สำนักงานวิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

CHAPTER IV
DISCUSSION

A proper understanding of the molecular basis of thermal stability in protein could have important consequences for their application in a range of biotechnological processes. The availability of enzymes with the appropriate specificity and capable of surviving for long periods at elevated temperatures could lead to the creation of novel applications of enzyme-based technology in industries such as those involved in the processing of starch, paper, pulp or fibres (Flam, 1994; Adams and Kelly, 1995). The opportunities afforded by such applications have generated significant interest in the field and a large number of studies have been undertaken to unravel the molecular mechanisms involved in generating thermostable enzymes (Yip *et al.*, 1998).

To improve thermostability of protein, a protein engineering approach is one of the many ways of studying. Engineering proteins for thermostability is a particular exciting and challenging field, as it is crucial for broadening the industrial use of recombinant proteins. Protein engineering is usually performed through site-directed mutagenesis, a favorite mode for genetic manipulation of the protein. Although the method is easy and highly useful, the rational mutagenesis requires detailed information on the tertiary structure and activities of the protein, the ability to predict the proper site of amino acid changes, and the intuition concerning the optimal amino acid to be changed (Kim *et al.*, 2003).

One strategy for the identification of amino acid sequence portions that correlate with thermostability involves the comparison of both structures and amino acid sequences of the more stable proteins with the less stable ones. In this study, it is hypothesized that at a given position in an amino acid sequence alignment of homologous proteins, the respective consensus amino acid sequence of the thermostable enzymes contributes more than average to the stability of the protein than the nonconsensus amino acids. Consequently, substitution of the nonconsensus sequences in the less thermostable enzymes by those consensus amino acid sequences

of the thermostable enzymes may be a feasible approach for improving the thermostability (Lehmann and Wyss, 2001). Thus, the amino acid sequence alignments between CGTases from *Bacillus circulans* A11 and other thermophilic CGTases from *Thermoanaerobacterium thermosulfurigenes* EM1 and *B. stearothermophilus* NO2, *Thermoanaerobacter* sp. ATCC53627 and *B. stearothermophilus* ET1 were carried out. The amino acid sequences were downloaded from the GenBank. Fig. 3.1 shows the sequence comparison performed using ClustalX software. Besides the several homologous sequences scattered all over the protein sequences, there are four major regions, designated as regions I-IV, respectively at positions 89-94, 265-271, 333-339 and 538-540 (A11 CGTase numbering) (Fig. 3.1) that are different from those of the thermostable CGTases. The first region is located in domain B while the second and third regions are in domain A, the two domains of CGTases that constitute the active site. The forth region are in domain D. The differences in the first three regions may relate to the thermostability of CGTase. The forth region is most likely not relate to thermostability because it is not found in some ordinary CGTases like the thermostable CGTases. To study the involvement of the first three regions in the thermostability of protein, the A11 CGTase gene was mutated and genetically manipulated such that the three mutant genes as well as the genes containing the various combinations of the three mutated regions were obtained. Their activity, thermostability, and products were determined thereafter.

To create the three mutations in the A11 CGTase gene, four mutagenic primers were designed according to the amino acid regions in the CGTase that are different from those of thermostable CGTases. Two mutagenic primers were designed for region I because there were two amino acid sequence patterns in the thermostable CGTases. Primer A replaced the wild type sequence YSGVHN with DSTFGGS that also contained an insertion of one amino acid, and primer B replaced the YSGVHN with DASGS that delete the sequence by one amino acid. For the region II, primer C put the sequence VDPNNHY in place of ISPEYHQ. For the region III, the sequences among the thermostable CGTases were somewhat variable. Therefore, the primer D is designed as mixed oligonucleotides by which the wild type HTSNGDR is replaced by $Y^{I}/_{N}{}^{D}/_{G}EGD^{P}/_{T}$. The actual mutant region III read YIGEGDT. The mutagenesis

changed the three regions in A11 CGTase in favor of the amino acid sequences in thermostable CGTases (Fig. 3.5). The four mutated plasmids, pRS1A, 1B, 2 and 3, containing the mutant regions I_A, I_B, II and III, respectively, were obtained as shown in Fig. 3.4. The mutated sequences as well as the sequences nearby were confirmed by DNA sequencing. To avoid the possibility of mutation in other regions in the plasmids, the mutant regions were subcloned into the wild-type CGTase gene replacing the equivalent sequences. Various CGTase mutants were also constructed such that they contained all possible combinations of the mutated regions. These resulted in pRS4A, 4B, 5, 6, 7A, 7B, 8A, 8 B, 9, 10A and 10B which contained mutation regions I_A, I_B, II, III, I_A+II, I_B+II, I_A+III, I_B+III, II+III and I_A+II+III and I_B+II+III, respectively.

The three-dimensional structures of A11 β-CGTase (wild type CGTase) and the mutant CGTases were predicted by using homology modeling (www.cbs.dtu.dk/services/CPHmodels). Since the amino acid sequence of A11 CGTase is about 98% homologous to that of CGTase from *Bacillus* sp. 1011 (PDB 1I75), whose three-dimensional structure is already known, the predicted structure of the A11 CGTase was found to be similar to that of CGTase from *Bacillus* sp. 1011 (Fig. 4.1). The predicted three-dimensional structures of pRS10A and 10B are also very similar to that of the wild type A11 CGTase (Fig. 4.2). Comparison of the conformation of the amino acids around the mutated residues showed clearly that no large structural rearrangements had taken place as a result of amino acid changes. The structures of mutant CGTases from pRS10A and 10B also revealed the location of the three mutation regions. Region I_A was located on the surface of the enzyme and resided on the loop regions, which upon amino acid changes seemed to have no effect on the folding of the main chain α-helices and β-sheets. Likewise, the mutation region I_B, located at the same region as mutation region I_A, had no influences on the folding of the A11 CGTase. The mutation regions II and III were also located on the surface of the enzyme, and retained more or less the wild type conformation. These mutation regions resided outside the active site, and perhaps did not have any influence on the active site structure of the enzyme. This might be the reason why all the CGTase mutants retained their catalytic activity though some alterations in their activity did occur.

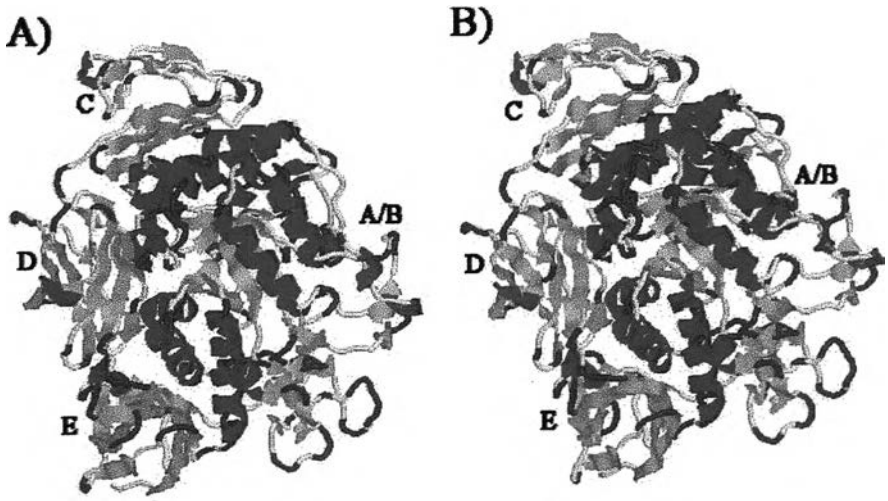

Fig. 4.1. Comparison of the three-dimensional structures between the CGTase from *Bacillus* sp.1011 (PDB 1I75) (A) and the CGTase from *B. circulans* A11 (B).

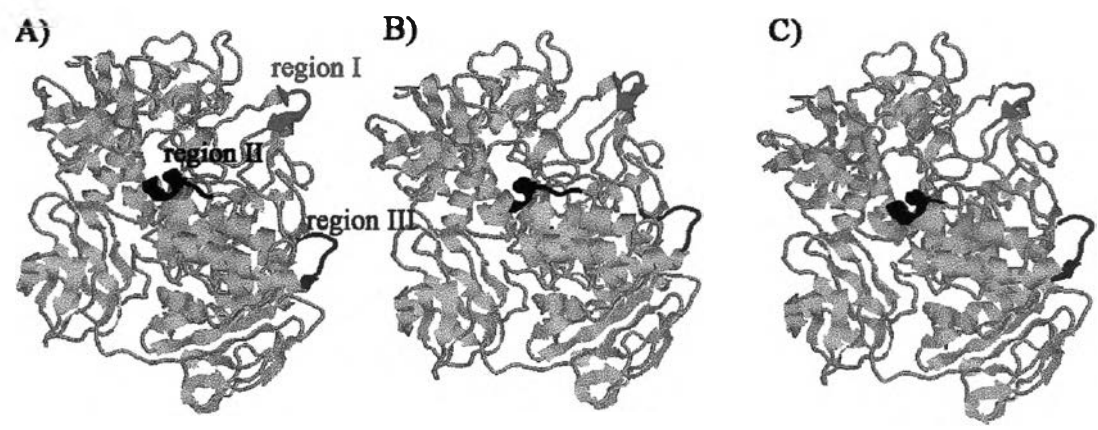

Fig. 4.2. Comparison of the three dimensional structures between the wild type CGTase from *B. circulans* A11 (A) and the mutant CGTases from the recombinant plasmids pRS10A (B) and pRS10B (C).

It was observed that all the mutants exhibited the dextrinizing activity. The *E. coli* JM109 transformants harboring each of the mutant CGTase gene were tested for dextrinizing activity on the LB-starch agar plates (Figs. 3.9-3.10 and Table 3.1). The halos that appeared around the colonies after exposure to a KI-I_2 indicator solution showed that starch around the colonies had been degraded by the hydrolytic activity

of the CGTases. At 40 °C, mutant CGTases have slightly higher dextrinizing activity than the wild type (Fig. 3.10 and Table 3.1). By studying the crystal structure of CGTase from *Thermoanaerobacterium thermosulfurigenes* EM1, Knegtel *et al.* (1996) had observed 3 amino acid sequence regions that were different from those of *B. circulans* 251 CGTase (similar to A11 CGTase with ~86% homology). They hypothesized that the differences might contribute to thermostability in EM1 CGTase. Two of the three regions were at the same positions as regions I and III in this study. The EM1 CGTase had a much higher hydrolytic activity than the 251 CGTase. Knegtel *et al.* (1996) speculated that the lack of Tyr89 in EM1 CGTase rendered it unable to sufficiently bind substrate strong enough for efficient cyclization. By changing the A11 CGTase region I amino acid sequence from YSGVHN to either DSTFGGS or DASGS, the Tyr89 (Y) were absent in the mutants. The results seemed to agree well with the speculation of Knegtel *et al.* (1996). However, other mutants with the Tyr89 also exhibited higher hydrolytic activity (see below). We argued here that the lacking of Tyr89 was not the major cause of higher hydrolytic activity in the A11 CGTase mutants in this study and perhaps not in EM1 CGTase either.

Mutations in regions II and III also resulted in higher dextrininzing activity in A11 CGTase. Region II was very closed to the acceptor subsite +3 that was possibly only found in *T. thermosulfurigenes* EM1 and other thermostable enzymes whereas region III was far away from the reactive cleft (Fig. 4.3 and 4.4). It had been proposed that the acceptor subsite +2 had the large effect on the hydrolytic activity of CGTase (Leemhuis *et al.*, 2003). One might think that mutation in region II (subsite +3), which was closed to subsite +2, could disturb the integrity of the subsite +2 and resulted in an increase in hydrolytic activity. However, the data presented here did not support the above notion since mutation in region III also resulted in comparable higher hydrolytic activity. The combination of mutant regions I+II, I+III, II+III and I+II+III also provided better dextrinizing activity than the wild type enzyme.

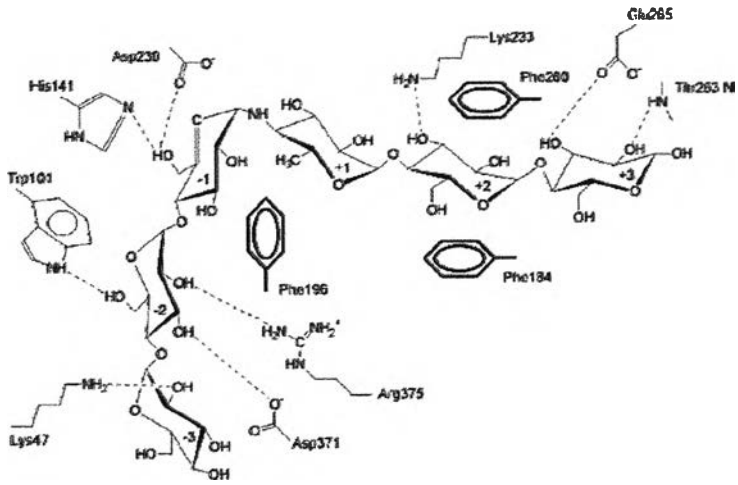

Fig. 4.3. Schematic view of the interactions between the EM1 CGTase and a maltohexaose inhibitor bound from subsites –3 to +3 (Leemhuis *et.al.*, 2003).

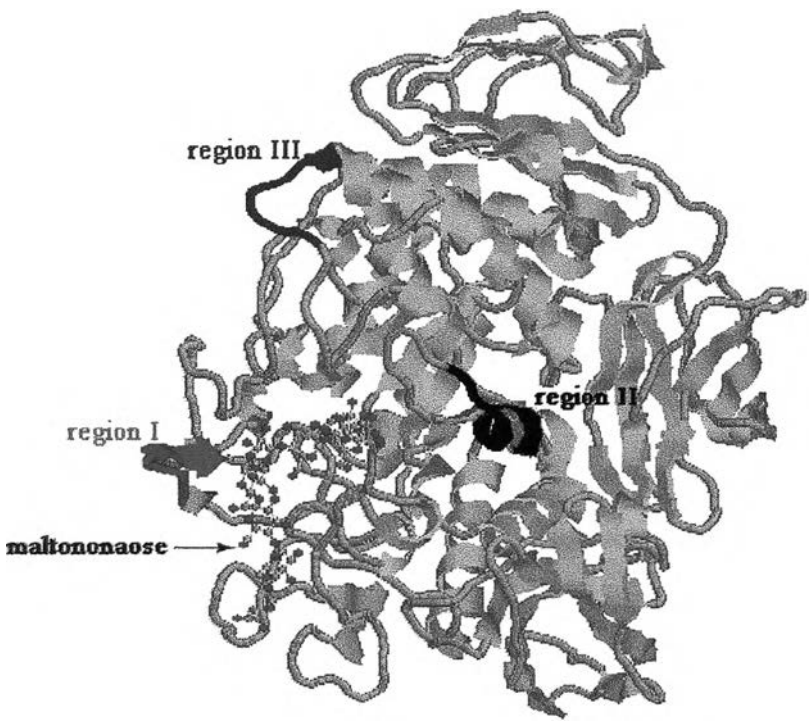

Fig. 4.4. Location of the three mutations regions relative to the binding subsites, presented by the maltononaose.

The crude mutant enzymes were prepared by culturing each clone in LB medium. The crude enzyme was obtained after the removal of cell pellet. The enzyme preparation was subjected to activity assay to determine the optimum temperature and thermostability. The optimum temperature was determined by measuring the dextrinizing activity at various temperatures from 40-80 °C (Table 3.1 and Fig. 3.11). Mutation at regions I and II, but not III, slightly reduced the optimum temperatures. The combination of mutant regions, I+II and I+II+III, but not I+III or II+III, seemed to be the most deleterious. It was probably the combined effect of mutant regions I and II.

In literature, the stability of thermopilic proteins has been described in a number of ways, such as the temperature at which a protein is active (activity temperature), stable (stable temperature) or the half-life for certain duration of time. Much less frequently, the stability is described in terms of melting or mid-point transition temperature (*T*m) (Kumar *et al.*, 2000). In this study, the thermostability of the mutant CGTases was described in terms of temperature as a function of the incubation time. It was determined by incubating the enzymes at various temperatures for 10 minutes, and then the residual starch-hydrolyzing activity of CGTase was assayed at certain optimum temperature, e.g. 40 °C. The thermostability profiles in Fig. 3.12 reveal that mutations in regions I and II but not III reduce the thermostability of the enzyme. The combination of mutant regions does not seem to give more deleterious effect on the thermostability of the enzyme.

Leemhuis, *et al.* (2004) attempted to localize the amino acid regions involved in the thermostability of *B. circulans* 251 CGTase (251 CGTase) by comparing the three-dimensional structures of the five CGTases with known crystal structures and mutating the 251 CGTase in different regions to mimic the thermostable enzymes. The approach was very similar to this study. Five structural differences between the 251 CGTase and the EM1 CGTase were found in loop regions 88-94, 334-339, 494-498, 536-542, and 658-660 (A11 CGTase numbering). Two of them were identified in this study, namely regions I (89-95) and III (333-339). Almost identical change was made in region I_A (88-NYSGVNN to PDSTFGGS in 251 CGTase and 89-YSGVHN to DSTFGGS in A11 CGTase) but different change in region III (334-ASNANR to NGGST in 251 CGTase and 333-HTSNGDR to YIGEGDT in A11 CGTase). The changes in all three regions in A11 CGTase did not improve the thermostability of

A11 CGTase. Similar to our finding, Leemhuis, *et al.* (2004) also found that all these changes towards thermostable CGTase had no effect on the thermostability of 251 CGTase.

Nevertheless, by providing a salt-bridge between positions 188 (N188D) and 192 (K192 or K192R), the half-life of the 251 CGTase at 60 °C was moderately increased from 9.7 min to an average of 45 min whereas the EM1 CGTase retained full activity after 24 hours of incubation. Introduction of the other 3 salt-bridges found in the thermostable enzymes at positions R47-D89, D244-K510 and E275-K560 had no or deleterious effect on the half-life of 251 CGTase. Likewise, the A11 CGTase also contained the N188 and K192, and would give similar result upon mutation to introduce the salt-bridge. Mutation in region I of A11 CGTase also resulted in Y89D that possibly formed salt-bridge with R47 and also did not improve the thermostability of the enzyme. Thus, it was reasonable to conclude that most of the structural and salt-bridge differences between the mesophilic CGTases and thermostable CGTases were not of primary importance for the thermostability.

To determine the CD forming activity by the mutant CGTases, the crude enzyme from each mutant plasmid was incubated with soluble starch at 37 °C for 12 hours. The remaining oligosaccharides were digested with glucoamylase, and the reaction products were subjected to HPLC analysis. HPLC profiles of the CGTase reactions indicated that the mutant CGTases produced the cyclodextrins at various ratios, and some of them were different from that of the wild type (Fig. 3.13 and Table 3.3). For all the CGTase mutants, the proportion of γ–cyclodextrin produced was relatively unchanged or insignificantly decreased. The presence of region I mutants (pRS4A, 4B, 7A, 7B, 8A, 8B, 10A and 10B) resulted in an increase in α-cyclodextrin production with the expense of β-cyclodextrin production. Mutation in region II (pRS5) slightly decreased the proportion of γ-cyclodextrin and increased those of either β-cyclodextrin or α-cyclodextrin or both. Mutation in region III (pRS6) slightly decreased the proportion of γ-cyclodextrin and increased that of α-cyclodextrin. The mutants that had region I mutations in combination with the other two mutation regions (pRS7A, 7B, 8A, 8B, 10A and 10B) all had increased proportion of α-cyclodextrin. The influence of mutation in regions II and III was not obvious in these mutants; the proportion of γ-cyclodextrin was slightly decreased in

pRS7A, 7B, 8A and 8B or unchanged in pRS10A and 10B. Interestingly, the combination of mutated regions II and III (pRS9) reduced the production of α-cyclodextrin, and increased that of β-cyclodextrin.

By using site-specific mutagenesis, van der Veen *et al.* (2000b) had reported that the mutations Y89D and S146P in 251 CGTase resulted in an increase in α-cyclodextrin production. The Y89 was found to hydrophobically interact the substrate in subsite –3, whereas the S146 was involved in hydrogen bonding in subsite -7. It was therefore concluded that changes in subsites –3 and –7 resulted in changes in product specificity. In our study, mutation in region I, 89-YSGVHN to DSTFGGS in pRS4A and DASGS in pRS4B possibly created a salt-bridge with R47 like that found in thermostable enzymes but did not increase the thermostability of the enzyme. The mutation in turn resulted in a slight shift towards α-cyclodextrin production in agreement with the result of van der Veen *et al.* (2000b). Although the size of the loop region I in pRS4B was smaller (Fig. 4.4), the observed cyclodextrin product ratio was relatively similar to that of pRS4A. On the other hand, Chotechuang (2003) had previously replaced the sequence around region I of the A11 CGTase from 87-INYSGVHN to HPGGF, which reduced the size of the loop region I by two amino acids more than the pRS4B CGTase. She found that the mutated enzyme produced mainly β-cyclodextrin and reduced the production of both the α- and γ-cyclodextrins. We argue here then that the subsite -3 was involved in the product specificity, not just for the α-cyclodextrin production but also β-cyclodextrin production. The size of the loop, but not the hydrophobic interaction of Y89 with the substrate, might play an important role in this respect.

For mutation in region II, it was interesting to find that the mutation in this region (pRS5) which resides very close to the supposedly subsite +3 had no influence on the product specificity. The acceptor binding subsite +3 was observed in *T. thermosulfurigenes* EM1 but not other mesophilic CGTases. Since the binding was very weak, it was believed to be less relevant to the enzyme activity (Wind *et al.*, 1998). Our result showed that the change in region III sequence of the A11 CGTase towards those of the thermostable enzymes provided very little effect, if any, on the enzyme activity. Contrarily, the mutation in region III (pRS6) which lay far away from the active site slightly increased the α-cyclodextrin production and slightly

decreased the γ-cyclodextrin without affecting the β-cyclodextrin production. The influence of mutation in region III on the enzyme activity is not understood.

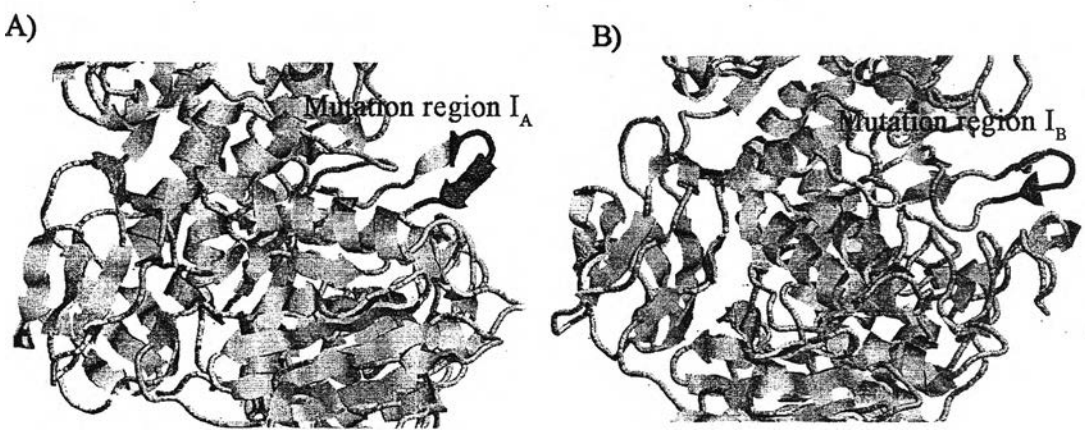

Fig. 4.5. Comparison of the three-dimensional structures of CGTase at mutation region I (subsite-3) between pRS10A (A) and pRS10B (B).

From our results, we had achieved our twofold goals. First, we were able to determine that the differences in the amino acid sequences, the structure as well, between the A11 CGTase and the thermostable CGTases were not the major determinants for thermostability of the enzyme. Second, we had tested the effect of mutation on the product specificity. We had found that mutations in region I and III but not II altered the enzyme product specificity. In other and our studies, we believed that neither structural feature nor disulfide bridge alone was adequate for the thermostability of enzyme. Thermostability seemed rather to be caused by a complex and subtle interplay of many different factors, and it was often dependent on the function and environment of the protein (Fontana, 1991 and Jaenicke, 1991). The discrepancies in the relative importance of individual factors to stability arose from a combination of different molecular interactions that were differentially weighted on a case by case basis (Yip, 1998). Nevertheless, the key factor lies out there perhaps in the primary sequence and will soon be unveiled.