Pan Y. J. and Li W. F. (2008) Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. Amino Acids 34:1, 25-33.

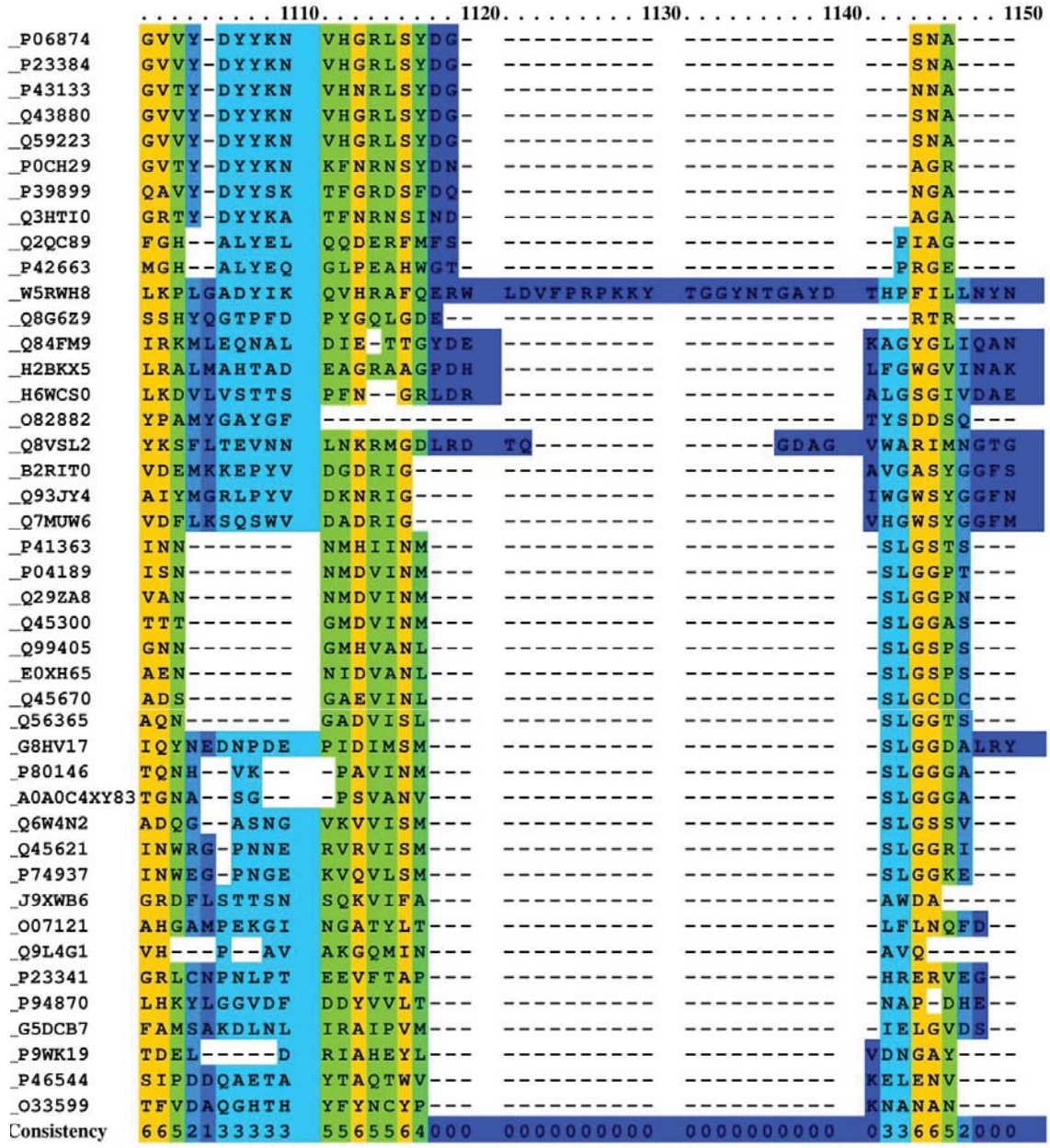
Open Access Statement:
This is an open access article distributed under the Creative Commons Attribution License (CC-BY), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

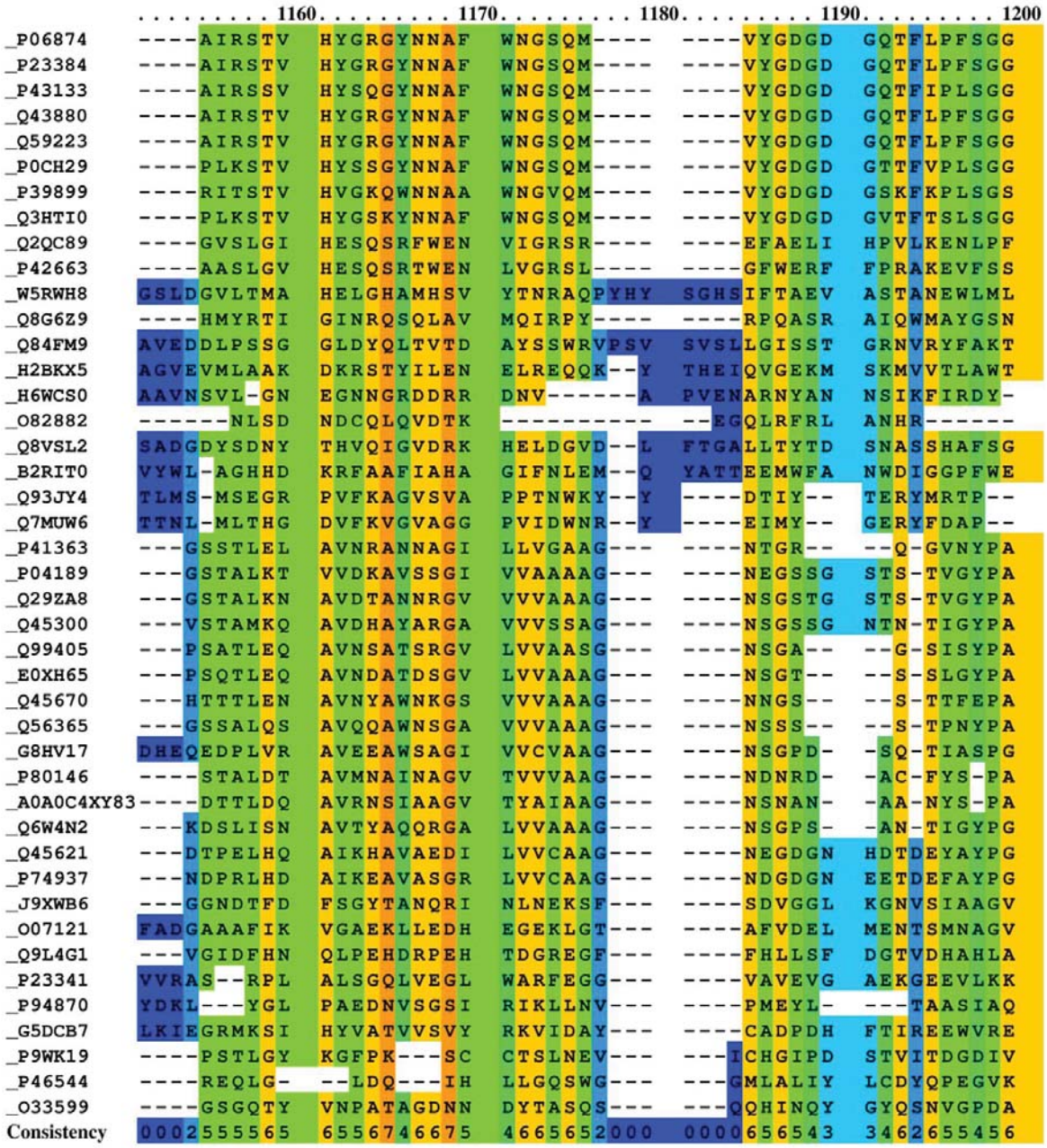
Supplementary Materials:

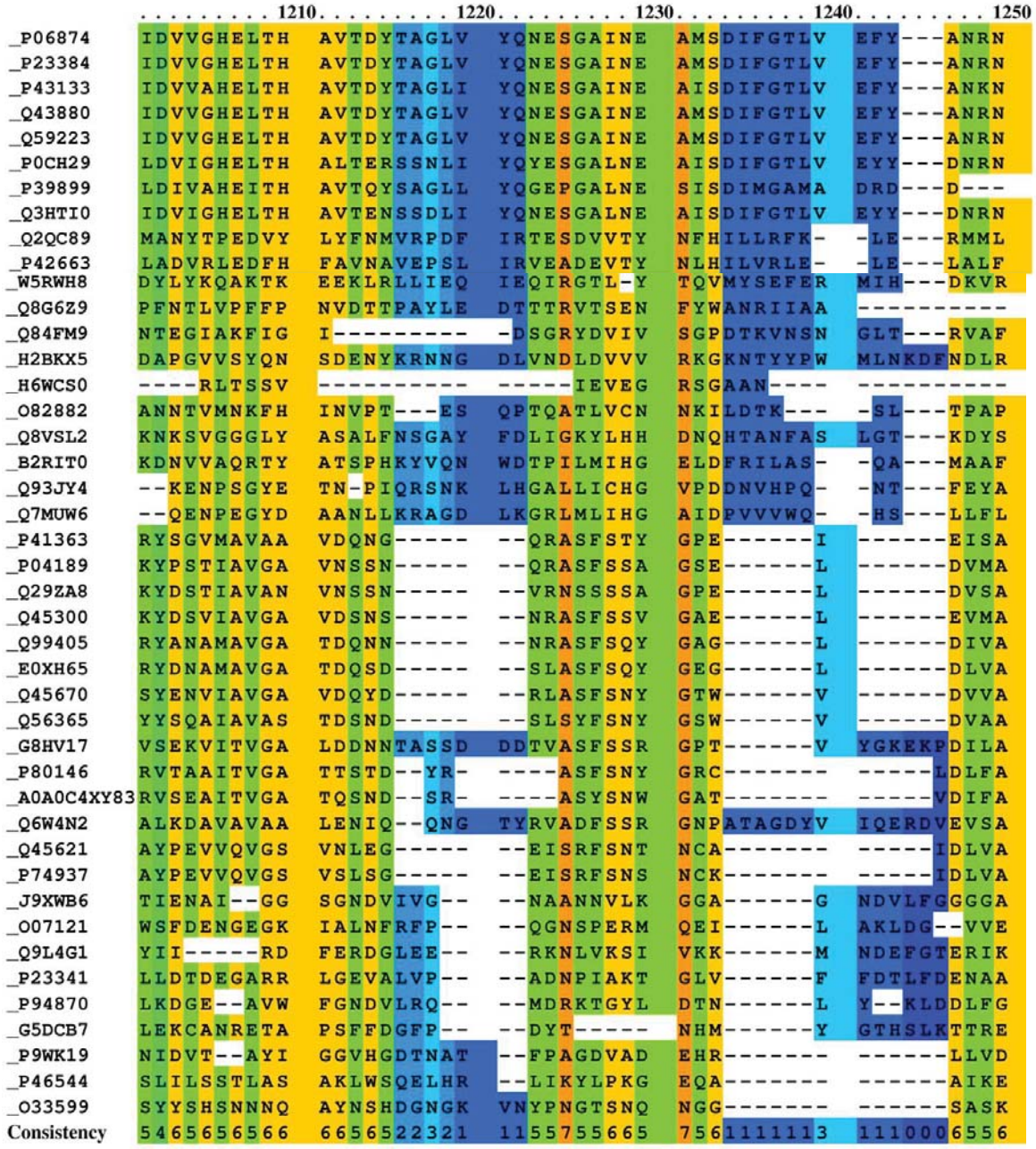
Supplementary file 1

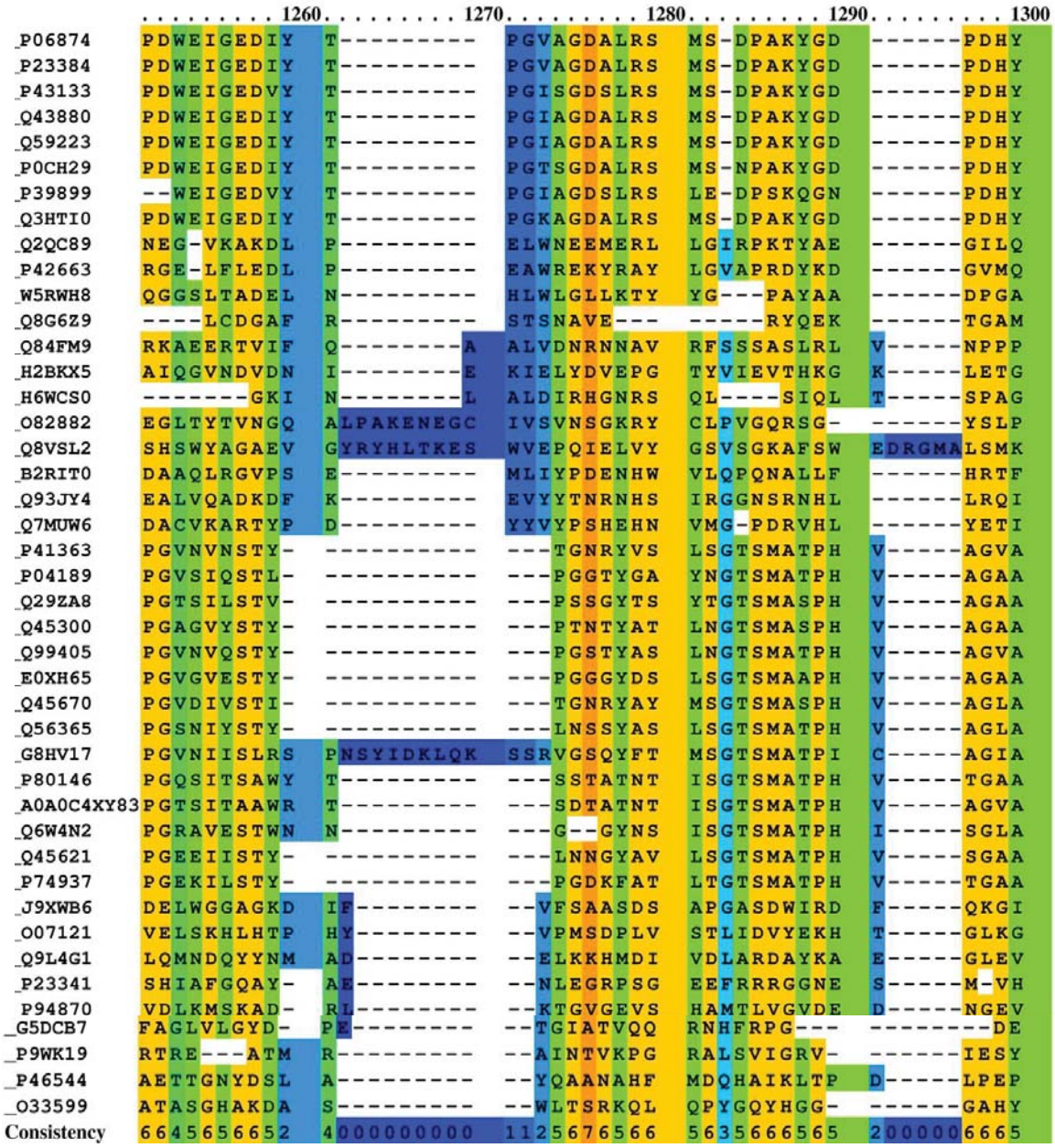



1060
1070
1080
1090
1100
_P06874
_P23384
_P43133
_Q43880
_Q59223
_P0CH29
_P39899
_Q3HTI0
_Q2QC89
_P42663
_W5RWH8
_Q8G6Z9
_Q84FM9
_H2BKX5
_H6WCS0
_O82882
_Q8VSL2
_B2RIT0
_Q93JY4
_Q7MUW6
_P41363
_P04189
_Q29ZA8
_Q45300
_Q99405
_E0XH65
_Q45670
_Q56365
_G8HV17
_P80146
_A0A0C4XY83
_Q6W4N2
_Q45621
_P74937
_J9XWB6
_O07121
_Q9L4G1
_P23341
_P94870
_G5DCB7
_P9WK19
_P46544
_O33599
Consistency


1110
1120
1130
1140
1150
_P06874 GVVY-DYYKN VHGRLSYDG- ---SNA----
_P23384 GVVY-DYYKN VHGRLSYDG- ---SNA----
_P43133 GVTY-DYYKN VHNRLSYDG- ---NNA----
_Q43880 GVVY-DYYKN VHGRLSYDG- ---SNA----
_Q59223 GVVY-DYYKN VHGRLSYDG- ---SNA----
_P0CH29 GVTY-DYYKN KFNRNSYDN- ---AGR----
_P39899 QAVY-DYYSK TFGRDSFDQ- ---NGA----
_Q3HTI0 GRTY-DYYKA TFNRNSIND- ---AGA----
_Q2QC89 FGH--ALYEL QQDERFMFS- --PIAG----
_P42663 MGH--ALYEQ GLPEAHWGT- --PRGE----
_W5RWH8 LKPLGADYIK QVHRAFQERW LDVFPRPKKY TGGYNTGAYD THPFILLNYN
_Q8G6Z9 SSHYQGTPFD PYGQLGDE-- ---RTR----
_Q84FM9 IRKMLEQNAL DIE-TTGYDE KAGYGLIQAN
_H2BKX5 LRALMAHTAD EAGRAAGPDH LFGWGVINAK
_H6WCS0 LKDVLVSTTS PFN--GRLDR ALGSGIVDAE
_O82882 YPAMYGAYGF TYSDDSQ---
_Q8VSL2 YKSFLTEVNN LNKRMGDLRD TQ------ ------GDAG VWARIMNGTG
_B2RIT0 VDEMKKEPYV DGDRIG---- AVGASYGGFS
_Q93JY4 AIYMGRLPYV DKNRIG---- IWGWSYGGFN
_Q7MUW6 VDFLKSQSWV DADRIG---- VHGWSYGGFM
_P41363 INN------- NMHIINM--- -SLGSTS---
_P04189 ISN------- NMDVINM--- -SLGGPT---
_Q29ZA8 VAN------- NMDVINM--- -SLGGPN---
_Q45300 TTT------- GMDVINM--- -SLGGAS---
_Q99405 GNN------- GMHVANL--- -SLGSPS---
_E0XH65 AEN------- NIDVANL--- -SLGSPS---
_Q45670 ADS------- GAEVINL--- -SLGCDC---
_Q56365 AQN------- GADVISL--- -SLGGTS---
_G8HV17 IQYNEDNPDE PIDIMSM--- -SLGGDALRY
_P80146 TQNH--VK-- -PAVINM--- -SLGGGA---
_A0A0C4XY83 TGNA--SG-- -PSVANV--- -SLGGGA---
_Q6W4N2 ADQG--ASNG VKVVISM--- -SLGSSV---
_Q45621 INWRG-PNNE RVRVISM--- -SLGGRI---
_P74937 INWEG-PNGE KVQVLSM--- -SLGGKE---
_J9XWB6 GRDFLSTTSN SQKVIFA--- -AWDA-----
_O07121 AHGAMPEKGI NGATYLT--- -LFLNQFD--
_Q9L4G1 VH---P--AV AKGQMIN--- -AVQ------
_P23341 GRLCNPNLPT EEVFTAP--- -HRERVEG--
_P94870 LHKYLGGVDF DDYVVLT--- -NAP-DHE--
_G5DCB7 FAMSAKDLNL IRAIPVM--- -IELGVDS--
_P9WK19 TDEL-----D RIAHEYL--- VDNGAY----
_P46544 SIPDDQAETA YTAQTWV--- KELENV----
_O33599 TFVDAQGHTH YFYNCYP--- KNANAN----
Consistency 6652133333 5565564000 0000000000 0000000000 0336652000


1160
1170
1180
1190
1200
_P06874 ----AIRSTV HYGRGYNNAF WNGSQM---- ----VYGDGD GQTFLPFSGG
_P23384 ----AIRSTV HYGRGYNNAF WNGSQM---- ----VYGDGD GQTFLPFSGG
_P43133 ----AIRSSV HYSQGYNNAF WNGSQM---- ----VYGDGD GQTFIPLSGG
_Q43880 ----AIRSTV HYGRGYNNAF WNGSQM---- ----VYGDGD GQTFLPFSGG
_Q59223 ----AIRSTV HYGRGYNNAF WNGSQM---- ----VYGDGD GQTFLPFSGG
_P0CH29 ----PLKSTV HYSSGYNNAF WNGSQM---- ----VYGDGD GTTFVPLSGG
_P39899 ----RITSTV HVGKQWNNAA WNGVQM---- ----VYGDGD GSKFKPLSGS
_Q3HTI0 ----PLKSTV HYGSKYNNAF WNGSQM---- ----VYGDGD GVTFTSLSGG
_Q2QC89 ----GVSLGI HESQSRFWEN VIGRSR---- ----EFAELI HPVLKENLPF
_P42663 ----AASLGV HESQSRTWEN LVGRSL---- ----GFWERF FPRAKEVFSS
_W5RWH8 GSLDGVLTMA HELGHAMHSV YTNRAQPYHY SGHSIFTAEV ASTANEWLML
_Q8G6Z9 ----HMYRTI GINRQSQLAV MQIRPY---- ----RPQASR AIQWMAYGSN
_Q84FM9 AVEDDLPSSG GLDYQLTVTD AYSSWRVPSV SVSLLGISST GRNVRYFAKT
_H2BKX5 AGVEVMLAAK DKRSTYILEN ELREQQK--Y THEIQVGEKM SKMVVTLAWT
_H6WCS0 AAVNSVL-GN EGNNGRDDRR DNV------A PVENARNYAN NSIKFIRDY-
_O82882 ------NLSD NDCQLQVDTK ---------- --EGQLRFRL ANHR------
_Q8VSL2 SADGDYSDNY THVQIGVDRK HELDGVD--L FTGALLTYTD SNASSHAFSG
_B2RIT0 VYWL-AGHHD KRFAAFIAHA GIFNLEM--Q YATTEEMWFA NWDIGGPFWE
_Q93JY4 TLMS-MSEGR PVFKAGVSVA PPTNWKY--Y ----DTIY-- TERYMRTP--
_Q7MUW6 TTNL-MLTHG DVFKVGVAGG PVIDWNR--Y ----EIMY-- GERYFDAP--
_P41363 ---GSSTLEL AVNRANNAGI LLVGAAG--- ----NTGR-- --Q-GVNYPA
_P04189 ---GSTALKT VVDKAVSSGI VVAAAAG--- ----NEGSSG STS-TVGYPA
_Q29ZA8 ---GSTALKN AVDTANNRGV VVVAAAG--- ----NSGSTG STS-TVGYPA
_Q45300 ---VSTAMKQ AVDHAYARGA VVVSSAG--- ----NSGSSG NTN-TIGYPA
_Q99405 ---PSATLEQ AVNSATSRGV LVVAASG--- ----NSGA-- --G-SISYPA
_E0XH65 ---PSQTLEQ AVNDATDSGV LVVAAAG--- ----NSGT-- --S-SLGYPA
_Q45670 ---HTTTLEN AVNYAWNKGS VVVAAAG--- ----NNGS-- --S-TTFEPA
_Q56365 ---GSSALQS AVQQAWNSGA VVVAAAG--- ----NSSS-- --S-TPNYPA
_G8HV17 DHEQEDPLVR AVEEAWSAGI VVCVAAG--- ----NSGPD- -SQ-TIASPG
_P80146 ----STALDT AVMNAINAGV TVVVAAG--- ----NDNRD- -AC-FYS-PA
_A0A0C4XY83 ----DTTLDQ AVRNSIAAGV TYAIAAG--- ----NSNAN- -AA-NYS-PA
_Q6W4N2 ---KDSLISN AVTYAQQRGA LVVAAAG--- ----NSGPS- -AN-TIGYPG
_Q45621 ---DTPELHQ AIKHAVAEDI LVVCAAG--- ----NEGDGN HDTDEYAYPG
_P74937 ---NDPRLHD AIKEAVASGR LVVCAAG--- ----NDGDGN EETDEFAYPG
_J9XWB6 ---GGNDTFD FSGYTANQRI NLNEKSF--- ----SDVGGL KGNVSIAAGV
_O07121 FADGAAAFIK VGAEKLLEDH EGEKLGT--- ----AFVDEL MENTSMNAGV
_Q9L4G1 ---VGIDFHN QLPEHDRPEH TDGREGF--- ----FHLLSF DGTVDHAHLA
_P23341 VVRAS--RPL ALSGQLVEGL WARFEGG--- ----VAVEVG AEKGEEVLKK
_P94870 YDKL---YGL PAEDNVSGSI RIKLLNV--- ----PMEYL- ---TAASIAQ
_G5DCB7 LKIEGRMKSI HYVATVVSVY RKVIDAY--- ----CADPDH FTIREEWVRE
_P9WK19 ----PSTLGY KGFPK---SC CTSLNEV--- ---ICHGIPD STVITDGDIV
_P46544 ----REQLG- --LDQ---IH LLGQSWG--- ---GMLALIY LCDYQPEGVK
_O33599 ----GSGQTY VNPATAGDNN DYTASQS--- ---QQHINQY GYQSNVGPDA
Consistency 0002555565 6556746675 4665652000 0000656543 3462655456


1210
1220
1230
1240
1250
_P06874 IDVVGHELTH AVTDYTAGLV YQNESGAINE AMSDIFGTLV EFY---ANRN
_P23384 IDVVGHELTH AVTDYTAGLV YQNESGAINE AMSDIFGTLV EFY---ANRN
_P43133 IDVVAHELTH AVTDYTAGLI YQNESGAINE AISDIFGTLV EFY---ANKN
_Q43880 IDVVGHELTH AVTDYTAGLV YQNESGAINE AMSDIFGTLV EFY---ANRN
_Q59223 IDVVGHELTH AVTDYTAGLV YQNESGAINE AMSDIFGTLV EFY---ANRN
_P0CH29 LDVIGHELTH ALTERSSNLI YQYESGALNE AISDIFGTLV EYY---DNRN
_P39899 LDIVAHEITH AVTQYSAGLL YQGEPGALNE SISDIMGAMA DRD---D---
_Q3HTI0 IDVIGHELTH AVTENSSDLI YQNESGALNE AISDIFGTLV EYY---DNRN
_Q2QC89 MANYTPEDVY LYFNMVRPDF IRTESDVVTY NFHILLRFK- -LE---RMML
_P42663 LADVRLEDFH FAVNAVEPSL IRVEADEVTY NLHILVRLE- -LE---LALF
_W5RWH8 DYLYKQAKTK EEKLRLLIEQ IEQIRGTL-Y TQVMYSEFER MIH---DKVR
_Q8G6Z9 PFNTLVPFFP NVDTTPAYLE DTTTRVTSEN FYWANRIIAA ----------
_Q84FM9 NTEGIAKFIG I--------- -DSGRYDVIV SGPDTKVNSN GLT---RVAF
_H2BKX5 DAPGVVSYQN SDENYKRNNG DLVNDLDVVV RKGKNTYYPW MLNKDFNDLR
_H6WCS0 ----RLTSSV ---------- -----IEVEG RSGAAN---- ----------
_O82882 ANNTVMNKFH INVPT---ES QPTQATLVCN NKILDTK--- -SL---TPAP
_Q8VSL2 KNKSVGGGLY ASALFNSGAY FDLIGKYLHH DNQHTANFAS LGT---KDYS
_B2RIT0 KDNVVAQRTY ATSPHKYVQN WDTPILMIHG ELDFRILAS- -QA---MAAF
_Q93JY4 --KENPSGYE TN-PIQRSNK LHGALLICHG VPDDNVHPQ- -NT---FEYA
_Q7MUW6 --QENPEGYD AANLLKRAGD LKGRLMLIHG AIDPVVVWQ- -HS---LLFL
_P41363 RYSGVMAVAA VDQNG----- --QRASFSTY GPE------I ------EISA
_P04189 KYPSTIAVGA VNSSN----- --QRASFSSA GSE------L ------DVMA
_Q29ZA8 KYDSTIAVAN VNSSN----- --VRNSSSSA GPE------L ------DVSA
_Q45300 KYDSVIAVGA VDSNS----- --NRASFSSV GAE------L ------EVMA
_Q99405 RYANAMAVGA TDQNN----- --NRASFSQY GAG------L ------DIVA
_E0XH65 RYDNAMAVGA TDQSD----- --SLASFSQY GEG------L ------DLVA
_Q45670 SYENVIAVGA VDQYD----- --RLASFSNY GTW------V ------DVVA
_Q56365 YYSQAIAVAS TDSND----- --SLSYFSNY GSW------V ------DVAA
_G8HV17 VSEKVITVGA LDDNNTASSD DDTVASFSSR GPT------V YGKEKPDILA
_P80146 RVTAAITVGA TTSTD--YR- ----ASFSNY GRC------- -----LDLFA
_A0A0C4XY83 RVSEAITVGA TQSND--SR- ----ASYSNW GAT------- -----VDIFA
_Q6W4N2 ALKDAVAVAA LENIQ--QNG TYRVADFSSR GNPATAGDYV IQERDVEVSA
_Q45621 AYPEVVQVGS VNLEG----- --EISRFSNT NCA------- -----IDLVA
_P74937 AYPEVVQVGS VSLSG----- --EISRFSNS NCK------- -----IDLVA
_J9XWB6 TIENAI--GG SGNDVIVG-- --NAANNVLK GGA------G NDVLFGGGGA
_O07121 WSFDENGEGK IALNFRFP-- --QGNSPERM QEI------L AKLDG--VVE
_Q9L4G1 YII-----RD FERDGLEE-- --RKNLVKSI VKK------M NDEFGTERIK
_P23341 LLDTDEGARR LGEVALVP-- --ADNPIAKT GLV------F FDTLFDENAA
_P94870 LKDGE--AVW FGNDVLRQ-- --MDRKTGYL DTN------L Y--KLDDLFG
_G5DCB7 LEKCANRETA PSFFDGFP-- --DYT----- NHM------Y GTHSLKTTRE
_P9WK19 NIDVT--AYI GGVHGDTNAT --FPAGDVAD EHR------- ------LLVD
_P46544 SLILSSTLAS AKLWSQELHR --LIKYLPKG EQA------- ------AIKE
_O33599 SYYSHSNNNQ AYNSHDGNGK VNYPNGTSNQ NGG------- ------SASK
Consistency 5465656566 6656522321 1155755665 7561111113 1110006556



Multiple sequence alignment of the thermostable proteolytic enzymes. More conserved parts of the alignment have been presented from amino acid residues from 1000 to 1300