Journal of Advanced Bioinformatics Applications and Research
ISSN 0976-2604.Online ISSN 2278–6007
Vol 5, Issue3, 2014, pp215-227
http://www.bipublication.com

Comparative Analysis of B- Glucosidases Thermostability: Differences in Amino Acids Composition and Distribution among Mesostable and Thermostable B- Glucosidases

Puneet Gupta, Shalini Verma and Jyoti Vakhlu*

School of Biotechnology
University of Jammu, Jammu-06.
*Corresponding author e-mail: puneet29gupta@gmail.com

[Received 29/10/2014, Accepted-09/11/2014]

ABSTRACT

Out of various industrial enzymes, thermostable β- glucosidases are of great interest, due to their utility in conversion of lingo- cellulosic material into bioethanol. Determination of the factors responsible for thermostability in β-glucosidases by sequence and structure comparison will be useful in identification, isolation and design of novel thermostable β- glucosidases. In the present study thermostable and mesostable β-glucosidases sequences retrieved from Uniport database were subjected to in-silico physio-chemical characterisation, multiple sequence alignment, phylogenetic tree construction and motif prediction using various bioinformatic tools. Present study revealed that thermostable β- glucosidases and mesophillic β- glucosidases are similar in percentage amino acids composition, share physico-chemical characteristics and secondary structures. But thermostable β- glucosidases have low GRAVY value, low percentage of glutamine (hydrophilic amino acid), conserved amino acids Phenylalanine (Phe)- 86, Proline (Pro) – 87, 282, 396, Glutamic Acid (Glu) -102, Asparatic acid (Asp)-147, 193, 633, Tryptophan (Trp) – 222, 261, 631, 722 , Glycine (Gly) - 126, 174, 255, 524, 604, 622, Leucine (leu) – 321 and Tyrosine (Tyr)- 605) along with 8 signature amino acid sequences that are absent in the mesophillic β-glucosidases. Out of 8 motifs, one conserved motif having sequence (WG[VT][AM][TC][AS][SY][YN][QK][IL][EN]G[AE][YV][CN][ES][DE][GN]R[GT][LP]) was 100 % conserved in thermostable β- glucosidases analysed in the present study.

Key words: Thermostable β- Glucosidases, In Silico, Enzyme, Motif, Bioinformatics

1. INTRODUCTION

β-glucosidases (3.2.1.21) are glycoside hydrolases (GH) present among simple as well as complex organisms. They catalyse the hydrolysis of terminal, non-reducing β-D-glucosyl residues with the release of β-D-glucose. In addition to biodiesel[1] production, β-glucosidases are also used in wine production [2,3], biopolishing of cotton and non- denim fabrics[5], increases the aroma of certain fuits [6,7] Production of biodiesel by β-glucosidases involves saccharification of cellulosic biomass by enzymes, which usually occurs at elevated

temperatures, making thermostable β-glucosidases (designated hereafter as TB) of great industrial importance [8]. During the last few decades TB have been isolated from various organisms but microorganisms, especially thermophilic bacteria are good source of TB [9-12]. Isolation of TB from thermophiles involve tedious steps that leads to increase in cost and labour. In addition, the amount of enzyme produced by the native organism is usually low, hence not commercially viable. However sequence and structure based differentiation of TB and MB can help in determination of factors responsible for enzyme thermostability. Information about, these factors can be used for conversion of already available mesophillic β-glucosidases (designated hereafter as MB) to TB or designing new TB.

Thermostable enzymes in general are reported to have higher number of salt bridges/ion pairs & network of salt bridges[13-16] , more hydrophobic residues[17-19] , compact packing[20] , better pie cationic interactions[21-22] , greater rigidity[23-25] , more hydrogen bonds[26-28] , more helix forming residues[19] , shortened / less loops[29-30] , less thermolabile residues [31] , high dipeptite composition[13,16] , increased helical content[19] , increased proline residues[23] and different packing of external residues[32].

Studies to determine the factors responsible for thermostability in various proteins has been carried out in past few years, as a result of availability of , i) Sequence/ structure data in various data banks and ii) efficient bioinformatic tools [33-36]. To our knowledge no such work has been carried out in β-glucosidases. Out of 132 families of glycoside hydrolases ,

β -glucosidases mainly belongs to glycoside hydrolases (GH) family 1 and 3, on the basis of signature sequence conservation [37-38]. GH family 1 have characteristic (α/β) 8 barrel fold with Glu as catalytic nucleophile as well as catalytic proton donor[39-40] .GH family 3 involves retaining mechanisn with aspartate as catalytic nucleophile and glutamate as proton donor41 . There is less information on GH family 3 and so far , few 3D structures have been resolved from this family (http://www.cazy.org).

In the present study comparative analysis of TB and MB had been carried out in order to identify factors that contribute to their thermostability. This information in turn can be used to identify and design novel thermostable β-glucosidases.

2 MATERIALS AND METHODS:

2.1 Sequence source:

Twelve well characterized bacterial β-glucosidase sequences were retrieved from UNIPROT knowledgebase database. Collected sequences were categorized into TB and MB on the basis their temperature stability. MB sequences were chosen as counterparts for comparison with TB to bring out the factors responsible for thermostability in TB. Respective families were assigned to selected β-glucosidases through search in CAZY database (http://www.cazy.org) as tabulated in table 1.

2.2 Multiple sequence alignment and phylogenetic tree construction

Multiple sequence alignment and evolutionary analysis of TB and MB sequences were carried out by CLC sequence viewer workbench (http://www.clcbio.com). Phylogenetic tree was constructed through neighbourhood joining method using CLC sequence viewer workbench. (http://www.clcbio.com).

2.3 Percentage amino acid composition of TB and MB

Comparative analysis of thermostable and mesostable β-glucosidases on the basis of amino acid composition was done to get quantitative estimation of different types of amino acids present in sequences used in present study. Predict protein tool (http://www.predictprotein.org/) was used for quantitative estimation of the 20 different amino acid residues present in TB and MB sequences.

2.4 Sequence based physico-chemical characterization:

ExPASy's ProtParam tool (http://web.expasy.org/protparam) was used to detect the different physico-chemical properties of both TB and MB sequences, namely, molecular weight , theoretical isoelectric point (pI), total number of positively and negatively charged residues, extinction coefficient42 , instability index 43 , aliphatic index44 , grand average hydropathicity (GRAVY)45.

2.5 Secondary Structural analysis:

Self-Optimized Prediction Method with Alignment (SOPMA) was used to predict the secondary structural features of the selected protein sequences considered in the present study46.

2.6 Statistical analysis

An analysis of variance (ANOVA) was conducted on various physiochemical parameter variables for each study with the statistical package 'Asistat version-7.4 beta 2008'. F-tests were used to determine the statistical significance. After significant effects were detected, a T- test was applied for all pairwise comparisons of mean responses.

2.7 Motif prediction:

The protein conserved motifs were deduced by MEME (Multiple Em for Motif Elicitation) Suite (http://meme.sdsc.edu/meme). The optimum width of each motif was set within the limits: minimum=6 and maximum=50. Maximum number of motifs to be found was set to 8. Since we were interested in finding the specific motif confined to thermostable β-glucosidase sequencs where as mesostable β-glucosidase sequences were chosen as negative sequences while using MEME tool

3 RESULTS:

3.1 Sequence source:

In the present study, only those β-glucosidase sequences were selected and analysed, that have been fully characterized in wet lab vis a vis temperature stability. The number of well characterized β-glucosidase sequences is very scarce in comparison to other hydrolases and only 12 wet laboratory characterized sequences could be collected from UNIPROT knowledge database .Out of these twelve sequences , Seven sequences were categorized as TB and five as MB respectively on the basis of enzyme thermostability and were assigned families, as tabulated **in table 1.**

3.2 Comparative study of percentage amino acid composition of TB and MB:

Results of amino acid analysis of β-glucosidases indicated, presence of all the common amino acids in both TB and MB (Table 6). Glutamine (Gln) and Threonine (Thr) were 49 and 0.79 times lower in TB as compared to MB, while Lysine (Lys) was 1.87 times higher in TB. Glutamic acid (Glu), Glycine (Gly), Leucine (Leu), Alanine (Ala) were the dominant amino acids while Cysteine (Cys) was found to be least dominant in TB as well as MB and no significant difference of composition was found. Proline was 0.88 times lower in TB as compared to MB . The percentage of charge amino acid was higher in TB than MB , and among the charged amino acids especially negatively charged amino acids were more (Table 7).

3.3 Multiple sequence alignment analysis

Multiple sequence alignment of TB and MB showed position specific conservation of certain amino acids. Conserved amino acids present in TB sequences were Phenylalanine (Phe)- 86, Proline (Pro) – 87, 282, 396, Glutamic Acid (Glu) -102, Asparatic acid (Asp)-147, 193, 633, Tryptophan (Trp) – 222, 261, 631, 722, - , Glycine (Gly) - 126, 174, 255, 524, 604, 622, Leucine (leu) - 321, Tyrosine (Tyr)- 605, whereas no such position specific conservation was found in MB.(Supplementary fig 1 ,fig 2).

3.4 Phylogenetic tree:-

Phylogenetic tree divided all the sequences in three major clades i.e upper clade, middle clade and lower clade (Fig :-1). Upper clade contained

specifically TB from same genus, *Thermoanaerobacter ethanolicus (*D3Y2V4) *Thermoanaerobacter brockii* (Q60026) with temperature optima of 80C and 60C respectively. Middle clade consisted of six enzymes out of which three were TB with temperature optima of 65C,65 C ,90C produced *by Clostridium thermocellum* P14002 ,*Clostridium stercorarium* O08331, *Thermotoga neapolitana* Q60038 respectively. The other three enzymes in the clad were of MB produced by *Cellvibrio gilvus* **(**P96316), *Terrabacter ginsenosidimutans* (D2KAT1) and*Uncultured bacterium* **(**Q3HXC2) (isolated using metagenomic cloning)with temperature optima of 35C ,37C and 40C respectively. Surprisingly MB from *Cellvibrio gilvus* **(**P96316) was close to TB of *Clostridium stercorarium* (O08331) and *Thermotoga neapolitana* (Q60038) instead of other MB in the same clad . Third clade was again formed mostly of TB from *Thermobispora bispora* (P38645) ,*Thermotoga neapolitana* (B9K7M5) and *Bifidobacterium breve* (P94248), showing optimal activity at 60C , 95C and 45C respectively and a single MB produced by *Paenibacillus sp.* HC1 **(**Q2WGB4**)** with temperature optima of 37C .

3.5 Sequence based physico-chemical characterization:

Molecular weight, theoretical isoelectric point (pI), total number of positively and negatively charged residues, extinction coefficient, instability index, aliphatic index and grand average hydropathicity (GRAVY) value was calculated for all the sequences using ExPASy's ProtParam tool (Tables 2 and Table 3). TB has almost similar number of amino acid residues (444 – 755) as compared to MB (amino acid residues ranging between 448 – 854). Molecular weight of TB ranged between 51535.2 Daltons - 83900.4 Daltons where as MB ranged between 51383.8 Daltons - 91714.6 Daltons. Theoretical pI was found to be between 5.27 – 7.36 for TB and 4.44 – 7.75 for MB. Out of seven TB sequences selected, six sequences had acidic PI except one sequence which has slightly basic PI. Total number of positively charged amino acids was found to be in the range of 57 – 117 for TB but in case of MB it was found to be in the range of 61 - 123. Total number of negatively charged amino acids was found between 49-103 for TB and between 42 - 77 for MB. Instability index plays an important role in determining half life of a particular protein. Sequences showing instability index of 40 or less than 40 have higher half life time as compared to sequences showing instability index more than 40. It was found to be in the range of 29.92 – 37.14 for TB and MB showed instability index in between 26.75 – 42.37. Aliphatic index ranged between 75.30 - 93.10 for TB and 65.63 - 94.65 for MB. GRAVY measures hydropathicity of a particular protein which means lower the value of GRAVY better will be the interaction with water. GRAVY value was in the range (-0.310) to (-0.555) for TB and (-0.198) to (0.355) for MB.

3.6 Structural analysis:

Apart from amino acid composition, type of secondary structure and % of each secondary structures plays a role in thermostability as they contribute to 3D structure of a protein. Random coil was found to be dominant structure in both TB and MB. Percentage of α-helix ranges between 34.17% - 38.44% for TB but for MB, α-helix fell in the range of 32.18 % - 39.70 %. TB showed percentage of β-turn ranging between 6.67 % - 10.89 % as compared to 5.56 % - 9.84 % β-turn found in MB. Percentage of extended strands ranges between 14.38 % - 17.90 % for TB which is 11.88% - 17.02% in case of MB. Percentage of random coils ranges between 36.22% - 42.64% for TB and 37.72 % - 48.46 % for MB (Table 4 and Table 5).

3.7 Statistical analysis:

The significant differences were calculated for amino acid composition ,various physiochemical properties and secondry structure composition .The parameters that were found to be stastically significant at a level of 5% of probability (<0.05) are percentage compositions of Glutamine (Glu)

and physiochemical characteristics i.e GRAVY as shown in table 6 and table 7 respectively.

3.7 Motif prediction:

Motif prediction using bioinformatics tools can be used to characterize a particular protein sequence. A total of 8 motifs were found in TB sequences (Table 8 and Table 9). Out of various motif , motif 6 with amino acid sequence WG[VT][AM][TC][AS][SY][YN][QK][IL][EN]G [AE][YV][CN][ES][DE][GN]R[GT][LP] was found to be uniformly distributed in TB where as it was totally absent in MB.

4. DISCUSSION

Twelve wet lab characterized MB and TB sequences from Uniprot database has been analysed . β-glucosidases mainly belongs to GH1 and GH3 families as per CAZY database classification, with catalytic sites as catalytic nucleophile/base and catalytic proton donor. These sites are conserved for both TB and MB (http://www.cazy.org). Charged amino acids are known to increase thermo stability of proteins by stabilizing secondary structure of proteins through electrostatic interaction. Most of the thermostable proteins are known to have increased percentage of charged amino acids [47,48,35]. In the present study also, the % of charged amino acids was more but we found in case of TB % of negatively charged amino acid was higher when compared to positive charged.

Glutamine (Gln) a hydrophilic, thermolabile residue found 49 times lower in TB as compared to MB. It had also been reported earlier that the decreased occurrence of thermolabile residues can enhance thermostability in proteins [49,50,35] . Glutamine (Gln) is thermolabile amino-acids due to it's tendency to undergo deamination at higher temperature [49] . Along with Gln, Cys and Met are also known as thermolabile amino acids due to their tendency to undergo oxidation at high temperature [51]. Thermophilic proteins have lower frequency of Met and Cys compared with mesophilic proteins[13,52] . β-glucosidase A (BglA) from thermostable bacteria shows lower content of Met than do BglA and other enzymes of the family from mesophilic organisms [53] .Present study revealed in β-glucosidases % compositionis of Glutamine (Gln) is lower and statistically significant where as other thermolabile residues Cysteine (Cys) and Methionine (Met) are compararable in TB and MB. It seems that lower percentage of Gln residue in β-glucosidases contribute towards their thermostability.

Hydrophobicity in a protein leads to compact core /low residual packing /voids and increased thermostability [54,25,13] .Percentage composition of hydrophobic residues like Phenylalanine (Phe), Leucine (leu) and Tyrosine (Tyr) increase the hydrophobicity of a protein and increased hydrophobicity results in thermostability [55] .

In TB , we found that although difference in hydrophobic residues like Phenylalanine (Phe), Leucine (leu), Tyrosine (Tyr) was not statistically significant in % composition but Phenylalanine (Phe)- 86, Leucine (Leu) – 321 and Tyrosine (Tyr)- 605 were found to be conserved . We did not come across any report, wherein importance of conservation of hydrophobic residues was important for thermostability. We propose in case of TB in addition to increased proportion of hydrophobic residues, conservation of Phenylalanine (Phe)- 86, Leucine (Leu)- 321and Tyrosine (Tyr)- 605 is also an important factor for their thermostability.

Proline residue effects thermostability either by increased occurrence that effects rigidity of helices or by substitution of other amino acids with proline54,56 (Haney *et al.,* 1997; Zhou *et al.,* 2010). In present study contrary to the literature the number of proline residues in TB is slightly less than that present in MB but is 100% conserved at positions 87, 282 and 396 indicating that in case of TB , the thermostability may be effected by conservation/substitution of proline rather than by higher number. It had been reported that amino acid composition difference is only the

fundamental factor of the factors that support the protein thermostability. Other factors including additional hydrogen bonds, electrostatic interactions, hydrophobic interactions, disulfide bonds and more rigid & compact packing are important [31]. The statistical analysis comparing amino acid compositions in MB and TB indicated that the properties responsible for thermostability of proteins include higher residue volume, higher residue hydrophobicity, more charged amino acids especially Glu, Arg, and Lys. Lysine gives the most important contribution to the charge excess. In hyperthermophilic proteins Lysine (Lys) is one of the dominant residue [27]. In β-glucosidases again, lysine (Lys) residue is more in TB than MB but not the dominant one.

Lu *et al*, compared the difference of amino acid composition between 110 pairs of homologous thermophilic and mesophilic proteins and found that thermophilic proteins have higher average hydropathy and aliphatic index due to higher Leucine (Leu) composition. The present study revealed similar findings, the frequent occurrence and conservation of Leucine (Leu) residues in case of TB and the value for GRAVY (grand average of hydropathicity) was found to be less in case of TB than MB.

Secondry structure comparison showed random coils followed by α-helix , were dominant structure in both TB and MB. β-turns was the least dominant structure present in both TB as well as MB. Pradeep *et al*, also found similar results in comparison of secondry structure but no statistical evaluation was done. In the present study the difference in secondary structure composition between MB and TB was found to be statistically insignificant.

Phylogenetic tree construction divided the sequences in three clads but two clads , middle and lower were had mixed sequences of TB and MB. In case of insilico analysis on lipases [35] , were not able to distinguish the thermophillic and mesophillic lipases by phylogenetic tree construction. Conserved motif WG[VT][AM][TC][AS][SY][YN][QK][IL][EN]G [AE][YV] [CN][ES][DE][GN]R[GT][LP] was found to be uniform in all thermostable TB but absent in MB.

In the present study, all the parameters checked were in corroboration with characteristics of thermostable proteins reported in literature except composition of Proline (Pro) which is slightly low in case of TB. However it is 100% conserved at positions 87, 282, 396. The proline conservation suggests that in case of β-glucosidases the thermostability may be effected by conservation/substitution of proline rather than by higher number. The results obtained were analysed statistically and it was found that % compositionis of Glutamine (Gln) is statistically significant. In addition to composition of Glutamine (Gln) and the value for GRAVY (grand average of hydropathicity) was found to be less in case of TB than MB and statstically significant also. In the literature though various factors have been reported important for thermostability using sequence data but we did come across any report where this data has been subjected to statistical analysis. The reason could be less number of sequences that are available for analysis . In case of TB we propose , high Glutamine (Gln) composition, low GRAVY value ,low proline composition and conservation of amino acids is responsible for thermostability.

5. CONCLUSION

Insilico comparison of various mesophilic and thermostable proteins has been done by the many workers in order to highlight the specific features in thermophiles that can be used to engineer the already existing mesophilic proteins to thermostable one. In the present study similar attempt was made to identify the specific factors in thermostable.

β-glucosidases in comparison to their mesophilic counterparts contributing to thermostability . The sequence and structural data available on β-glucosidases is very scarce and we have only

selected those sequences that have been characterized in the wet lab. On the basis of the limited data set we infer that though % amino acid composition , secondry structure compsosition , various physiochemical characteristics were almost similar in both TB and MB but displayed statistically significant differences in Glutamine (Gln) composition, GRAVY value . We further propose the signature sequence: WG[VT][AM][TC][AS][SY][YN][QK][IL][EN]G [AE][YV][CN][ES][DE][GN]R[GT][LP] conserved in all the TB can be further used either as a primer to fish out novel TB or for site directed mutagenesis for engineering MB to TB.

ACKNOWLEDGEMENT

We are thankful to Department of Biotechnology, New Delhi, India for financial support in the form DBT - Major project (BT/PR-11727/BCE/08/720/2008). PG and SV are also thankful to DBT for Financial support in the form of fellowship and stipend. The bioinformatic analysis carried out at Bioinformatics centre of the school funded by DBT is also acknowledged.

REFERENCES

1. Kumar R, Singh S, & Singh O V, Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives, J Ind Microbiol Biotechnol, 35 (2008) 377- 391.
2. Gunata Y Z, Bayonove C L, Cordonnier R E, Arnaud A & Galzy P, Hydrolysis of grape monoterpenyl glycosides by Candida molischiana and Candida wickerhamii β-glucosidases, J Sci Food Agric, 50 (1990) 499–506.
3. Caldini C, Bonomi F, Pifferi P G, Lanzarini G & Galante Y M, Kinetic and immobilization studies on fungal glycosidases for aroma enhancement in wine, Enzyme Microb Technol, 16 (1994) 286–291.
4. Kumar A, Lepola M & Purtell C, Enzyme finishing of man-made cellulosic fabrics, Textile Chem Colourist, 26 (1994) 25–28.
5. Galante Y M, Monteverdi R, Inama S, Caldini C, De Conti A et al, New applications of enzymes in wine making and olive oil production, Italian Biochem Soc Trans, 4 (1993) 34.
6. Pabst A, Barron D, Etievant P & Schrier P, Enzymatic hydrolysis of bound aroma constituents from raspberry fruit Pulp, J Agric Food Chem, 39 (1991)173–175.
7. Marlatt C, Ho C-T & Chien M, Studies of aroma constituents bound as glycosides in tomato, J Agric Food Chem,40 (1992) 249–252.
8. Bhatia Y, Mishra S & Bisaria V S, Microbial beta-glucosidases: cloning, properties, and applications, Crit Rev Biotechnol, 22 (2002) 375-407.
9. Park T H, Choi K W, Park C S, Lee S B, Kang H Y et al , Substrate specificity and transglycosylation catalyzed by a thermostable b-glucosidase from marine hyperthermophile Thermotoganea politana, Appl Microbiol Biotechnol, 69 (2005) 411–422.
10. Murray P, Aro N, Collins C, Grassick A, Penttil M et al , Expression in Trichoderma reesei and characterization of a thermostable family 3 b-glucosidase from the moderately thermophilic fungus Talaromyces emersonii, Protein Expr Purif, 38(2004) 248–257.
11. Yoon J J, Kim K Y &Cha C J, Purification and characterization of thermostable b-glucosidase from the brown-rot Basidiomycete Fomitopsis palustris grown on microcrystalline cellulose, J Microbiol, 46 (2008) 51–55.
12. Yang S, Jiang Z, Yan Q & Zhu H, Characterization of a thermostable extracellular b-glucosidase with activities of exoglucanase and transglycosylation from Paecilomyces thermophila, J Agric Food Chem, 56 (2008) 602–608.
13. Kumar S, Tsai C J & Nussinov R, Factors enhancing protein thermostability, Protein Eng,13 (2000) 179-191.
14. Natesh R, Manikandan K, Bhanumoorthy P, Viswamitra M A & Ramakumar S, Thermostable xylanase from Thermoascus aurantiacus at ultrahigh resolution (0.89 A) at 100 K and atomic resolution (1.11 A) at 293 K refined anisotropically to small-molecule accuracy, Acta Crystallogr D Biol Crystallogr, 59 (2003) 105-17.
15. Ihsanawati, Kumasaka T, Kaneko T, Morokuma C, Yatsunami R et al, Structural basis of the substrate subsite and the highly thermal stability of

Xylanase10B from Thermotoga maritima MSB8, Proteins, 61(2005) 999-1009.
16. Yang H M, Yao B, Meng K, Wang Y R, Bai Y G et al, Introduction of a disulfide bridge enhances the thermostability of a Streptomyces olivaceoviridis xylanase mutant. J Ind Microbiol Biotechnol, 34(2007) 213-218.
17. Spassov V Z, Karshikoff A D & Ladenstein R, The optimization of protein-solvent interactions: thermostability and the role of hydrophobic interactions, Protein Sci, 4(1995) 1516–1527.
18. Haney P J, Badger J H, Buldak G L, Reich C I, Woese C R et al, Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic Methanococcus species, Proc Natl Acad Sci USA, 96 (1999) 3578-3583.
19. Szilagyi A & Zavodszky P, Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comparative survey, Structure, 8(2000) 493–504.
20. Pack S P & Yoo Y J, Protein thermostability: structure based difference of residual properties between thermophilic and mesophilic proteins, J Mol Catal B Enzym, 26 (2003) 257–264.
21. Gromiha M M, Oobatake M & Sarai, A Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins, Biophys Chem 82 (1999) 51–67.
22. Chakravarty S & Varadarajan R, Elucidation of determinants of protein stability through genome sequence analysis, FEBS Lett, 470 (2000) 65–69.
23. Watanabe K, Hata Y, Kizaki H, Katsube Y & Suzuki Y, The refined crystal structure of Bacillus cereus oligo-1, 6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization, J Mol Biol, 269 (1997) 142–153
24. Van den Burg B, Vriend G, Veltman O R, Venema G & Eijsink V G, Engineering an enzyme to resist boiling. Proc Natl Acad Sci USA, 95 (1998) 2056-2060.
25. Bogin O, Peretz M, Hacham Y, Korkhin Y, Frolow F et al, Enhanced thermal stability of Clostridium beijerinckii alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic Thermoanaerobacter brockii alcohol dehydrogenase, Protein Sci, 7 (1998) 1156-1163.
26. Vogt G & Argos P, Protein thermal stability: hydrogen bonds or internal packing?, Fold Des. 2 (1997) S40–S46.
27. Berezovsky I N & Shakhnovich E I, Physics and evolution of thermophilic adaptation, Proc Nat Acad Sci USA, 102 (2005) 12742-12747.
28. Glyakina A V, Garbuzynskiy S O, Lobanov M Y & Galzitskaya O V, Different packing of external residues can explain differences in the thermostability of proteins from thermophilic and mesophilic organisms, Bioinformatics, 23 (2007) 2231-2238.
29. Russell R J, Hough D W, Danson M J & Garry L T, The crystal structure of citrate synthase from the thermophilic archaeon Thermoplasma acidophilum, Structure, 2(1994) 1157–1167.
30. Thompson M J & Eisenberg D, Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability, J Mol Biol, 290 (1999) 595-604.
31. Vieille C, Epting K L, Kelly R M & Zeikus J G, Bivalent cations and amino-acid composition contribute to the thermostability of Bacillus licheniformis xylose isomerase, Eur J Biochem, 268 (2001) 6291-6301.
32. Robinson-Rechavi M, Alibés A & Godzik A, Contribution of electrostatic interactions,compactness and quaternary structure to protein thermostability: lessons from structural genomics of Thermotoga maritime, J Mol Biol, 356 (2006) 547–557.
33. Gromiha M M & Suresh M X, Discrimination of mesophilic and thermophilic proteins using machine learning algorithm, Proteins, 70 (2008) 1274- 1279.
34. Coleman R G & Sharp K A, Shape and evolution of thermostable protein structure, Proteins, 78 (2009) 420-433.
35. Chakravorty D, Parameswaran S, Dubey V K & Patra S, In silico characterization of thermostable lipases, Extremophiles, 15 (2010) 89-103.
36. Sterpone F & Melchionna S, Thermophilic proteins: insight and perspective from in silico experiments,
a. Chem Soc Rev, 4 (2012) 1665-1676.

37. Henrissat B & Davies G, Structural and sequence-based classification of glycoside hydrolases, Curr Opin Struct Biol, 7 (1997) 637–644.
38. Cantarel B L, Coutinho P M, Rancurel C, Bernard T, Lombard V et al, The carbohydrate-active
39. enzymes database (CAZy), an expert resource for glycogenomics, Nucleic Acids Res, 37 (2009) D233-D238.
40. Wang Q, Trimbur D, Graham R, Warren R A & Withers S G, Identification of the acid/base catalyst in Agrobacterium faecalis beta-glucosidase by kinetic analysis of mutants, Biochemistry, 34(1995) 14554-14562.
41. Nijikken Y, Tsukada T, Igarashi K , Samejima M, Wakagi T et al, Crystal structure of intracellular family 1b-glucosidase BGL1A from the Basidiomycete Phanerochaete chrysosporium, FEBS Lett, 81(2007) 1514–1520.
42. Paal K, Ito M & Withers S G, Paenibacillus sp. TS12 glucosylceramidase, kinetic studies of a novel sub-family of family 3 glycosidases and identification of the catalytic residues, Biochem J, 378 (2004) 141–149.
43. Gill SC &Von Hippel PH, Extinction coefficient, Anal Biochem, 182 (1989) 319-328.
44. Guruprasad K, Reddy B V P & Pandit M W, Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, Prot Eng, 4 (1990) 155-161.
45. Ikai AJ, Thermostability and aliphatic index of globular proteins, J Biochem, 88 (1980) 1895-1898.
46. Kyte J & Doolottle R F, A simple method for displaying the hydropathic character of a protein, J Mol Biol, 157 (1982) 105-132.
47. Sahay A, Shakya M, In silico Analysis and Homology Modelling of Antioxidant Proteins of Spinach. J Proteomics Bioinform, 3 (2010) 148-154.
48. Fukuchi S & Nishikawa K, Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria, J Mol Biol, 309 (2001) 835–843.
49. Alsop E, Silver M & Livesay D R, Optimized electrostatic surfaces parallel increased thermostability: a structural bioinformatic analysis, Protein Eng, 16 (2003) 871–874.
50. Russell R B & Sternberg M J, Two new examples of protein structural similarities within the structure-function twilight zone, Protein Eng, 10 (1997) 333-338.
51. Trivedi S, Gehlot H S & Rao S R, Protein thermostability in Archaea and Eubacteria, Genet Mol Res, 5 (2006) 816- 827.
52. Russell R J M, Gerike U, Danson M J, Hough D W &Taylor G L, Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium, Structure 6,(1998) 351-361.
53. Xu Z, Liu Y, Yang Y, Jiang W, Arnold E et al , Crystal structure of D-Hydantoinase from Burkholderia pickettii at a resolution of 2.7 Angstroms: insights into the molecular basis of enzyme thermostability, J Bacteriol, 185 (2003) 4038–4049
54. Lopez-Camacho C, Salgado J, Lequerica J L, Madarro A, Ballestar E et al, Amino acid substitutions enhancing thermostability of Bacillus polymyxa beta-glucosidase A, Biochem J, 314 (1996) 833–838.
55. Haney P, Konisky J, Koretke K K, Luthey-Schulten Z & Wolynes P G, Structural basis for thermostability and identification of potential active site residues for adenylate kinases from the archaeal genus Methanococcus, Proteins, 28 (1997) 117-130 .
56. Tyndall J, Sinchaikul S, Fothergill-Gilmore L, Taylor P & Walkinshaw M D, Crystal structure of a thermostable lipase from Bacillus stearothermophilus P1, J Mol Biol, 323 (2002) 859–869
57. Zhou C, Xue Y & Ma Y, Enhancing the thermostability of α-glucosidase from Thermoanaerobacter tengcongensis MB4 by single proline substitution, J Biosci Bioeng, 110 (2010) 12–17.
58. Lu B, Wang G & Huang P, A comparison of amino acid composition of proteins from thermophiles and mesophiles, Wei Sheng Wu Xue Bao 38 (1998) 20–25.
59. Pradeep N V, Anupama, Vidyashree K G & Lakshmi P, In Silico Characterization of Industrial Important Cellulases using Computational Tools, Adv Sci Tech, (2010) Vol 4.
60. Yernool D A, McCarthy J K, Eveleigh D E & Bok J D, Cloning and characterization of the

glucooligosaccharide catabolic pathway beta-glucan glucohydrolase and cellobiose phosphorylase in the marine hyperthermophile Thermotoga neapolitana, J Bacteriol, 182 (2000) 5172-5179.

61. Wright R M, Yablonsky M D, Shalita Z P, Goyal A K & Eveleigh D E, Cloning, characterization, and nucleotide sequence of a gene encoding Microbispora bispora BglB, a thermostable beta-glucosidase expressed in Escherichia coli, Appl Environ Microbiol, 58 (1992) 3455-3465.
62. Breves R, Bronnenmeier K, Wild N, Staudenbauer W L & Hofemeister J, Genes encoding two different beta-glucosidases of Thermoanaerobacter brockii are clustered in a common operon , Appl Environ Microbiol, 63 (1997) 3902-3910.
63. Bronnenmeier K & Staudenbauer W, Purification and properties of an extracellular beta-glucosidase from the cellulolytic thermophile Clostridium stercorarium. Appl Microbiol Biotechnol, 28 (1988) 380-386.
64. Song X , Xue Y, Wang Q & Wu X, Comparison of three thermostable beta-glucosidases for application in the hydrolysis of soybean isoflavone glycosides, J Agric Food Chem, 59 (2011) 1954-1961.
65. Graebnitz F, Ruecknagel K P, Seiss M & Staudenbauer W L, Nucleotide sequence of the Clostridium thermocellum bgIB gene encoding thermostable beta-glucosidase B: homology to fungal beta-glucosidases, Mol Gen Genet, 217 (1989) 70-76.
66. Zverlov V V, Volkov I Y, Velikodvorskaya T V & Schwarz W H, Thermotoga neapolitana bglB gene, upstream of lamA, encodes a highly thermostable beta-glucosidase that is a laminaribiase, Microbiology, 143 (1997) 3537-3542.
67. Harada K M, Tanaka K, Fukuda Y, Hashimoto W, Murata K et al, Degradation of rice bran hemicellulose by Paenibacillus sp. strain HC1: gene cloning, characterization and function of beta-D-glucosidase as an enzyme involved in degradation, Arch Microbiol, 184 (2005) 215-224
68. An D S, Cui C H, Lee H G, Wang L, Kim S C et al, Identification and characterization of a novel Terrabacter ginsenosidimutans sp. nov. beta-glucosidase that transforms ginsenoside Rb1 into the rare gypenosides XVII and LXXV, Appl Environ Microbiol, 76 (2010) 5827-5836.
69. Singh A, Hayashi K, Hoa T, Kashiwagi Y & Tokuyasu K, Zverlov V V, Volkov I Y, Velikodvorskaya T V & Schwarz W H, Thermotoga neapolitana bglB gene, upstream of lamA, encodes a highly thermostable beta-glucosidase that is a laminaribiase, Microbiology, 143 (1997) 3537-3542.
70. Feng Y, Duan C J, Pang H, Mo X C, Wu CF, Yu Y et al, Cloning and identification of novel cellulase genes from uncultured microorganisms in rabbit cecum and characterization of the expressed cellulases, Appl Microbiol Biotechnol, 75 (2007) 319-328.
71. Nunoura N, Ohdan K, Yano T, Yamamoto K & Kumagai H, Purification and characterization of beta-D-glucosidase (beta-D-fucosidase) from Bifidobacterium breve clb acclimated to cellobiose, Biosci Biotechnol Biochem, 60 (1996).

Table 1: Protein sequences considered for the study

S.No.	Source Organisms	CAZy Family classification	Temperature Stability /optimum range (^{0}C)	Accession No.	References
1	*Thermotoga neapolitana*	GH1	95^{0C}	B9K7M5	59
2	*Thermobispora bispora*	GH1	60^{0C}	P38645	60
3	*Thermoanaerobacter brockii*	GH1	60^{0C}	Q60026	61
4	*Clostridium stercorarium*	GH3	65^{0C}	O08331	62
5	*Thermoanaerobacter ethanolicus*	GH1	80^{0C}	D3Y2V4	63
6	*Clostridium thermocellum*	GH3	65^{0C}	P14002	64
7	*Thermotoga neapolitana*	GH1	90^{0C}	Q60038	65

8	*Paenibacillus sp. HC1*	GH1	37^{0C}	Q2WGB4	66
9	*Terrabacter ginsenosidimutans*	GH3	37^{0C}	D2KAT1	67
10	*Cellvibrio gilvus*	GH3	35^{0C}	P96316	68
11	*Uncultured bacterium*	GH3	40^{0C}	Q3HXC2	69
12	*Bifidobacterium breve*	GH1	45^{0C}	P94248	70

Table 2: Various parameters of TB Protein sequences calculated using Expasy's ProtParam tool considered for the study.

Accession number	Sequence length	Molecular weight	Theoretical pI	Total number of positively charged residues (Asp + Glu)	Total number of negatively charged residues (Arg + Lys)	Extinction coefficient	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
B9K7M5	444	51535.2	5.66	61	49	122730	33.45	80.99	-0.423
P38645	473	52227.8	7.36	57	57	95800	29.92	85.05	-0.311
Q60026	450	52024.8	5.28	75	57	120210	36.86	77.42	-0.532
O08331	754	85207.1	5.67	117	103	93740	35.37	86.21	-0.424
D3Y2V4	447	51849.7	5.58	71	58	120210	35.08	75.30	-0.555
P14002	755	83900.4	5.38	114	93	74260	37.14	89.59	-0.310
Q60038	720	81370.0	5.27	114	94	119180	34.11	81.83	-0.492

Table 3: Various parameters of MB Protein sequences calculated using Expasy's ProtParam tool considered for the study.

Accession number	Sequence length	Molecular weight	Theoretical PI	Total number of positively charged residues (Asp + Glu)	Total number of negatively charged residues (Arg + Lys)	Extinction coefficient	Instability index	Aliphatic index	Grand average of hydropathicity (GRAVY)
2WGB4	448	51383.8	4.90	61	40	116770	36.37	79.91	-0.335
D2KAT1	648	70622.5	4.88	99	59	56840	42.37	88.92	-0.198
P96316	752	79942.6	7.75	76	77	102330	33.34	80.55	-0.210
Q3HXC2	854	92714.6	4.44	123	62	95120	34.13	86.60	-0.174
P94248	460	51513.8	5.00	66	42	94770	26.75	77.63	-0.330

Table 4: Calculated secondary structure elements by SOPMA tool for TB.

Accession number	Alpha helix	Extended strand	Beta turn	Random coil
B9K7M5	36.26%	16.44%	8.11%	39.19%
P38645	37.00%	14.38%	9.30%	39.32%
Q60026	38.44%	15.56%	9.78%	36.22%
O08331	36.87%	17.90%	6.50%	38.73%
D3Y2V4	36.69%	14.54%	9.84%	38.93%
P14002	36.29%	16.95%	7.95%	38.81%
Q60038	34.17%	17.64%	5.56%	42.64%

Table 5: Calculated secondary structure elements by SOPMA tool for MB.

Accession number	Alpha helix	Extended strand	Beta turn	Random coil
Q2WGB4	37.50%	15.62%	9.15%	37.72%
D2KAT1	35.34%	11.88%	4.32%	48.46%
P96316	32.18%	17.02%	6.12%	44.68%
Q3HXC2	39.70%	16.04%	5.97%	38.29%
P94248	37.17%	15.00%	9.35%	38.48%

Table 6: Showing statistically analysed values for various amino acids present in TB and MB sequences.

S. No.	Amino acids	Mean value %		t-value	P-value (<0.05)
		TB	MB		
1	**Ala**	7.1	9.6	1.7423	0.1154
2	**Cys**	5.85	5.58	0.3505	0.7342
3	**Asp**	4.2	3.7	0.4725	0.6478
4	**Gln**	1.71	3.12	5.0399	0.0007
5	**Phe**	4.250	4.892	1.5392	0.1380
6	**Gly**	8.042	8.217	0.3257	0.7477
7	**His**	2.383	3.108	2.1840	0.0399
8	**Ile**	6.275	5.633	0.9274	0.3638
9	**Lys**	6.850	4.633	2.2296	0.0363
10	**Leu**	8.392	8.083	0.4627	0.6482
11	**Met**	2.033	2.108	0.3199	0.7521
12	**Asn**	4.917	4.608	0.4032	0.6907
13	**Pro**	4.183	4.733	1.2165	0.2367
14	**Glu**	2.533	3.092	1.2769	0.2149
15	**Arg**	4.917	5.517	0.5997	0.5548
16	**Ser**	5.342	4.667	1.3672	0.1854
17	**Thr**	4.200	5.192	1.7919	0.0869
18	**Val**	6.800	5.950	1.2606	0.2206
19	**Trp**	2.217	2.058	0.4440	0.6614
20	**Tyr**	4.975	5.192	0.3141	0.7564

*significative at a level of 5% of probability (P=<0.05).

Table 7: Showing statistically analyzed values for various Physic-chemical properties for TB and MB.

S. No.	Parameters	Mean value %		t-value	P-value (<0.05)
		TB	MB		
1	**Theoretical pI**	5.7429	5.492	0.3761	0.7156
2	**Random coils**	39.120	42.287	1.5026	0.1672
3	**Total no. of Positively charged residues (Arg + Lys)**	87	89.75	0.1628	0.8743
4	**Total no. of Negatively charged residues (Asp+Glu)**	73.00	59.50	1.0556	0.3187
5	**Instability index**	34.56	36.55	1.0282	0.3307
6	**GRAVY**	-0.4352	-0.2292	3.6275	0.0055
7	**Extended strand**	16.201	15.140	0.9730	0.3560
8	**Beta-turn**	8.1486	6.390	1.5778	0.1491
9	**Alpha-helix**	36.531	36.180	0.2642	0.7976
10	**Aliphatic index**	82.34	83.99	0.545	0.5985

*****Significative at a level of 5% of probability (P=<0.05).

Table 8: Motif analysis of TB sequences.

Accession number	Motif 1	Motif 2	Motif3	Motif4	Motif5	Motif6	Motif 7	Motif8
B9K7M5	+	+	+	+	-	+	+	+
P38645	+	+	+	+	-	+	+	+
Q60026	+	+	+	+	-	+	+	+
O08331	-	-	-	-	+	+	-	-
D3Y2V4	+	+	+	+	-	+	+	+
P14002	-	-	-	-	+	+	-	-

Q60038	-	-	-	-	+	+	-	-

Table 9: Showing amino acid sequences of motifs found in TB.

Motif description	**Amino acid sequence of the motifs**
Motif 1	W[DH][TV]FS[HK]T[EP]G[KN][TV][YAK][NKS]G[HD]TGD[VI]ACDHY[HN]R[YW][KA][ED]D[VI][ER][IL][LIM][KAE][EGK][IL]G[VAD][KR][AV]YRFS[IV][AS]WPR
Motif 2	[RK]VHD[DPQ][EN]RI[EDR]Y[IL][KRT][AE][HT][LI][KER][AQ][AV][AHW][KDR][AF]I[GMQ][EAD][GR][GAV][NDP]L[KR]GYFVWS[LV][LM]DNFEWA[HEY]GY[SH]KR[FG]
Motif 3	G[KP][VY]N[PQS][KA]G[LM]DFY[KDN][RK][LI][IV]D[ET]L[LQ][KEG][KHR][DGN]I[TMV]P[AFTY][AIP]T[IL]YHWDLP
Motif 4	GGW[AL][NA]R[DE][SIT][AIV][KDY][WR][FY][AV]EY[AS][TLR][KAV][LV][FH][ER][ENR][LF]GD[RAV][IV][PKR][LCH]WIT[HL]NEPW[CV][ASV][AS][IF][LV][SAG][YHT][GHL][IRY]G[EAV][HP][AG]
Motif 5	PL[CS]GRNFEY[FY]SEDP[YV]LS[SG]E[ML]A[ARS]S[HF][IV]KGVQS[QR]GVG[AT][CS][LI]KHF[AV]ANNQE[HT][RN]RM
Motif 6	WG[VT][AM][TC][AS][SY][YN][QK][IL][EN]G[AE][YV][CN][ES][DE][GN]R[GT][LP]
Motif 7	IVYVDY[TEN]T[QM][KR]RI[LIP][KR][DE]S[AG][LY]WY[KRS][EDN]V[IV][KLQR][DKNR][ND][GS]
Motif 8	P[ED][GS][LI][YR][DTW][LI]L[KL][RG][LV][DKS][RE][ED]Y[TNP][KGP][LQV][PEG][MLV][YI]ITENGAAF

Fig 1. phlogenetic tree between thermostable and mesostable β- glucosidases