ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 7 (3): 418-429, 2008
ISSN 1682-296X
© 2008 Asian Network for Scientific Information

Molecular Modeling of a Predominant β-CGTase G1 and Analysis of Ionic Interaction in CGTase

[1,2]Kian Mau Goh, [3]Nor Muhammad Mahadi, [4]Osman Hassan,
[5]Raja Noor Zaliha Raja Abdul Rahman and [1]Rosli Md. Illias
[1]Department of Bioprocess Engineering, Faculty of Chemical and Natural Resources Engineering,
Universiti Teknologi Malaysia, 81310 Johor, Malaysia
[2]Faculty of Biosciences and Bioengineering, Universiti Teknologi Malaysia, 81310 Johor, Malaysia
[3]Malaysia Genome Institute, UKM-MTDC Smart Technology Center, 43600 Bangi Selangor, Malaysia
[4]School of Chemical Science and Food Technology, Faculty of Science and Technology,
Universiti Kebangsaan Malaysia, 43600 Selangor, Malaysia
[5]Enzyme and Microbial Technology Research Group, Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia, 43400 Selangor, Malaysia

Abstract: The protein 3D structure for CGTase G1 was determined by homology modeling and a good structure was generated after three rounds of energy minimization process. The ERRAT and Verify3D scores for the predicted structure were determined as 98.07 and 90.991%, respectively. The presences of ionic interactions inside the CGTase G1 structure were compared with five CGTases crystal-structures. Mesophilic CGTases have lesser numbers of ionic pairs (average of 72.3 pairs) than of the thermophilic CGTases (average of 78.6 pairs). Most of the interactions in CGTases were involved in the form of networking. The average number for networking ionic pairs in thermostable and mesophilic CGTases is 69.3 and 62.7, respectively. Most of the ionic interactions in CGTases were found in Domain A and the most complex ionic networking was located in this catalytic domain as well. These charged-residues generate interlinking networking that covers a huge area that surrounds the active site groove. A few numbers of secondary structures strands were interlinked by the ionic interactions and this creates a natural protection for the catalytic TIM-barrel structure (Domain A) against heat. Most of the residues involved are consensus, however, slight variations were found. The triad Asp181-Arg185-Asp175 might plays an important factor in the networking which causes the half life of CGTase G1 to be slightly higher compared to other CGTases originally produced by mesophilic strains.

Key words: Cyclodextrin glucanotransferase, homology modeling, ionic interaction, protein salt bridge, protein stability

INTRODUCTION

Cyclodextrin glycosyltransferase (CGTase) is an industrial enzyme which produces cyclodextrin (CD) when it reacts with starch. Cyclodextrins have various applications in the food, cosmetic, pharmaceutical and agrochemical industries. Bacteria *Bacillus* sp. G1 has been screened previously and was found to produce CGTase (Sian *et al*., 2005). CGTase from *Bacillus* sp. G1 produced about 89% β-CD and 11% γ-CD, with no α-CD detected. The gene encoded the enzyme was earlier fished out and the nucleotide sequence of the CGTase gene was submitted to the National Center for Biotechnology Information (NCBI) database (accession number AY770576).

Domain A contains the $(\beta/\alpha)_8$ tertiary arrangement or so called TIM barrel structure. The fold of the caronical $(\beta/\alpha)_8$ -barrel consists of a closed eight-stranded parallel β-strands, forming the central barrel, which is surrounded by eight α-helices. The active site of CGTase is located in Domain A. CGTase is mainly used in starch industry. High operation temperature is normally applied in liquefying starch paste and CGTase that can resist to heat is of great advantages.

There are many reasons that contribute to the thermostability of proteins and these factors have been extensively reviewed by various researchers (Szilagyi and Zavodszky, 2000; Kumar *et al*., 2000; Kumar and Nussinov, 2001; Li *et al*., 2005; Greaves and Warwicker, 2007). Generally, proteins that can withstand higher

Corresponding Author: Rosli Md. Illias, Department of Bioprocess Engineering,
Faculty of Chemical and Natural Resources Engineering, Universiti Teknologi Malaysia,
81310 Johor, Malaysia Tel: +607-5535564 Fax:+607-5581463

temperature have greater hydrophobicity (Haney *et al.*, 1999), better efficiency in the packing of folded protein (Russell *et al.*, 1997), deletion or shortening of loops (Chakravarty and Varadarajan, 2002; Zhou, 2004), smaller and less structural cavities (Russell *et al.*, 1997), higher percentage of proline residues and reduced amount of thermolabile residues (Kumar *et al.*, 2000), increase in the presence of interactions such as the hydrogen, ionic, disulfide bond, hydrophobic, electrostatic and hydrophobic interactions (Vogt *et al.*, 1997; Li *et al.*, 2005; Ramazzotti *et al.*, 2006; Greaves and Warwicker, 2007). Due to the contribution of these factors, protein or enzyme can exhibits enhanced thermodynamic and kinetic stabilities (Kumar *et al.*, 2001).

The purpose of this study is to homology model the structure of CGTase G1 and to study the relationship between ionic interactions and thermostability. The ionic interactions in the mesophilic and thermophilic CGTase homologous were compared.

MATERIALS AND METHODS

Cell culturing, purification and determination of activity half life: The alkalophilic bacteria identified as *Bacillus* sp. G1 was isolated from soil and the nucleotide sequence of the CGTase gene was submitted to the National Center for Biotechnology Information (NCBI) database (accession No. AY770576). The recombinant CGTase G1 was expressed by *E. coli* JM109 in 1L LB/amp media and the crude enzyme was harvested after overnight incubation at 37°C, 200 rpm. The method used to purify the recombinant CGTase G1 is similar to the procedure used to purify the wild type CGTase (Sian *et al.*, 2005). The β-CD cyclization activity of CGTase was measured according to the method established by Kaneko *et al.* (1987) with modification. The reaction mixture containing 40 mg of soluble starch in 1.0 mL of 0.1 M phosphate buffer (pH 6.0) and 0.1 mL of enzyme solution was incubated at 60°C for 10 min. The reaction was stopped by adding 3.5 mL of 30 mM NaOH solution. Then, 0.5 mL of 0.02% (w/v) phenolphthalein in 5 mM Na_2CO_3 was added to the mixture and mixed well. After leaving the mixture to stand for 15 min at room temperature, the reduction in colour intensity was measured at 550 nm. One unit of enzyme activity is defined as the amount of enzyme that formed 1 μmol of β-CD per minute under the conditions mentioned above. The half life activity of CGTase G1 was determined by soaking the samples at optimum temperature (60°C) without substrate for various periods of time and the residual activity was determined using the assay mentioned above.

Homology modeling and determination of ionic interaction in the 3D structure: All computational methodology was carried out using Accelrys Discovery Studio (DS) Modeling 1.1 *Windows*-based modules, unless specified. Blasting to NCBI database was done using the BLAST program and automated sequence alignment between CGTase G1 and the template sequence was carried out. Adjustment on the gap was done manually to improve the alignment. Homology modeling was performed using Modeler software and the internal algorithm for the optimization level option for model building was set to high where the program used a through molecular dynamic simulation annealing step when building the first model. The 3D structure of CGTase G1 thus obtained was further optimized by gradual energy minimization in three stages. Constraints in both stages were positional harmonic constraints of 20 kcal mol^{-1}. A, applied to all of the backbone atoms. The refined model was used as the final structure for further analysis and comparison. Display of structures were achieved using the DS Modeling Visualizer.

The presence of ionic interactions in the deduced CGTase G1 3D structure was compared with the five known 3D structure of CGTases. The residue pairs that are involved in the ionic interactions were preliminary predicted using the WHAT- IF web interface program (Rodriguez *et al.*, 1998). The residues numbers involved in the interactions were extracted and the coordinates of the charged atoms were subsequently identified using the Yasara View software (Krieger and Vriend, 2002). Then, the centroid distance between the opposite charged residues were computed manually (using Microsoft Excel) by taking the average of the coordinates of the heavy (non hydrogen) atoms in the side chain charged group. The calculation was based on the reported method by Kumar and Nussinov (2002a).

RESULTS AND DISCUSSION

Homology prediction and structural refinement: From the ExPASy database, it was found that more than forty CGTase gene/amino acid sequence from various strains have been deposited. The number of known sequence is expected to grow. Three dimensional (3D) structures of five CGTase are known from the *Bacillus circulans* strain 8 and 251 (BC251), *Thermoanaerobacterium thermosulfurigenes* strain EM1 (*Tabium*), *Bacillus stearothermophiles* and alkalophilic *Bacillus* sp. 1011. From these five main structures, approximately 40 crystal structures were further developed where the crystals were diffracted at different conditions such as in the presence of ligands typically calcium ions, genuine and pseudo-

Table 1: Energy levels and validation scores for CGTase G1 models at different stages

Explanation	Structure energy level (kcal mol^{-1})	Structure validation by ERRAT plot	Structure validation by VERIFY3D (%)
Initial CGTase G1 model build	9,291.83	97.63	79.880
CGTase G1 model after first minimization	-5,108.74	96.59	88.138
CGTase G1 model after second minimization	-5,519.16	97.93	88.438
CGTase G1 model after third minimization	-5,961.00	98.07	90.991

substrates, cyclodextrins as well as inhibitors (data retrieved from NCBI protein structure database). Homology modeling for protein CGTase G1 was executed to understand the structure of this enzyme and to determine the ionic interactions that exist within the structure. Comparative protein modeling is based on the reasonable assumption that two homologous proteins will share very similar structures (Rodriguez *et al*., 1998). Because a protein's fold is more evolutionarily conserved than its amino acid sequence, CGTase G1 can be modeled with reasonable accuracy on a very distantly related template, provided that the relationship between target and template can be discerned through sequence alignment.

Most of the computational methodologies used in the experiments were carried out using Accelrys Discovery Studio (DS) Modeling 1.1 *Windows* based modules, unless specified. Blasting to NCBI database was done using the BLAST program. CGTase from *Bacillus stearothermophiles* (1CYG) was chosen as template because it turned out to have the highest E value 6.3 e^{-227} and has the closest sequence identity of 62% and similarity of 78% to CGTase G1. The amino acid alignment between CGTase G1, template sequence (1CYG) and four other sequences with known 3D structures is shown in Fig. 1. Automated sequence alignment between CGTase G1 and template sequence was carried out using Align123 program. Subsequently, initial model for CGTase G1 was built by homology modeling performed using Modeler software. The 3D structure of CGTase G1 thus obtained was further optimized by gradually minimizing energy in three stages.

Table 1 resumes the energy level and structure quality for the models in different steps. Preliminary model generated by the Modeler software has high energy level (9,291.83 kcal mol^{-1}). Normally this energy level is considered too high and due to this reason, a series of energy minimization steps is required to reach a local or more preferential global minimum level of total energy. *In silico*, protein structures are more stable in low energy level. After three rounds of minimization process, the total energy of CGTase G1 model reduced to -5,108.74, -5,519.16 and -5,961.00 kcal mol^{-1}, respectively for models obtained in the first, second and third round minimization, as shown in Table 1.

Verification of the models in each stage of minimization processes were determined using the ERRAT and Verify3D programs (Colovos and Yeates, 1993). ERRAT program was designed to detect local errors within the geometry in the 3D structure using a nine residue sliding window approach. The quality of CGTase G1 initial model was considered good based on ERRAT plot program (97.63). The ERRAT score for CGTase G1 structure was further improved to 98.07 after three rounds of minimization processes (Table 1) resulting in a high quality of CGTase G1 model. Some of the predicted CGTase structures deposited in Swiss-Model Repository were examined, CGTase *Bacillus licheniformis* (UniProt AC P14014) has ERRAT score of 82.5, CGTase *Brevibacillus brevis* (O30565) 81.4, CGTase *Bacillus ohbensis* (P27036) 79.3, CGTase *Bacillus* sp. strain 17-1 (P30921) 93.4 and CGTase *Bacillus* sp. strain 38-2 (P09121) 85.

The structures of CGTase G1 generated at different optimization steps were also evaluated using the VERIFY 3D program in DS Modeling package. While the ERRAT plot justified the geometrical errors, VERIFY 3D program validates the structure by a statistical approach that measures the compatibility of an amino acid sequence with the model structure. The comparisons were based on criteria such as residue environments (area of buried and fraction polar of each residue) and the local secondary structure. As shown in Table 1, the VERIFY3D scoring for the initial CGTase G1 model was 79.88%. The score improved gradually to 88.138 and 88.438, respectively for models in first and second round of minimizations. The score for model refined in the third minimizations process was 90.991% which simply means that 90.991% of the total amino acids in the 3D structure were located in a theoretically correct position. Only 9.009% of the structure was inaccurately predicted. This might be due to a few reasons. First, there were no compatible amino acids in the template that were useful to generate estimated positions of coordination in space. Second, the flexibility of these amino acids was higher. Based on the scoring from ERRAT plot and VERIFY3D program, it was concluded that a reasonably good 3D structure was predicted for CGTase G1. The Ramachandran plot (Lovell *et al*., 2003) for the model generated after three cycles of energy minimization is shown in Fig. 2. There are

```
G1       ----DVTNKVNYSKDVIYQVVTDRFSDGNPGNNPSGAIFSQNCIDLHKYCGGDWQGIIDK 56
1CYG     ---AGNLNKVNFTSDVVYQIVVDRFVDGNTSNNPSGALFSSGCTNLRKYCGGDWQGIINK 57
1CIU     ASDTAVSNVVNYSTDVIYQIVTDRFVDGNTSNNPTGDLYDPTHTSLKKYFGGDWQGIINK 60
1CXI     APDTSVSNKQNFSTDVIYQIFTDRFSDGNPANNPTGAAFDGTCTNLRLYCGGDWQGIINK 60
1CGT     DPDTAVTNKQSFSTDVIYQVFTDRFLDGNPSNNPTGAAYDATCSNLKLYCGGDWQGLINK 60
1PAM     APDTSVSNKQNFSTDVIYQIFTDRFSDGNPANNPTGAAFDGSCTNLRLYCGGDWQGIINK 60
                  *   .::.**:**:..*** ***..***:*   :.     .*:  * ******:*:*

G1       INDGYLTDLGITALWISQPVENVYAL----HPSGY--TSYHGYWARDYKKTNPYYGNFDD 110
1CYG     INDGYLTDMGVTAIWISQPVENVFSVMNDA--SG--SASYHGYWARDFKKPNPFFGTLSD 113
1CIU     INDGYLTGMGVTAIWISQPVENIYAVLPDSTFGG--STSYHGYWARDFKRTNPYFGSFTD 118
1CXI     INDGYLTGMGVTAIWISQPVENIYSII---NYSGVNNTAYHGYWARDFKKTNPAYGTIAD 117
1CGT     INDNYFSDLGVTALWISQPVENIFATI---NYSGVTNTAYHGYWARDFKKTNPYFGTMAD 117
1PAM     INDGYLTGMGITAIWISQPVENIYSVI---NYSGVNNTAYHGYWARDFKKTNPAYGTMQD 117
         ***.*::.:*:**:********:::           .*   ::********:*:.** :*.: *

G1       FDRLMSTAHSNGIKVIMDFTPNHSSPALETNPNYVENGAIYDNGTLLGNYSNDQQNLFHH 170
1CYG     FQRLVDAAHAKGIKVIIDFAPNHTSPASETNPSYMENGRLYDNGTLLGGYTNDANMYFHH 173
1CIU     FQNLINTAHAHNIKVIIDFAPNHTSPASETDPTYAENGRLYDNGTLLGGYTNDTNGYFHH 178
1CXI     FQNLIAAAHAKNIKVIIDFAPNHTSPASSDQPSFAENGRLYDNGTLLGGYTNDTQNLFHH 177
1CGT     FQNLITTAHAKGIKIVIDFAPNHTSPAMETDTSFAENGRLYDNGTLVGGYTNDTNGYFHH 177
1PAM     FKNLIDTAHAHNIKVIIDFAPNHTSPASSDDPSFAENGRLYDNGNLLGGYTNDTQNLFHH 177
         *..*: :**::.**:::**:***:*** . :..: *** :****.*:*.*:** :   ***

G1       NGGTDFSSYEDSIYRNLYDLADYDLNNTVMDQYLKESIKFWLDKGIDGIRVDAVKHMSEG 230
1CYG     NGGTTFSSLEDGIYRNLFDLADLNHQNPVIDRYLKDAVKMWIDMGIDGIRMDAVKHMPFG 233
1CIU     YGGTDFSSYEDGIYRNLFDLADLNQQNSTIDSYLKSAIKVWLDMGIDGIRLDAVKHMPFG 238
1CXI     NGGTDFSTTENGIYKNLYDLADLNHNNSTVDVYLKDAIKMWLDLGIDGIRMDAVKHMPFG 237
1CGT     NGGSDFSSLENGIYKNLYDLADFNHNNATIDKYFKDAIKLWLDMGVDGIRVDAVKHMPLG 237
1PAM     YGGTDFSTIENGIYKNLYDLADLNHNNSSVDVYLKDAIKMWLDLGVDGIRVDAVKHMPFG 237
          **: **: *:.**:**:**** : :*. :* *:*.::*.*:* *:****:******. *

G1       WQTSLMSEIYSHKPVFTFGEWFLGSGEVDPQNHHFANESGMSLLDFQFGQTIRNVLKDRT 290
1CYG     WQKSLMDEIDNYRPVFTFGEWFLSENEVDANNHYFANESGMSLLDFRFGQKLRQVLRNNS 293
1CIU     WQKNFMDSILSYRPVFTFGEWFLGTNEIDVNNTYFANESGMSLLDFRFSQKVRQVFRDNT 298
1CXI     WQKSFMAAVNNYKPVFTFGEWFLGVNEVSPENHKFANESGMSLLDFRFAQKVRQVFRDNT 297
1CGT     WQKSWMSSIYAHKPVFTFGEWFLGSAASDADNTDFANKSGMSLLDFRFNSAVRNVFRDNT 297
1PAM     WQKSFMATINNYKPVFTFGEWFLGVNEISPEYHQFANESGMSLLDFRFAQKARQVFRDNT 297
         **.. *  :  ::**********.    . :   ***:********:* .  *:*::.:

G1       SNWYDFNEMITSTEKEYNEVIDQVTFIDNHDMSRFSVGSSSNRQTDMALAVLLTSRGVPT 350
1CYG     DNWYGFNQMIQDTASAYDEVLDQVTFIDNHDMDRFMIDGGDPRKVDMALAVLLTSRGVPN 353
1CIU     DTMYGLDSMIQSTASDYNFINDMVTFIDNHDMDRFY-NGGSTRPVEQALAFTLTSRGVPA 357
1CXI     DNMYGLKAMLEGSAADYAQVDDQVTFIDNHDMERFHASNANRRKLEQALAFTLTSRGVPA 357
1CGT     SNMYALDSMINSTATDYNQVNDQVTFIDNHDMDRFKTSAVNNRRLEQALAFTLTSRGVPA 357
1PAM     DNMYGLKAMLEGSEVDYAQVNDQVTFIDNHDMERFHTSNGDRRKLEQALAFTLTSRGVPA 357
         .. * :. *: .:   *  : * *********.**  .  . *  : ***. *******

G1       IYYGTEQYVTGGNDPENRKPLKTFDRSTNSYQIISKLASLRQTNSALGYGTTTERWLNED 410
1CYG     IYYGTEQYMTGNGDPNNRKMMSSFNKNTRAYQVIQKLSSLRRNNPALAYGDTEQRWINGD 413
1CIU     IYYGTEQYMTGNGDPYNRAMMTSFNTSTTAYNVIKKLAPLRKSNPAIAYGTTQQRWINND 417
1CXI     IYYGTEQYMSGGTDPDNRARIPSFSTSTTAYQVIQKLAPLRKCNPAIAYGSTQERWINND 417
1CGT     IYYGTEQYLTGNGDPDNRAKMPSFSKSTTAFNVISKLAPLRKSNPAIAYGSTQQRWINND 417
1PAM     IYYGSEQYMSGGNDPDNRARLPSFSTTTTAYQVIQKLAPLRKSNPAIAYGSTHERWINND 417
         ****:***::*. ** **  : :*. .* ::::*.**:.**: *.*:.** * :**:* *

G1       IYIYERTFGNSIVLTAVN-SSNSNQTITNLNTSLPQGNYTDELQQRLDGNTITVNANGAV 469
1CYG     VYVYERQFGKDVVLVAVNRSSSSNYSITGLFTALPAGTYTDQLGGLLDGNTIQVGSNGSV 473
1CIU     VYIYERKFGNNVALVAINRNLSTSYNITGLYTALPAGTYTDVLGGLLNGNSISVASDGSV 477
1CXI     VLIYERKFGSNVAVVAVNRNLNAPASISGLVTSLPQGSYNDVLGGLLNGNTLSVGSGGAA 477
1CGT     VYVYERKFGKSVAVVAVNRNLSTSASITGLSTSLPTGSYTDVLGGVLNGNNIT-STNGSI 476
1PAM     VIIYERKFGNNVAVVAINRNMNTPASITGLVTSLPRGSYNDVLGGILNGNTLTVGAGGAA 477
         : :*** **..:.:.*:* . .:  .*:.* *:** *.*.* *   *:**.:   :.*:

G1       NSFQLRANSVAVWQVSNPSTSPLIGQVGPMMGKSGNTITVSGEGFGDERGSVLFDSTS-- 527
1CYG     NAFDLGPGEVGVWAYSATESTPIIGHVGPMMGQVGHQVTIDGEGFGTNTGTVKFGTTA-- 531
1CIU     TPFTLSAGEVAVWQYVSSSNSPLIGHVGPTMTKAGQTITIDGRGFGTTSGQVLFGSTAGT 537
1CXI     SNFTLAAGGTAVWQYTAATATPTIGHVGPMMAKPGVTITIDGRGFGSSKGTVYFGTTAVS 537
1CGT     NNFTLAAGATAVWQYTTAETTPTIGHVGPVMGKPGNVVTIDGRGFGSTKGTVYFGTTAVT 536
1PAM     SNFTLAPGGTAVWQYTTDATTPIIGNVGPMMAKPGVTITIDGRGFGSGKGTVYFGTTAVT 537
         . *  *  ..  ..**       :* **:*** * :  *  :*:.*.***   * * *.:*:

G1       -SEIISWSNTEISVKVPNVAGGYYDLSVVTAANLKSPTYKEFEVLSGNQVSVRFGVNNAT 586
1CYG     -ANVVSWSNNQIVVAVPNVSPGKYNITVQSSSGQTSAAYDNFEVLTNDQVSVRFVVNNAT 590
1CIU     ---IVSWDDTEVKVKVPSVTPGKYNISLKTSSGATSNTYNNINILTGNQICVRFVVNNAS 594
1CXI     GADITSWEDTQIKVKIPAVAGGNYNIKVANAAGTASNVYDNFEVLSGDQVSVRFVVNNAT 597
1CGT     GAAITSWEDTQIKVTIPSVAAGNYAVKVA-ASGVNSNAYNNFTILTGDQVTVRFVVNNAS 595
1PAM     GADIVAWEDTQIQVKIPAVPGGIYDIRVANAAGAASNIYDNFEVLTGDQVTVRFVINNAT 597
             : :*.:.:: * :* *. * * : :  ::.  *  *.:: :*:.:*: *** :***:

G1       TSPGTNLYIVGNVSELGNWDA-DKAIGPMFNQVMYQYPTWYYDISVPAGKNLEYKYIKKD 645
1CYG     TNLGQNIYIVGNVYELGNWDT-SKAIGPMFNQVVYSYPTWYIDVSVPEGKTIEFKFIKKD 649
1CIU     TVYGENVYLTGNVAELGNWDT-SKAIGPMFNQVVYQYPTWYYDVSVPAGTTIQFKFIKKN 653
1CXI     TALGQNVYLTGSVSELGNWDP-AKAIGPMYNQVVYQYPNWYYDVSVPAGKTIEFKFLKKQ 656
1CGT     TTLGQNLYLTGNVAELGNWSTGSTAIGPAFNQVIHQYPTWYYDVSVPAGKQLEFKFFKKN 655
1PAM     TALGQNVFLTGNVSELGNWDP-NNAIGPMYNQVVYQYPTWYYDVSVPAGQTIEFKFLKKQ 656
         *  * *:::.*.* *****..  .**** :***::.**.** *:*** *  :::*::**:

G1       QNGNVVWQSGNNRTYTSPTTGTDTVMINW-------------------------- 674
1CYG     SQGNVTWESGSNHVYTTPTNTTGKIIVDWQN------------------------ 680
1CIU     GN-TITWEGGSNHTYTVPSSSTGTVIVNWQQ------------------------ 683
1CXI     GS-TVTWEGGSNHTFTAPSSGTATINVNWQP------------------------ 686
1CGT     GS-TITWESGSNHTFTTPASGTATVTVNWQ------------------------- 684
1PAM     GS-TVTWEGGANRTFTTPTSGTATVNVNWQP------------------------ 686
          . .:.*:.* *:.:* *:. * .: ::*
```

Fig. 1: Amino acids alignment of CGTase G1 with some other CGTases with known 3D structure. Abbreviation used: CGTase from *Bacillus stearothermophiles* (1CYG, template used in modeling CGTase G1), CGTase from *Bacillus irculans* strain 8 (1CGT), CGTase from alkalophilic *Bacillus* sp. 1011 (1PAM), CGTase from *Thermoanaerobacterium thermosulfurigenes* strain EM1 (1CIU) and CGTase from *Bacillus* circulans 251 (1CXI). α-helices were indicated by red-bars while β-sheets were shown as blue-arrows

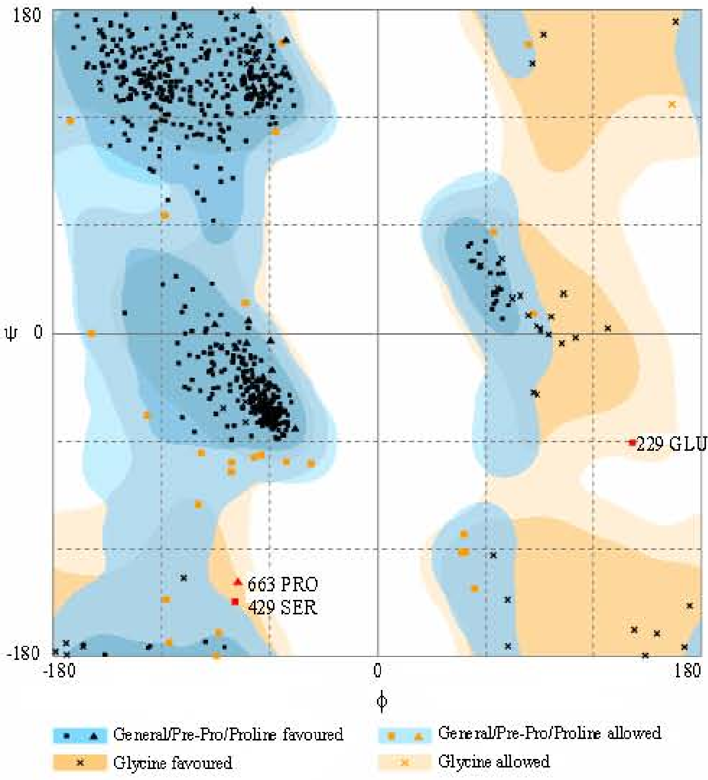

Fig. 2: Full Ramachandran plot of the main-chain conformation angles of final CGTase G1 model. Glu 229, Ser 429 and Pro 663 (red colour) are those energetically unfavorable residues

only three amino acid- Glu229, Ser429 and Pro663 that fall into energetically unfavorable region in the Ramachandran plot. Number of residues in CGTase G1 3D structure that were in the favoured region was 95.5%, allowed region (4%) and outlier region (0.4%). Based on the ERRAT plot, VERIFY3D scoring and Ramachandran plot, it is shown that a good model for CGTase G1 was generated.

General amino acid comparison: *Bacillus* sp. G1 was originally isolated locally from soil at a rubber estate. Although CGTase G1 was produced from a mesophilic *Bacillus* strain, it has an activity half life of 30 min at 60°C which is considered above the average half life of known CGTases. It can be observed from Table 2 that CGTase G1 is more stable than some other CGTases from the mesophilic microorganism. The CGTase G1 can withstand higher temperature in a longer period of time.

In folded proteins, pairs of neighbouring oppositely charged residues often interact to form ionic interactions. These interactions stabilize the folded conformations of proteins. Kumar and Nussinov coworkers collected abundant protein structure in the PBD database and analyzed the statistical correlation of ionic interactions with the stability of protein structure based on the half life of each proteins/enzymes (Kumar *et al*., 2000; Kumar and Nussinov, 2002b). Such statistical works on ionic interactions were also performed by other researches, for examples in reports by Musafia *et al*. (1995), Szilagyi and Zavodszky (2000) and Hendsch (1994). It was concluded from the statistical data that ionic interactions have important influences on the stability of a protein. Thermophilic protein or enzyme normally has more ionic interaction than of the same type of protein produced from the mesophilic strains (Kumar *et al*., 2000).

In Table 3, ionic interactions that existed in CGTase G1 structure were compared with five known crystal structures for CGTase *Thermosulfurigenes thermosulfurigenes EM1* (PDB code: 1CIU), CGTase *Bacillus stearothermophiles* (1CYG), CGTase *Bacillus circulans* 251 (1CXI), CGTase *Bacillus circulans* No. 8 (1CGT) and CGTase from *Bacillus* sp. 1011 (1PAM). Among them, two CGTases were thermostable enzymes while the rest were mesophilic enzymes.

Generally, it can be observed from Table 3 (Row No. 1) that thermostable CGTases exhibited slightly shorter amino acid sequence than of the mesophilic CGTases. Interestingly, CGTase G1 has the shortest

Table 2: Thermostability behavior of selected CGTases

CGTase source	Optimum temperature (°C)	Temperature stability	Reference
Thermophilic CGTase			
Thermococcus sp. B1001	>90	90°C, >3 h	Yamamoto *et al.* (2000)
Pyrococcus furiosus DSM3638	95	95°C, 46 min	Lee *et al.* (2007)
Thermococcus kodakaraensis KOD1	80	No activity loss after 80 min, at 85°C $t_{½}$: at 100°C 20 min, (present of 10 mM $CaCl_2$	Rashid *et al.* (2002)
Tabium (EM1)	60	$t_{½}$: 90°C, 15 min	Leemhuis *et al.* (2004)
Bacillus stearothermophilus	80	$t_{½}$: 75°C, 10 min	Leemhuis *et al.* (2004)
Mesophilic CGTase			
Bacillus sp. *TS1-1*	60	$t_{½}$: 70°C, 30 min	Rahman *et al.* (2006)
Bacillus firmus var. alkalophilus	50	t½: ~65°C, 10 min	Hyun-Dong *et al.* (2000)
Bacillus stearothermophilus NO_2	60	Retain 65 % activity at 70°C for 1 h	Jeon *et al.* (1998)
Brevibacillus brevis CD162	55	$t_{½}$: 55°C, 30 min	Kim *et al.* (1998)
Bacillus sp. G1	60	$t_{½}$: 60°C, 30 min	Present study
Anaerobranca gottschalkii	65	DSC: 66°C (Tm)	Thiemann *et al.* (2004)
Alkalophilic *Bacillus clarkii* 7364	60	$t_{½}$: 55°C, 15 min	Takada *et al.* (2003)
Bacillus agaradhaerens LS-3C	50	$t_{½}$: 45°C, 1 h	Martins and Hatti-Kaul (2002)
Bacillus circulan 251	50	$t_{½}$: 60°C, 10 min	Leemhuis *et al.* (2004)
Bacillus macerans	60	$t_{½}$: 57°C, 10 min, in 1.2 mM $CaCl_2$	Chang *et al.* (1998)
Paenibacillus sp. F8	50	$t_{½}$: 65°C, 1 h (present of $CaCl_2$)	Larsen *et al.* (1998)
Bacillus circulan E 192	60	Heat stability up to 45°C (present of $CaCl_2$)	Larsen *et al.* (1998)
Klebsiella oxytoca M5A1	N.A	Heat stability up to 45°C (present of $CaCl_2$)	Larsen *et al.* (1998)
Bacillus autolyticus	60	Up to 40°C	Martins and Hatti-Kaul (2002)

Table 3: Summary of ionic interactions in various CGTases. Negatively charged amino acids are Asp (D) and Glu (E) while positively charged residues are Arg (R), His (H) and Lysine (K)

Row No.	Description	G1	1CIU*	1CYG*	1CXI	1CGT	1PAM
1.	Total residue number in primary sequence	674	683	680	686	686	686
2.	Total numbers of Asp (D)+ Glu (E) residues in primary sequence	74	58	70	58	63	63
3.	Total numbers of D+E involved in ionic interactions	50	48	52	48	41	45
4.	Total numbers of Arg (R) +His (H) + Lysine (K) residues in primary sequence	58	54	60	61	60	60
5.	Total numbers of R+H+K residues involved in ionic interactions	47	45	47	51	48	47
6.	Total ionic pairs	78	73	85	76	70	71
7.	Isolated ionic pairs	12	9	7	11	8	10
8.	Networking ionic pairs	66	64	78	65	62	61
9.	Ionic interactions in domain A only	52	45	52	47	44	45
10.	Ionic interactions in domain B only (number in bracket refers to cross interaction between Domain B with other domains)	7 (+3)	7 (+5)	7 (+6)	5 (+4)	7 (+4)	6 (+4)
11.	Ionic interactions in Domain C only	2	3	3	3	4	4
12.	Ionic interactions in Domain D only	2	3	1	4	3	2
13.	Ionic interactions in Domain E only	3	5	4	4	3	3
14.	Cross domains interactions	12	10	18	13	9	11
15.	No. of salt bridges	16	18	15	19	21	19
16.	No. of N-O bridge	12	14	18	14	12	21
17.	No. of long pair	50	41	52	43	37	31

Abbreviation/PDB-code used: G1-CGTase G1, 1CIU-CGTase *T. thermosulfurigenes EM1*, 1CYG- CGTase *Bacillus stearothermophiles*, 1CXI-CGTase *B. circulans* 251, 1CGT-CGTase *B. circulans* No. 8 and 1PAM-CGTase *B.* sp. 1011. *CGTase from *T. thermosulfurigenes EM1* and *Bacillus stearothermophiles* are thermophilic enzymes while the rest are from mesophilic strains

primary sequence of 674 residues. In row number 2 of Table 3, it shows that the overall negatively charged amino acids (Asp and Glu) in the primary sequence were relatively more for thermophilic CGTase (average number 64) compared to mesophilic CGTase (average number 61 without considering CGTase G1). Surprisingly, CGTase G1 has 74 negatively charged residues and this number is greater than the two other thermostable CGTases. The presence of higher quantity of charged residues can increase the changes of ionic interaction forming. Moving downward to row No. 3 of Table 3, the values actually refer to the numbers of Asp+Glu that were found involved in the ionic interactions for each respectively protein structures. By taking CGTase G1 as an example, 50 out of a total 74 negatively charged amino acid (Asp and Glu) were involved in the ionic interactions. The rest of the residues have no charge-charge interactions with any adjacent amino acid in the structure of CGTase G1. The total number of Asp+Glu that were involved in the ionic interaction for CGTase G1, CGTase *Thermosulfurigenes thermosulfurigenes EM1* and CGTase *Bacillus stearothermophiles* once again is slightly higher than that of the mesophilic CGTases, as shown in Table 3 (row No. 3). However, the total numbers of positively charged amino acid (Arg+His+Lys) in protein primary sequence and those that were involved in the ionic interactions

for CGTase G1, CGTase *Thermosulfurigenes thermosulfurigenes EM1* and CGTase *Bacillus stearothermophiles* are lower than that of the mesophilic CGTases, typically CGTase *Bacillus circulans* 251, CGTase *Bacillus circulans* No. 8 and CGTase from *Bacillus* sp. 1011 as shown in Table 3 (row 4 and 5).

Ionic interaction comparison: The statistical analysis showed that ionic interactions have direct correlation with the stability of protein structure (Kumar *et al*., 2000; Kumar and Nussinov, 2002b). The total numbers of ionic pairs in the CGTases protein structure are summarized in Table 3 (row No. 6). CGTase G1 and CGTase *Bacillus stearothermophiles* have total interactions of 78 and 85, respectively. Mesophilic CGTase *Bacillus circulans* 251, CGTase *Bacillus circulans* No. 8 and CGTase from *Bacillus* sp. 1011 however have lesser numbers of ionic pairs, respectively 76, 70 and 71. If more than three charged residues were linked together, it is defined as networking ionic complex. Otherwise, isolated ionic pair is defined as only a negatively charged residue linked with another positively charged residue. Most of the interactions in CGTases were involved in networking systems as shown in Table 3 (row No. 6-8).

Networking pairs for CGTase G1 and thermostable CGTases (average No. 69.3) were found higher than mesophilic CGTase *Bacillus circulans* 251, CGTase *Bacillus circulans* No. 8 and CGTase from *Bacillus* sp. 1011 (average No. 62.7). From the data comparison discussed since the beginning of this section, it appears that CGTase G1 is distinct from the other three mesophilic CGTases. The ionic interactions characteristic for CGTase G1 has higher similarity to the thermostable CGTases rather than the mesophilic CGTases. From statistical analysis, Kumar and Nussinov (2002b) mentioned that normally thermostable proteins have slightly more pairs of interactions compared to the mesophilic proteins of the same kind. Besides that, complex ionic networking can stabilize the protein structure more than single isolated pairs (Musafia *et al*., 1995). We mentioned that the half life of our CGTase G1 is higher than some other mesophilic CGTases (Table 3). One of the possible reasons is because of richer ionic interactions was found in CGTase G1 than other mesophilic CGTases.

Examination on the ionic pair's location and the distance of interactions: The locations of the ionic pairs in each CGTases were further examined. The numbers of pairs in each domain of CGTase was shown in Table 3 (row No. 9-13). It was observed that most of the ionic pairs were located in Domain A of the protein structure. Assuming CGTase G1 is categorized as a pseduthermostable enzyme, the average number of total interactions in Domain A of thermostable CGTase (CGTase G1, CGTase *T. thermosulfurigenes EM1* and CGTase *Bacillus stearothermophiles*) is 49.7 while the average number for mesophilic CGTase is 45.3. Domain A is known as catalytic domain consisting of the active site and active site cleft. This domain has a tertiary structure of a TIM-barrel. To be able to stand inactivation as well as to maintain the enzyme activity at high temperature, the secondary and tertiary structures of Domain A must be stable. One of the possible ways to stabilize the conformation is to have a higher quantity of ionic interactions. It is observed from Table 3 that thermostable CGTases indeed have more ionic interactions.

Besides that, earlier studies (Leemhuis *et al*., 2004) suggested that Domain B is also important for the stability of CGTase. CGTase G1, CGTase *T. thermosulfurigenes EM1* and CGTase *Bacillus stearothermophiles* have an average total interaction of 7 pairs in Domain B while the average pair is 6 for mesophilic CGTases. If the cross domains interactions (between Domain B and other domains) were taken into consideration, the thermostable CGTases still have higher ionic pairs than the normal CGTases. In Table 3 (row No. 10), the number in brackets refers to the numbers of interaction between Domain B with other domains in the protein structure. The average number of interaction for thermophilic CGTase is 11.7 and for mesophilic CGTase it is 10. On the other hand, no apparent difference was observed for the interactions in Domain C-E as shown in Table 3 (row No. 11-13).

The ion pairs can actually be divided into three basic geometrical categories: salt bridges, N-O bridges and longer-range ion pairs. An ion pair is classified as a salt bridge when the side chain charged group centroids are within 4 Å distance and has at least one pair of Asp/Glu side chain carbonyl oxygen and Arg/Lys/His side chain nitrogen atoms are within 4 Å distance (Kumar and Nussinov, 2002a). Table 3 shows that CGTase G1, CGTase *T. thermosulfurigenes EM1*, CGTase *Bacillus stearothermophiles*, CGTase *B. circulans* 251, CGTase *B. circulans* No. 8 and CGTase *B.* sp. 1011 have salt bridges of 16, 18, 15, 19, 21 and 19, respectively.

An ionic pair is called N-O bridges when at least a part of the side chain nitrogen and oxygen atoms of the ion-pairing residues are within 4 Å but the centroid distance is greater than 4 Å. In the longer-range ion pairs, the side-chain charged group centroids as well as the side chain nitrogen and oxygen atoms are more than 4 Å apart. Previous continuum electrostatics studies indicate that most of the ion pairs have stabilizing electrostatic contributions when the centroids are within a 5 Å distance (Kumar and Nussinov, 2002a). Unfortunately,

there is no clear indication of correlation between the stabilizing ion pairs (centroids ≤ 5 Å) with the thermostability of CGTases as shown in Table 3. This could probably be due to the lack of sufficient data as there are only limited known CGTase structures to date and the statistical comparison might not be sufficiently representative. Nevertheless, all the earlier data revealed that there are differences between the thermophilic and mesophilic CGTases.

Ionic networking comparison: Protein stabilization by ionic networking is much greater than the contribution by individual interactions (Kumar *et al*., 2000). Location of networking is believed to be important as well. By comparing the networking pattern of various CGTases, important insight of enzyme stability might be gained. Figure 3 shows the amino acids that were involved in the largest ionic networking found in CGTase G1 and other five CGTases with known crystal structure. The more consensus residues from various CGTase were painted with the same color. The bolded-line indicates the salt-bridges, while the normal-line represents N-O pairs. Long ionic pairs are denoted by the dashed-line. As an overall picture, almost all residues in the networking were conserved. Those residues that are less conserved are not color painted. In the case for CGTase G1, twenty two residues (11 Asp+Glu and 11 His+Lys+Arg) were cluster-linked by three salt-bridges, six N-O pairs and seventeen long pairs. Generally in the case for ionic networking, Arg is an important key connector because its geometry allows 3 possible directions of interaction (Musafia *et al*., 1995) because it contains guanidium group in its side chain (Kumar and Nussinov, 2002a). Apart from that, Arg often forms a triad Asp-Arg-Asp in the ionic networking for proteins. There are four Arg (Arg368, Arg220, Arg185 and Arg579) in CGTase G1 and all these Arg were in the form of triad system (Fig. 3). Residues Arg368 and Arg220 were both involved in a multiple triad systems connecting Asp364- Arg368- Asp321- Arg220-Asp128-Asp222. For Arg185, it acts as a connector between Asp181 and Asp175. The fourth Arg found in CGTase G1 biggest networking was Arg579 and its connecting Asp194 and Asp628, acting as a conduit between Domain B and E. While Arg is the key connector for positively charged amino acid, Glu which is respectively longer than Asp serves as a connector for negatively charged residue (Musafia *et al*., 1995). Although statistical analysis showed that Glu is an important connector in ionic interactions (Musafia *et al.* 1995), however in the case of CGTase, only one Glu (Glu250) is present in the biggest networking in CGTase G1. This applies to the other five CGTases in Fig. 3.

Several differences could be pointed out between the six diagrams shown in Fig. 3. First, variation in distance (verifies by the bold, normal or dash lines) can be noticed for the other five CGTases. Second, besides the differences in distance, dissimilarity is noted at the top-left portion of each diagram in Fig. 3, as indicated by the arrow in CGTase G1. Lys44 and Asp41 of CGTase were not consensus. Interestingly CGTase *B. stearothermophilus* has a few extra residues not found in other CGTases. These amino acids were located near the TIM-barrel N-terminal of the CGTase.

The third variation is located at the bottom-right of each diagram, as shown by the arrow in the diagram for CGTase G1 in Fig. 3. Arg579-Asp628 (CGTase G1 numbering) were joined to the networking main body by Asp 194 (non consensus) which was linked to a histidine (His169). As in other five CGTases, for example in CGTase *T. thermosulfurigenes EM1*, Arg583- As 632 was directly linked to His172 (orange coloured).

The fourth variation is observed at the top-right region of CGTase G1, as shown by the dashed-oval in Fig. 3. Triad Asp181-Arg185-Asp175 was linked to the main networking body by His 226. However, the linking was only found in CGTase G1. In CGTase *T. thermosulfurigenes EM1*, the consensus residues Asp184-Arg188- Glu183 were located too far away from the main networking body resulting in separation among the two networks. About the same observation could be found in CGTase *Bacillus stearothermophiles*, where the triad Asp189-Arg193-Asp183 was not ionic-linked with the major networking. In mesophilic CGTase from *B. circulans* 251, non ionic-interaction were able to form at this region as the corresponding residues in primary sequence were non-charged amino acid (Asn). As in the case for CGTase *B. circulans* No. 8 and CGTase *B.* sp. 1011, isolated pair Asp182-Arg192 also stayed apart from the main networking body.

Leemhuis *et al.* (2004) applied the concept of salt bridges in improving thermostability of *B. circulans* 251 (BC 251) CGTase. By comparing the salt bridges networking in mesophilic BC 251 CGTase with thermostable CGTase from *Thermoanaerobacterium thermosulfurigenes* EM1, four unique salt bridges in the latter CGTase was identified and was introduced to the structure of the former by site-directed mutagenesis method (Leemhuis *et al*., 2004). A mutant being introduced with a new salt bridge has an interaction between Asp188 and Arg192 (BC 251 numbering). The mutant CGTase showed a significant half life increased of 7.5 fold compared to the CGTase from the wild type strain.

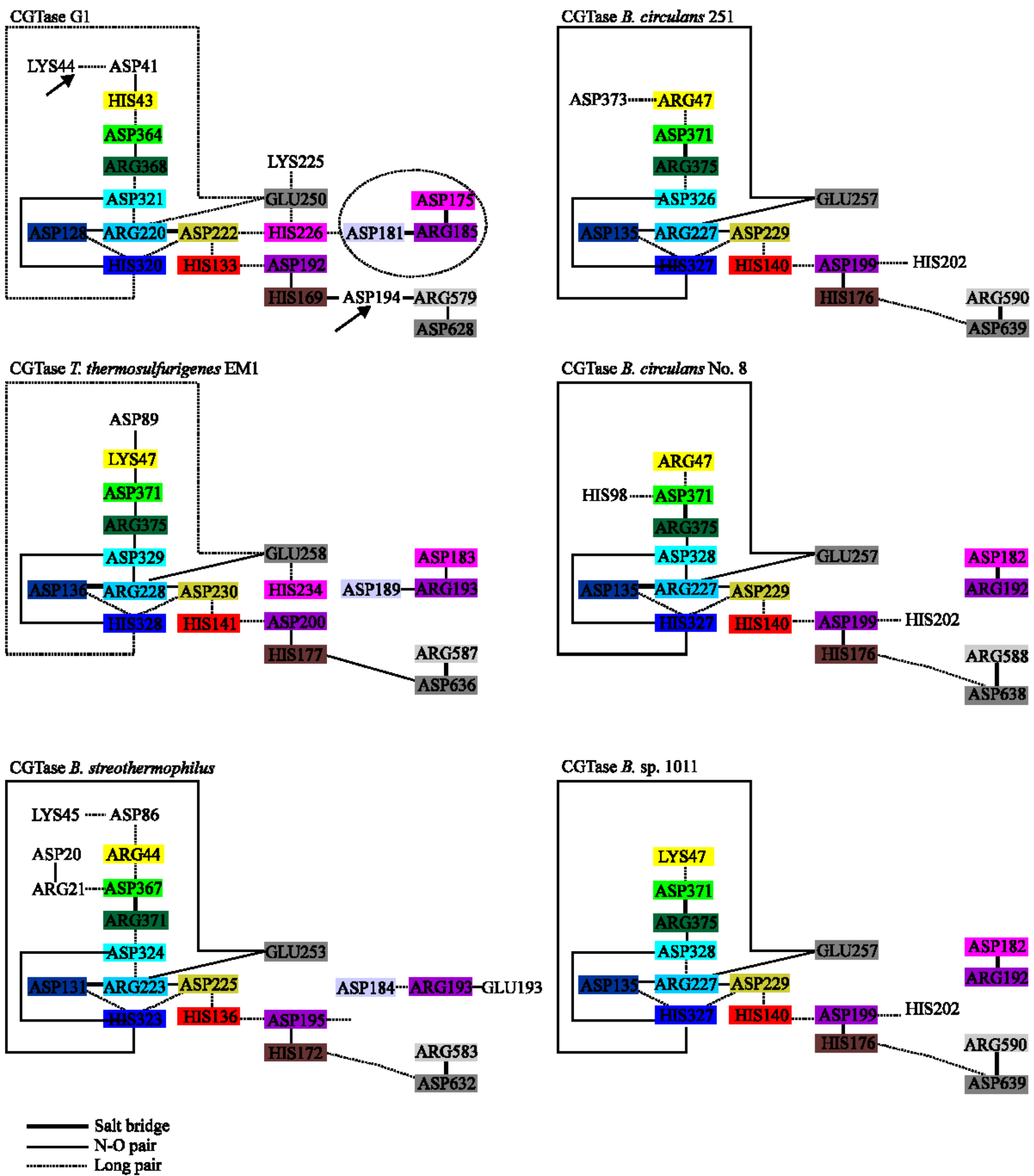

Fig. 3: The presence of biggest ionic interaction networking in CGTase G1 and five other CGTases with known crystal structures. The lines indicate the ionic interaction between the amino acids. The residues that are in the same position in amino acid alignment are indicated with the same background colors. The residues that are not consensus are not painted. Variations in distance can be observed for each CGTase. The two arrows and dashed-line oval indicate the other three main variations found in the structures

The new pair of ionic interaction introduced is actually the forth variation mentioned above for the biggest ionic interaction networking in CGTases. Before the mutation was made for CGTase BC 251, no naturally existing charged-interaction was observed because residue 188 was originally an Asn (uncharged amino acid). Substitution Asn188 to Asp and Lys192 to Arg provides a chance for residues 188 and 192 to form a close-distance

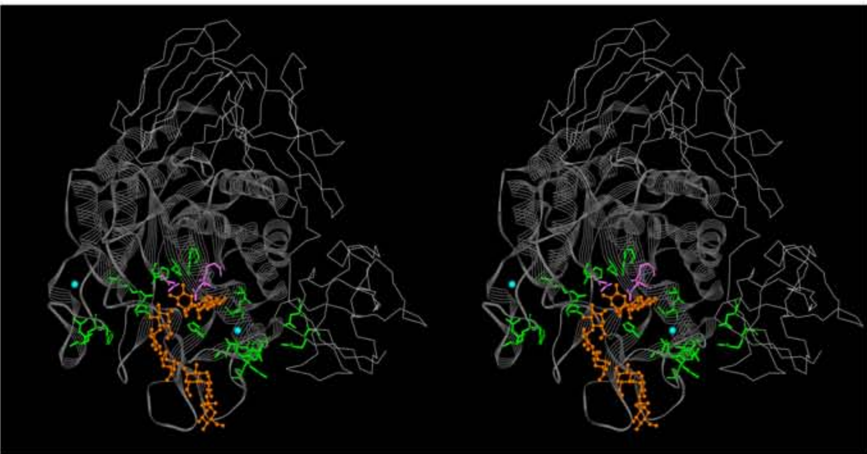

Fig. 4: Stereo representation of the biggest ionic networking in CGTase G1. Green: residues involved in the networking. Magenta: catalytic residues. Orange: substrate. TIM barrel (Domain A) is shown in faint ribbon representation. The ionic interactions grab hold of a few numbers of secondary strands and create a natural protection for the TIM- barrel structure against heat

ionic interaction. The location of the newly introduced salt bridge could be one of the crucial factors that determine the structural stability of CGTase. In CGTase G1, the corresponding residues are Asp181 and Arg185 (Fig. 3). Surprisingly, these two residues were located very closely in the 3D structure and have a strong salt bridge interaction with centroid distance of 3.49 Å (results not shown). Very likely a hydrogen bond is also present since the distance is less than 3.5 Å. This Asp181-Arg185-Asp175 ionic pairing of CGTase G1 is partially exposed to solvent and the theoretical bonds dissociate enthalpy can fall in the range of 12.5-21 kJ $mole^{-1}$ (Petsko and Ringe, 2002). The triad Asp181-Arg185-As 175 salt bridges pair is also in concert with the biggest ionic networking that exists in CGTase G1. Probably because of this, CGTase G1 has better tolerance ability to heat. As a conclusion, it was presumed that the triad Asp181-Arg185-Asp175 plays an important factor in the networking that caused the half life of CGTase G1 to be slightly higher compared to other CGTase originally produced by mesophilic strains, typically *B. circulans* 251 CGTase. CGTase G1 is indeed relatively more resistant to heat. Although it was originally produced from mesophilic *Bacillus* strain, it has a half life of 30 min at 60°C.

The locations of the residues that form the biggest networking in CGTase G1 is shown in Fig. 4. These 22 amino acids are in green colour. The catalytic residues of CGTase G1 are differentiated by magenta color. The substrate (starch) is in orange color. The locations of the two calcium binding sites are shown by blue dots. Twenty of the networking residues are actually located in Domain A and B of CGTase G1. Furthermore, it can be seen from Fig. 4 that the networking covers a huge area (from the left to the right) and it is placed surrounding the active site cleft. The ionic interactions grab hold of a few numbers of secondary structure (α-helice, β-sheet and loop) strands and create a natural protection for the TIM-barrel structure against heat (Hocker *et al*., 2001).

CONCLUSION

The thermostable CGTase is particularly important in converting starch to cyclodextrins. Although CGTase G1 was produced from a mesophilic *Bacillus* strain, it has a relatively higher half life than of other CGTases. Indeed ionic interactions can stabilize the structure of CGTase G1 and hence increases the life spends of the enzyme. The location of ionic interaction in the protein structure also plays an important factor in deciding the stability of the conformation.

ACKNOWLEDGMENTS

The author would like to thank Mohd Ghanee for technical advices. This research was financially supported by the National Biotechnology Division, Malaysia Ministry of Science, Technology and Innovation (MOSTI), under the project No. 09-02-05-006 BTK/ER/34.

REFERENCES

Chang, H.Y., P.M. Irwin and Z.L. Nikolov, 1998. Effects of mutations in the starch-binding domain of *Bacillus macerans* cyclodextrin glycosyltransferase. J. Biotechnol., 65 (2-3): 191-202.

Chakravarty, S. and R. Varadarajan, 2002. Elucidation of factors responsible for enhanced thermal stability of proteins: A structural genomics based study. Biochem. J., 41 (25): 8152-8161.

Colovos, C. and T.O. Yeates, 1993. Verification of protein structures: Patterns of nonbonded atomic interactions. Protein Sci., 2 (9): 1511-1519.

Greaves, R.B. and J. Warwicker, 2007. Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. BMC. Struct. Biol., 7 (1): 18-41.

Haney, P.J., J.H. Badger, G.L. Buldak, C.I. Reich, C.R. Woese and G.J. Olsen, 1999. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. Proc. Natl. Acad. Sci., 96 (7): 3578-3358.

Hendsch, Z.S., 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci., 3 (2): 211-226.

Hocker, B., C. Jurgens, M. Wilmanns and R. Sterner, 2001. Stability, catalytic versatility and evolution of the $(\beta\alpha)_8$-barrel fold. Curr. Opin. Biotechnol., 12 (4): 376-381.

Hyun-Dong, S., T.H. Park and Y.H. Lee, 2000. Site-directed mutagenesis and functional analysis of maltose-binding site of β-cyclodextrin glycosyltransferase from *Bacillus firmus* var. *alkalophilus*. Biotechnol. Lett., 22 (2): 115-121.

Jeon, S.J., S.W. Nam, J.W. Yun, S.K. Song and B.W. Kim, 1998. Effect of C or D-domain deletion on enzymatic properties of cyclodextrin glucanotransferase from *Bacillus stearothermophilus* NO_2. J. Microbiol. Biotechnol., 8 (2): 152-157.

Kaneko, T., T. Kato, N. Nakamura and K. Horikoshi, 1987. Spectrophotometric determination of cyclization activity of β-cyclodextrin-forming cyclomaltodextrin glucanotransferase. J. Jap. Soc. Starch Sci., 34 (1): 45-48.

Kim, M.H., C.B. Sohn and T.K. Oh, 1998. Cloning and sequencing of a cyclodextrin glycosyltransferase gene from *Brevibacillus brevis* CD162 and its expression in *Escherichia coli*. FEMS. Microbiol. Lett., 164 (2): 411-418.

Krieger, E. and G. Vriend, 2002. Models@ Home: distributed computing in bioinformatics using a screensaver based approach. Bioinformatics, 18 (2): 315-318.

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

Kumar, S. and R. Nussinov, 2001. How do thermophilic proteins deal with heat? Cellular Mol. Life Sci., 58 (9): 1216-1233.

Kumar, S., C.J. Tsai and R. Nussinov, 2001. Thermodynamic differences among homologous thermophilic and mesophilic proteins. Biochem. J., 40 (47): 14152-14165.

Kumar, S. and R. Nussinov, 2002a. Close-range electrostatic interactions in proteins. Chem. Biol. Chem., 3 (7): 604-617.

Kumar, S. and R. Nussinov, 2002b. Relationship between ion pair geometries and electrostatic strengths in proteins. Biophys. J., 83 (3): 1595-1612.

Larsen, K.L., H.J. Christensen, F. Mathiesen, L.H. Pedersen and W. Zimmermann, 1998. Production of cyclomaltononaose (deltacyclodextrin) by cyclodextrin glycosyltransferases from *Bacillus* sp. and bacterial isolates. Applied Microbiol. Biotechnol., 50 (3) 314-317.

Lee, M.H., S.J. Yang, J.W. Kim, H.S. Lee, J.W. Kim and K.H. Park, 2007. Characterization of a thermostable cyclodextrin glucanotransferase from *Pyrococcus furiosus* DSM3638. Extremophiles, 11 (3): 537-541.

Leemhuis, H., H.J. Rozeboom, B.W. Dijkstra and L. Dijkhuizen, 2004. Improved thermostability of *Bacillus circulans* cyclodextrin glycosyltransferase by the introduction of a salt bridge. Proteins, 54 (1): 128-134.

Li, W.F., X.X. Zhou and P. Lu, 2005. Structural features of thermozymes. Biotechnol. Adv., 23 (4): 271-281.

Lovell, S.C., I.W. Davis, W.B. Arendall, P.I. De Bakker, J.M. Word, M.G. Prisant, J.S. Richardson and D.C. Richardson, 2003. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. Proteins, 50 (3): 437-450.

Martins, R.F. and R. Hatti-Kaul, 2002. A new cyclodextrin glycosyltransferase from an alkaliphilic *Bacillus agaradhaerens* isolate: Purification and characterisation. Enzyme Microb. Tech., 30 (1): 116-124.

Musafia, B., V. Buchner and D. Arad, 1995. Complex salt bridges in proteins: Statistical analysis of structure and function. J. Mol. Biol., 254 (4): 761-770.

Petsko, G. and D. Ringe, 2002. Protein Structure and Function: From Sequence to Consequence. New Science Press, London.

Rahman, K., R.M. Illias, O. Hassan, N.A.N. Mahmood and N.A.A. Rashid, 2006. Molecular cloning of a cyclodextrin glucanotransferase gene from alkalophilic *Bacillus* sp. TS1-1 and characterization of the recombinant enzyme. Enzyme Microb. Tech., 39 (1): 74-84.

Ramazzotti, M., C. Parrini, M. Stefani, G. Manao and D. Degl'innocenti, 2006. The intrachain disulfide bridge is responsible of the unusual stability properties of novel acylphosphatase from *Escherichia coli*. FEBS. Lett., 580 (30): 6763-6768.

Rashid, N., J. Cornista, S. Ezaki, T. Fukui, H. Atomi and T. Imanaka, 2002. Characterization of an archaeal cyclodextrin glucanotransferase with a novel C-terminal domain. J. Bacteriol., 184 (3): 777-784.

Rodriguez, R., G. Chinea, N. Lopez, T. Pons and G. Vriend, 1998. Homology modeling, model and software evaluation: Three related resources. Bioinformatics, 14 (6): 523-528.

Russell, R.J., J.M. Ferguson, D.W. Hough, M.J. Danson, and G.L. Taylor, 1997. The crystal structure of citrate synthase from the hyperthermophilic archaeon *Pyrococcus furiosus* at 1.9 Å resolution. Biochem. J., 36 (333): 9983-9994.

Sian, H.K., M. Said, O. Hassan, K. Kamaruddin, A.F. Ismail, R.A. Rahman, N.A.N. Mahmood and R.M. Illias, 2005. Purification and characterization of cyclodextrin glucanotransferase from alkalophilic *Bacillus* sp. G1. Process Biochem., 40 (3-4): 1101-1111.

Szilagyi, A. and P. Zavodszky, 2000. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: Results of a comprehensive survey. Structure, 89 (5): 493-504.

Takada, M., Y. Nakagawa and M. Yamamoto, 2003. Biochemical and genetic analyses of a novel gamma-cyclodextrin glucanotransferase from an alkalophilic *Bacillus clarkii* 7364. J. Biochem., 133 (3): 317-324.

Thiemann, V., C. Donges, S.G. Prowe, R. Sterner and G. Antranikian, 2004. Characterisation of a thermoalkali-stable cyclodextrin glycosyltransferase from the anaerobic thermoalkaliphilic bacterium *Anaerobranca gottschalkii.* Arch. Microbiol., 182 (2-3): 226-235.

Vogt, G., S. Woell and P. Argos, 1997. Protein thermal stability, hydrogen bonds and ion pairs. J. Mol. Biol., 269 (4): 631-643.

Yamamoto, T., S. Fujiwara, Y. Tachibana, M. Takagi, K. Fukui and T. Imanaka, 2000. Alteration of product specificity of cyclodextrin glucanotransferase from *Thermococcus* sp. B1001 by site-directed mutagenesis. J. Biosci. Bioeng., 89 (2): 206-209.

Zhou, H.X., 2004. Loops, linkages, rings, catenanes, cages and crowders: Entropy-based strategies for stabilizing proteins. Accounts Chem. Res., 37 (2) 123-130.