ELSEVIER

Journal of Molecular Catalysis B: Enzymatic 4 (1998) 167–180

JOURNAL OF MOLECULAR CATALYSIS B: ENZYMATIC

Review

Protein thermostabilization by proline substitutions

Kunihiko Watanabe, Yuzuru Suzuki *

Laboratory of Microbial Physiology and Applied Microbiology, Department of Applied Biochemistry, Kyoto Prefectural University, Shimogamo, Sakyo, Kyoto 606, Japan

Received 22 July 1997; revised 20 August 1997; accepted 20 August 1997

Abstract

Many recent approaches involving site-directed mutants have succeeded in increasing the thermostability of proteins. It is well known that replacements with proline residues reduce the conformational degrees of freedom in the main polypeptide chain and thus can increase protein thermostabilization. We have studied protein thermostabilization by introducing proline substitutions in the homologous oligo-1,6-glucosidases from various *Bacillus* strains which grow within different temperature ranges. As a consequence, the 'proline rule' was proposed for protein thermostabilization. The principle of this rule is that an increase in the frequency of proline occurrence at β-turns and/or an increase in the total number of hydrophobic residues can enhance protein thermostability. We have generated several lines of evidence supporting the theory from the comparative analysis of oligo-1,6-glucosidases in their primary and secondary structures and molecular properties, the X-ray crystal structure analysis of the *Bacillus cereus* oligo-1,6-glucosidase, and the enhancement in thermostability of the oligo-1,6-glucosidase by cumulative replacements with prolines. As a new finding from the studies, two specific sites (second positions at β-turns and N1 positions of α-helices) were found to be the most critical to protein thermostabilization dependent on several structural prerequisites for proline substitution.

Keywords: Proline; Site-directed mutagenesis; Thermostabilization; Protein stability

1. Introduction

Enzyme proteins are useful biocatalysts for various industrial applications. One major limitation of these proteins is their susceptibility to heat inactivation. In the industrial use of such proteins, the costs of this limitation account for significant parts of the total expenses. Therefore, at the rise of protein engineering, excessive expectations were held for protein thermostabilization. This was because, by a newly devised method, the so-called site-directed mutagenesis, it became possible to freely engineer in vitro all primary structures of proteins [1]. Many people working in the fields of biochemistry and biotechnology had hoped that every useful or expensive protein could be thermostabilized, once the genes responsible for the proteins were cloned, by following a general strategy for site-directed mutagenesis. Many attempts for thermostabilization of various proteins, without regard to how specific mutations affect the thermostability of proteins, have been reported in this last decade. Although several successes showed the positive effect of specific mutations on protein thermostability, it has be-

* Corresponding author. Tel./Fax: (81-75) 703-5664; E-mail: y_suzuki@kpu.ac.jp

1381-1177/98/$19.00
PII S1381-1177(97)00031-3

come apparent that such mutations are seldom applicable to other enzyme proteins. Proteins are composed of different, long stretches of amino acid residues, which fold with enormous numbers of minute and various interactions. It is unrealistic to expect that one or two amino acid substitutions could cause the drastic enhancement in the thermostability of a protein while also preserving protein functions.

There has been much argument in the recent reviews about whether general rules or principles for thermostabilizing proteins exist in nature [2–6]. The conclusions have been controversial for a long time, but the studies to identify molecular determinants of protein structure and functional stability are well recognized as rather profitable for protein thermostabilization. These studies are on, for example, (i) increase of hydrophobic interaction [7], (ii) increase of packing efficiency [8], (iii) stabilization of the dipoles of α-helices [9], (iv) stabilization of salt bridges [10], (v) reduction of conformational strain [11], (vi) reduction of the area of water accessible hydrophobic surface [12] and (vii) reduction of the entropy of unfolding [13]. These are inferred from the comparison between homologous enzyme proteins from thermophiles and mesophiles or from the analysis of critical interactions to protein structures and functions. Independent of these studies for protein engineering, we focused on proline residues in oligo-1,6-glucosidases from *Bacillus* strains with different ranges of growth temperature. Through studies on these enzymes, we have succeeded in cumulative thermostabilization of the protein by following a steady strategy for proline substitutions. In this review, we introduce our molecular biological studies on the oligo-1,6-glucosidases and then show protein thermostabilization by proline substitutions.

2. Special features of proline

Proline differs from all other amino acids because the side chain curls back to the preceding peptide-bond nitrogen and forms the five-member pyrrolidine ring (Fig. 1). Strictly speaking, from such a structural viewpoint proline is an α-imino acid rather than an α-amino acid. Since the pyrrolidine ring imposes rigid constraints on the N–C$^{\alpha}$ rotation, the conformation of the proline residue in a polypeptide chain is ultimately limited [14,15]. Therefore, proline has less backbone configurational entropy than other amino acid. Schimmel and Flory demonstrated that the conformation energy of proline in polypeptide chains depends mainly on its conformational angles of φ and that there are two minima at $\varphi = -55°$ and $+145°$ for an isolated proline residue [15]. MacArthur and Thornton subsequently revealed from X-ray crystallographic analysis data of the Brookhaven Protein Data Bank [16] that the conformations of 963 *trans* proline residues are clustered in the mean values of $(\phi, \varphi) = (-61°, -35°)$ for the α region and of $(\phi, \varphi) = (-65°, 150°)$ for the β region [14]. The pronounced feature appears as well in the limited regions for proline residues in the Ramachandran plots [14,17]. It is also noted that the conformation adopted by proline is affected by the preceding residue [18].

Proline residues occurring in proteins have a unique propensity for secondary structures due to their structural peculiarity. Notably, they have a preference for second positions of β-turns

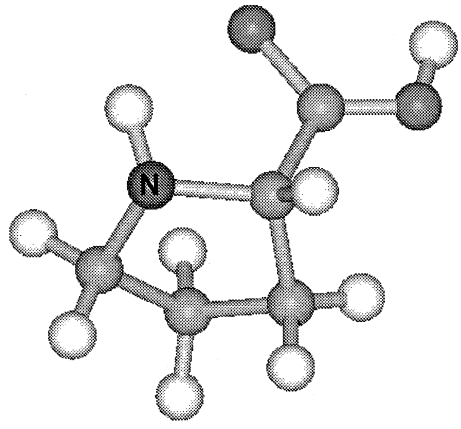

Fig. 1. Molecular structure of proline.

[19–21]. From an up-to-date analysis based on 3899 β-turns of 205 proteins, this preference has been extensively rationalized. In addition, prolines have been found to occur frequently at the first positions of type I, II and VIII β-turns [22]. On the other hand, it is generally understood that proline plays a significant role as a helix-breaker due to the absence of N-hydrogen to form a helix interior hydrogen-bond [20,23]. In globular proteins, helices containing proline in their middle positions are very rare. About one third of proline residues in proteins occur in random coils and flexible loops, which are structures other than α-helix, β-sheet and β-turn [14,22]. However, with the recent increase in the number of well-determined three-dimensional protein structures, the high frequency of proline occurrence in the N-terminal first turns of α-helices has been reported as an important exception [24–26]. In the first turns, proline residues are predominantly found at the N1 positions. At this N1 position, a kink in the peptide-backbone introduced by proline can thermodynamically stabilize an α-helix [26]. Furthermore, von Heijne reported that a proline residue within the first three positions does not destabilize an α-helix because amide nitrogens of the first four residues in an α-helix do not form hydrogen bonds [27]. Thus, proline can play the critical role of helix initiator as well as of helix breaker.

These structural features of proline are biologically significant for many proteins. For example, proline-rich motifs in human immunodeficiency virus (HIV-1) Nef protein are critical in binding to SH3 domains of a subset of Src kinases, and in the enhanced growth of Nef^+ viruses [28]. Despite of the general lack of proline residues in helices of globular proteins, it has been speculated that proline plays a crucial role in the transmembrane helices of hormone receptors and of energy- and signal-transducing proteins [29,30]. In these proteins, the mutation of critical proline residues to other amino acid residues leads to the drastic loss of protein function due to the increase of flexibility in the conformation of the residues around the binding sites [30]. Besides, peptidyl proline *cis–trans*-isomerase (PPIase) catalyzes the interconversion of the *cis* and *trans* isomers of the peptidyl-prolyl bonds in peptide and protein substrates [31]. Some of the PPIases bind clinically useful immunosuppressants such as cyclophilins and FK506 [31]. As the frequency of organ transplantation in the medical field increases, much attention has been paid to the peptidyl proline bond. Thus, the structural properties of proline, unique among the amino acids, are crucial for the function of a wide range of proteins.

Another important feature is that proline can stabilize proteins from their structural peculiarity. It is our great concern to determine how proline residues have contributed to the evolution of thermostable proteins and to discover the principles of protein thermostabilization used by nature.

3. Oligo-1,6-glucosidases from various *Bacillus* strains with different growth temperature

The molecular and catalytic properties of *p*-nitrophenyl-α-D-glucopyranoside-hydrolyzing oligo-1,6-glucosidases (dextrin 6-α-glucanohydrolase, E.C.3.2.1.10) from six *Bacillus* species strains with different ranges of growth temperature (*B. cereus* ATCC7064 which grows at 10–40°C [32]; *B. coagulans* ATCC7050 at 30–55°C [33]; *B. thermoamyloliquefaciens* KP1071 at 30–66°C [34]; *B. thermoglucosidasius* KP1006 at 42–69°C [35]; *B. caldotenax* KP1213 below 85°C, but optimum at 65–72°C [36]; *B. flavocaldarius* KP1228 at 51–82°C [37]) (see Fig. 2) have been progressively investigated. The first five oligo-1,6-glucosidases are molecularly homologous monomeric globular proteins with a similar molecular weight, while the most thermostable enzyme from the extreme thermophile *B. flavocaldarius* KP1228 consists of two identical subunits with a molecular weight similar to those of other monomeric

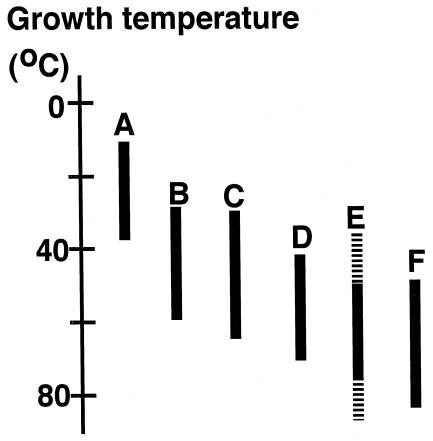

Fig. 2. Growth temperature range of six *Bacillus* strains. (A) *B. cereus* ATCC7064; (B) *B. coagulans* ATCC7050; (C) *B. thermoamyloliquefaciens* KP1071; (D) *B. thermoglucosidasius* KP1006; (E) *B. caldotenax* KP1213; (F) *B. flavocaldarius* KP1228.

enzymes [36,37]. The thermostability of each enzyme is well correlated with the range of growth temperature suitable for the corresponding strain. From the detailed comparison of those oligo-1,6-glucosidases in terms of thermostability, amino acid composition and structural parameters calculated from the amino acid composition, it was proposed that enhanced thermostability of *Bacillus* oligo-1,6-glucosidases would be gained by increasing the frequency of proline occurrence at β-turns (presumably their second sites) and by increasing the total numbers of hydrophobic residues (the proline rule, formerly named the proline theory) (Fig. 3) [34]. This relationship has been found to hold for sixteen bacterial enzymes from four other different groups (α-glucosidases, pullulanases, neopullulanases and 3-isopropylmalate dehydrogenases) [36]. Moreover, it was suggested that the proposal for *Bacillus* oligo-1,6-glucosidases has the nature of a general principle for increasing protein thermostability [38]. To provide additional lines of evidence to confirm the proposal we began analyses of the gene encoding each oligo-1,6-glucosidase, using molecular biological techniques.

Gene cloning and the following DNA sequencing were undertaken, focusing on the oligo-1,6-glucosidases of *B. cereus* ATCC7064 [39], *B. coagulans* ATCC7050 [40] and *B. thermoglucosidasius* KP1006 [41,42]. Then their primary structures were clarified (Fig. 4). The features of the three enzymes are summarized in Table 1. Total amino acid numbers and molecular weights are about 560 and 66 000 in all cases. The identities in primary structure among three enzymes are 57–72%. In particular, the *B. cereus* enzyme and the *B. thermoglucosidasius* enzyme have the highest identity (72%). As the high identity is reflected in the DNA sequence, the genes for these two enzymes could be fused and thus chimeric genes were constructed as mentioned at the end of this chapter. The number of proline residues, which was of greatest interest to us, was 19 for the *B. cereus* enzyme, 24 for the *B. coagulans* enzyme and 32 for the *B. thermoglucosidasius* enzyme [39,40,42]. The proline contents were comparable to those of the intact enzymes purified from the bacillary strains, respectively. The correlation of proline

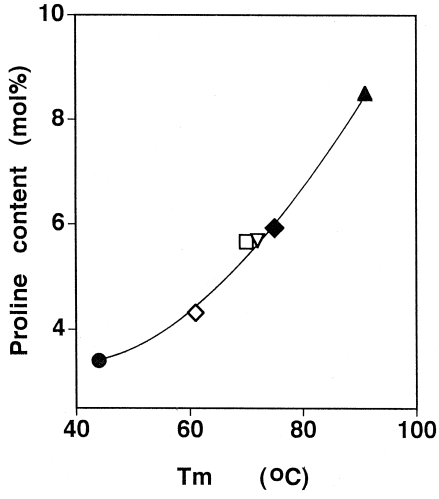

Fig. 3. Strong correlation between the increase in the proline contents of six *Bacillus* oligo-1,6-glucosidases and the increase in their thermostability (T_m). The proline contents of the oligo-1,6-glucosidases from *B. cereus* ATCC7064 (●) [39], *B. coagulans* ATCC7050 (◇) [40], *B. thermoamyloliquefaciens* KP1071 (□) [34], *B. thermoglucosidasius* KP1006 (▽) [42], *B. caldotenax* KP1213 (◆) [36], *B. flavocaldarius* KP1228 (▲) [37].

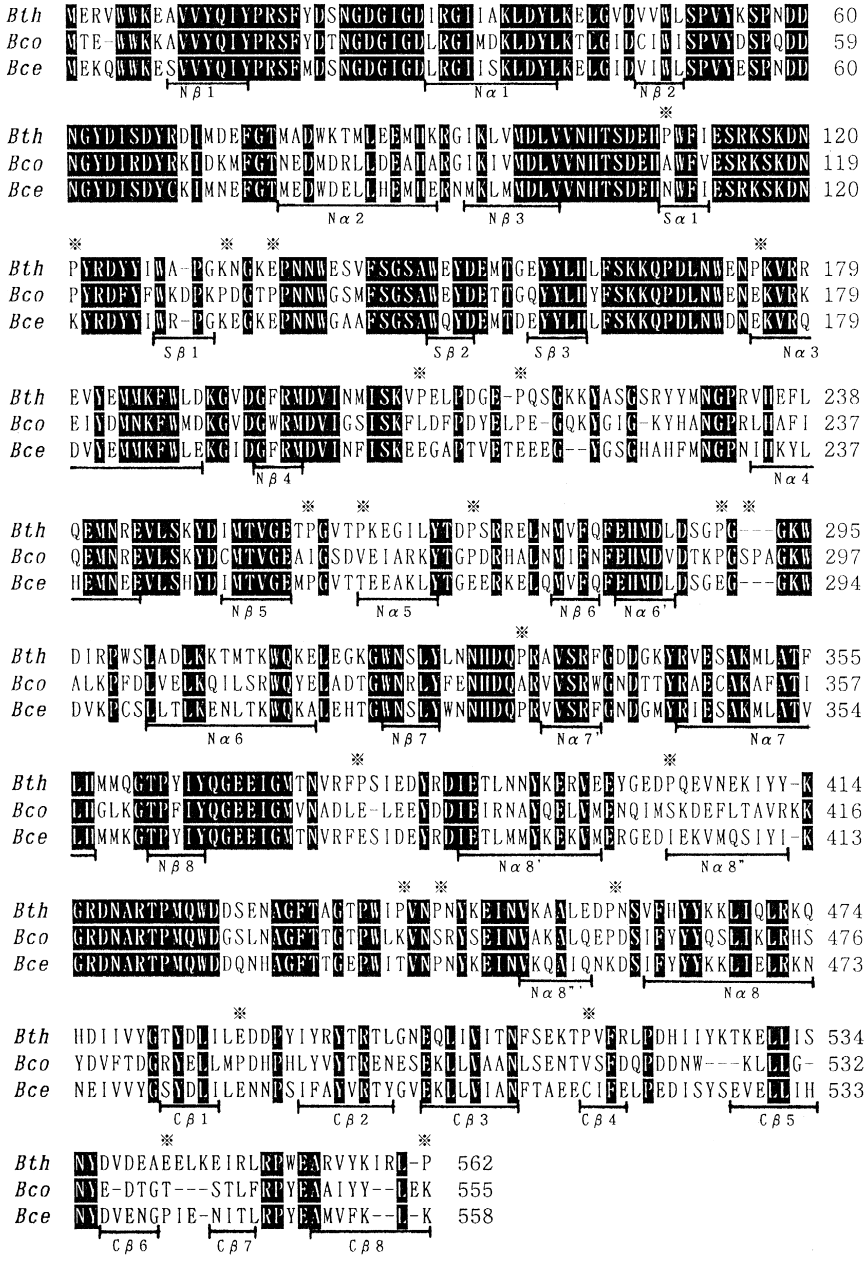

Fig. 4. Comparison of the primary sequence and secondary structural assignments of three bacillary oligo-1,6-glucosidases. The sequences are shown in the order. *B. thermoglucosidasius* KP1006 (*Bth*), *B. coagulans* ATCC7050 (*Bco*) and *B. cereus* ATCC7064 (*Bce*). Identical residues are black-boxed. The critical prolines are indicated with a dotted cross. Each secondary element is shown with a bar.

content with protein thermostability was proved to be consistent as shown in Fig. 3. The oligo-1,6-glucosidases expressed in *E. coli* cells are molecularly and catalytically identical with the corresponding intact enzymes. Throughout our studies we have used temperatures for 50% inactivation (T_m) by 10 min incubation as the standard for thermostability.

Table 1
Comparison of the oligo-1,6-glucosidases from *B. cereus* ATCC7064, *B coagulans* ATCC7050 and *B. thermoglucosidasius* KP1006

	B. cereu ATCC7064	*B. coagulans* ATCC7050	*B. thermoglucosidasius* KP1006
Molecular weight	66,010	64953	66,501
Total amino acid residues	558	555	562
Pro residue number (conten %)	19 (3.14)	24 (4.32)	32 (5.69)
Tm (°C)	45	61	72
Homology (%)	57		59
		72	

The proline residues of the three oligo-1,6-glucosidases were classified according to their primary sequence alignment (Fig. 5) [40]. As the enzymes showed high identity in primary structure, all proline residues could be assigned to corresponding residues in other primary sequences except for Pro293 unique in the *B. coagulans* enzyme (Fig. 4). These three enzymes share 15 common proline residues. The 15 proline residues are essential for maintaining the enzyme structure and function. On the other hand, the rest of the proline residues in these enzymes, representing 21 proline residues plus Pro293 unique to the *B. coagulans* enzyme [40], were supposed to occur independently of the basic requirements for the protein structure and function, but instead might be critical to protein thermostabilization. The 21 proline residues were replaced with rather hydrophilic residues such as Glu, Lys or Thr at equivalent sites in less thermostable enzyme [40,43].

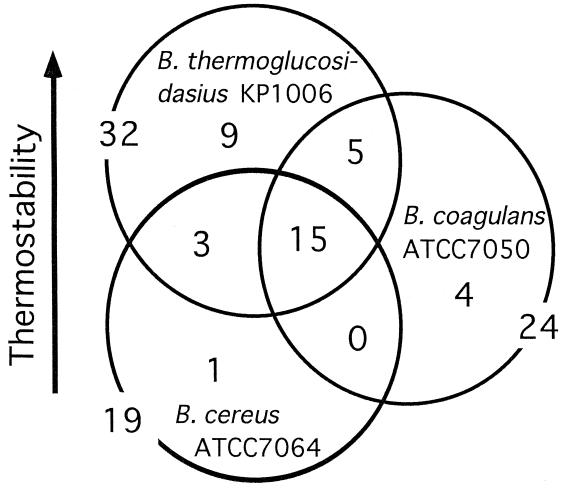

Fig. 5. Classification of common and uncommon proline residues of the three oligo-1,6-glucosidases. The vertical direction shows the increase in thermostability in an arbitrary unit. The three circles represent the *B. thermoglucosidasius* KP1006, *B. coagulans* ATCC7050 and *B. cereus* ATCC7064 oligo-1,6-glucosidases. The numbers shown indicate the proline residues in individual positions.

It was necessary to know whether other local spots or interactions in the oligo-1,6-glucosidases might dominate their thermostability. For example, two limited regions in the amino acid sequence of the α-amylase from *Bacillus licheniformis* were shown to be responsible for thermostability [44]. Since these regions play a crucial role in holding the structure with a Ca^{2+} ion, the motifs in the α-amylase were introduced into another protein to elevate the protein thermostability [45]. Oligo-1,6-glucosidases show significant similarities with such glucan hydrolyzing enzymes in structure [39,46]. So the existence of such local spots or interactions was examined by investigating the thermostability of the chimeric oligo-1,6-glucosidases [47]. As stated above, the genes for the *B. cereus* oligo-1,6-glucosidase and the *B. thermoglucosidasius* oligo-1,6-glucosidase were fused by the use of the commonly and equivalently existing restriction site *Dra*I [39,41,42]. An N-terminal seg-

ment consisting of 170 amino acid residues of the *B. cereus* enzyme or the *B. thermoglucosidasius* enzyme was ligated to the remainder of the other to generate gene fusions. The two types of chimeric enzymes thus constructed were produced in *E. coli* MV1184, purified to homogeneity and compared with the *B. cereus* enzyme and the *B. thermoglucosidasius* enzyme in terms of thermostability. The thermostability of the chimeric enzymes shifted from those of the *B. cereus* and *B. thermoglucosidasius* enzymes, such that there was a correlation between an increase in proline content and a rise in thermostability. Therefore, we concluded that there are no significant regions or sites in oligo-1,6-glucosidases that dominate thermostability.

4. X-ray crystallographic analysis of oligo-1,6-glucosidase from *Bacillus cereus* ATCC7064

From the classification of proline residues in the three oligo-1,6-glucosidases, our major concern was elucidating the structures of the 21 proline substituted-sites which are expected to be critical to protein thermostabilization. We needed a high resolution structure of the *B. cereus* oligo-1,6-glucosidase to help explain the significance of those residues for thermostability. One of the requirements for the work was the ability to produce pure samples of the protein in quantities large enough for us to obtain crystals suitable for X-ray crystallography.

A simple procedure to purify more than 100 mg of the *B. cereus* oligo-1,6-glucosidase was established by the use of an overproducing clone [48]. The purified enzyme was crystallized and grown at 25°C in two weeks by using a hanging drop vapor diffusion method with 53% saturated ammonium sulfate. The crystals, $1.0 \times 0.3 \times 0.3$ mm in size, have the shape of hexagonal bipyramids and belong to the space group $P\ 6_2$ with lattice constants of $a = 106.1$ Å, $c = 120.0$ Å and $\gamma = 120°$ [48]. The crystal structure was first determined by the X-ray diffraction method at 3.0 Å resolution [49] and then refined to 2.0 Å resolution by use of the new 2.0 Å data set [46].

The molecular structure of the *B. cereus* oligo-1,6-glucosidase consists of three domains: the N-terminal domain (residue 1 to 104 and 175 to 480), the subdomain (residues 105 to 174) and the C-terminal domain (residues 481 to 558) (Fig. 6). The overall chain folding is similar to those of *Aspergillus oryzae* α-amylase [50], pig pancreas α-amylase [51,52] and *Bacillus licheniformis* α-amylase [53] that are common to starch hydrolases and related enzymes [54]. The N-terminal domain takes a $(\beta/\alpha)_8$-barrel, the so called TIM-barrel structure [55]. The topological eight-fold symmetry in the N-terminal domain is disturbed by the addition of the subdomain and additional five α-helices (Nα6′, Nα7′, Nα8′, Nα8″, Nα8‴). The subdomain takes a loop-rich structure containing an α-helix (Sα1) and a three stranded antiparallel β-sheet (Sβ1, Sβ2, Sβ3). The subdomain and the two α-helices (Nα8′, Nα8″) in the N-domain make a much deeper cleft than those of α-amylase. Three catalytic residues (Asp199, Glu255, Asp329), deduced from the primary structure comparison of the oligo-1,6-glucosidase with α-amylases [39,56], are located at the bottom of the cleft. The mutation of the three catalytic residues of the oligo-1,6-glucosidase to asparagine or glutamine deprived the protein of its enzyme function [57]. In addition, His103 and His 328, predicted to be part of the substrate binding site, are located around the active site. The rigid specificity of oligo-1,6-glucosidases toward α-1,6-glucosidic bonds is distinct from those of α-amylases, cyclodextrin glucanotransferase and other α-glucosidases which all act on α-1,4-glucosidic bonds. It would be intriguing to clarify the determinants of substrate specificity from the point of view of potentially controlling substrate specificity by protein engineering. On the other hand, the C-terminal domain is located on the opposite side relative to the active site cleft. The C-terminal domain has a β-barrel structure of

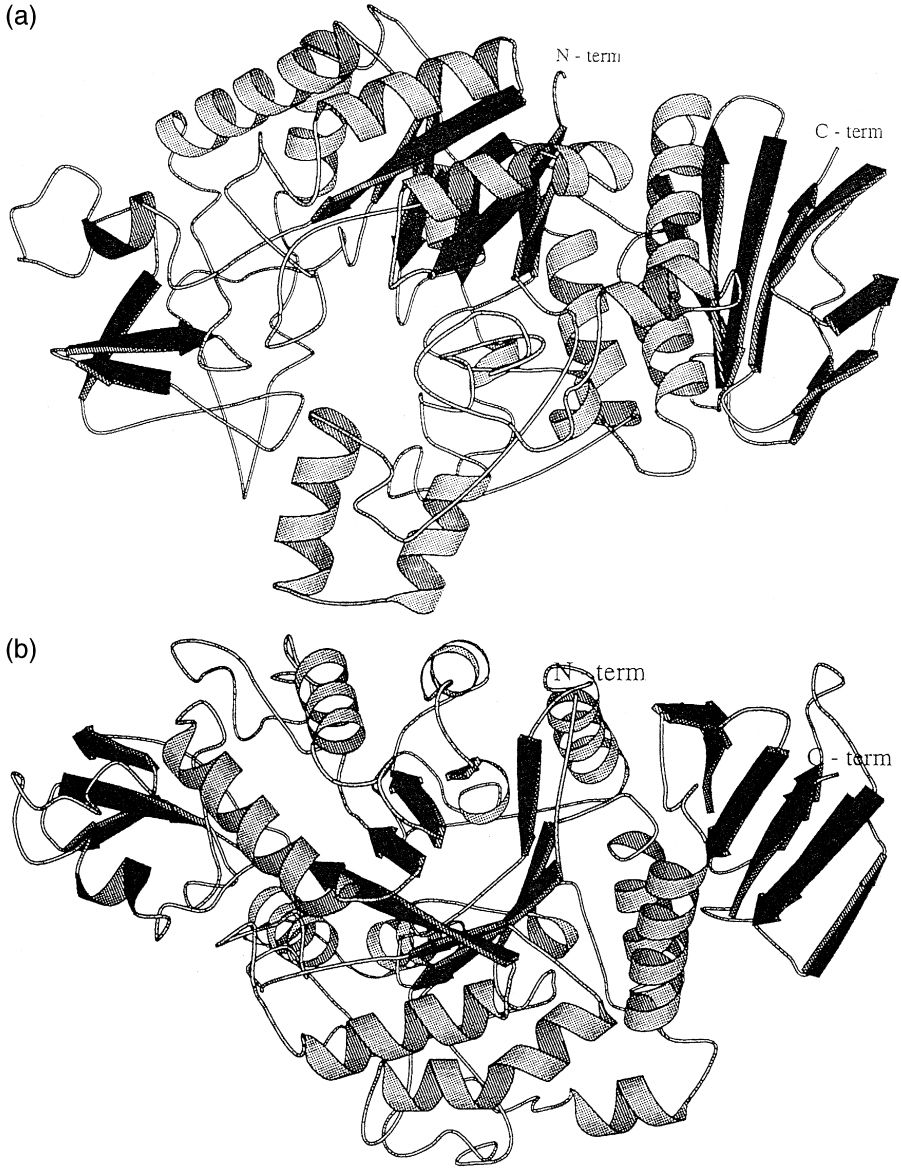

Fig. 6. Ribbon model of *B. cereus* oligo-1,6-glucosidase. (a) The view perpendicular to the axis of the $(\beta/\alpha)_8$-barrel; (b) the view along the axis of the $(\beta/\alpha)_8$-barrel.

eight antiparallel β-strands folded in the double Greek key motif with distortion of the isolated sixth β-strand Cβ6.

Several characteristics of the 21 sites are revealed from the refined structure of the *B. cereus* oligo-1,6-glucosidase. (i) Of the 21 residues, 7 residues occur in β-turns, 4 residues in α-helices and 9 residues in loops. Only one residue, Cys515, occurs in a β-strand. This distribution of the residues agrees well with the statistical data for proline residues in the previous reports by Thornton et al. [14,22]. (ii) In addition, the residues in β-turns are all present at the second positions, and the residues in α-helices are found at the N1 positions of the first helical turns. (iii) Most residues tend to

have large solvent accessibility. This tendency is prominent in the residue at β-turns and α-helices because the structures often occur at the surface of proteins. (iv) The conformations of the residues are within the regions of two favorable clusters for proline residues with the exceptions of three residues (Lys132, Glu216, Lys558) in loops and one residue (Cys515) in the β-strand (see Section 2). These structural characteristics for the 21 sites have given much information for cumulative thermostabilization of the *B. cereus* enzyme by site-directed mutagenesis.

5. Cumulative thermostabilization of oligo-1,6-glucosidase by proline substitutions

The fall in thermostability of a thermostable counterpart by proline-removing substitution is not direct evidence for the enhancement in thermostability by proline residues. The inference that the thermostability of a thermolabile counterpart is increased by proline-introducing substitutions should be tested directly. Thus, we have cumulatively introduced proline residues by site-directed mutagenesis on the oligo-1,6-glucosidase gene from *B. cereus* ATCC7064 [58].

Nine substitution sites were randomly selected from the 21 sites presumed to be critical to protein thermostabilization as in the previous chapters (Fig. 7). These sites were equally divided into three secondary structures; flexible loop (Glu216, Glu270, Glu378), β-turn (Lys121, Glu208, Glu290) and α-helix (Asn109, Glu175, Ile403). Mutations were cumulatively undertaken using the Kunkel method [59] in this order, at the positions Lys121 → Glu175 → Glu290 → Glu208 → Glu270 → Glu378 → Thr261 → Glu216 → Asn109. The resulting mutant enzymes were individually named as Mut-N where N was the number of proline residues

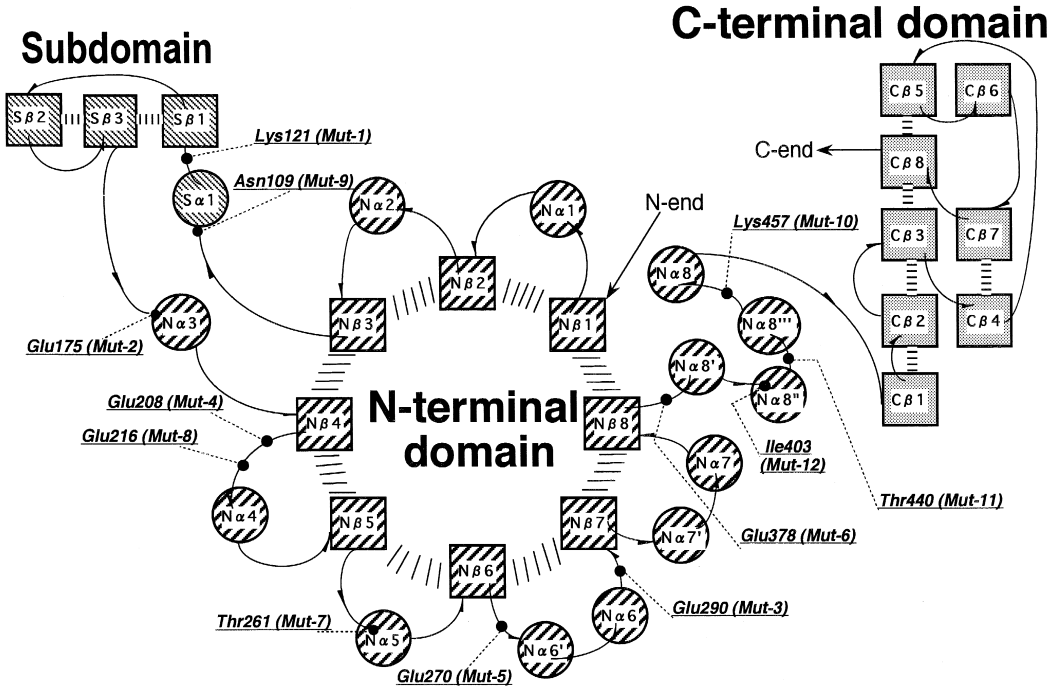

Fig. 7. Topology of secondary structure elements of *B. cereus* oligo-1,6-glucosidase. α-Helices and β-strands are shown by circles and rectangles, respectively. Positions of critical sites for proline substitutions are indicated with black dots.

introduced by site-directed mutagenesis. Furthermore, three additional residues (Lys457 (β-turn), Thr440 (loop) and Ile403 (α-helix)) of the mutant enzyme Mut-9 were replaced with proline residues in the same manner and then named as Mut-10, Mut-11 and Mut-12, respectively (Fig. 7). The last mutant enzyme Mut-12 contained 12 more proline residues than the wild-type enzyme.

It is an absolute requirement that in the attempts aimed at accomplishing protein thermostabilization by site-directed mutagenesis, mutations must not impair the functions of a target protein. The specific activities of the mutant oligo-1,6-glucosidases after purification were comparable to that of the wild-type eyzyme [39]. The retention of specific activity by 12 mutant enzymes suggests that in all cases the proline substitutions were smoothly incorporated into the enzyme structure without any drastic change in function. Since most of the 21 sites have the well-fitted conformations for proline residues, the mutations at those sites presumably do not affect the catalytic conformation of the enzyme.

The effect of proline substitutions on thermostability of the *B. cereus* oligo-1,6-glucosidase was investigated by the comparison in ΔT_m (ΔT_m, the difference between these T_m values of the *B. cereus* enzyme and of its mutants) (Fig. 8). As a whole, the proline substitutions incrementally improved the thermostability of mutant enzymes; in other words, the common effect of the cumulative proline introduction by site-directed mutagenesis was additive on thermostabilization. The overall enhancement in thermostability achieved was 5.6°C. Moreover, a striking result was that the extent of thermostabilization was dependent on the secondary structure. In contrast to the modest effect of proline introductions on the positions in flexible loops (Glu216, Glu270, Glu378, Thr440), the increase in thermostability was most remarkable when proline residues were introduced at second positions of β-turns (Lys121, Glu208, Glu290, Lys457) and at N1

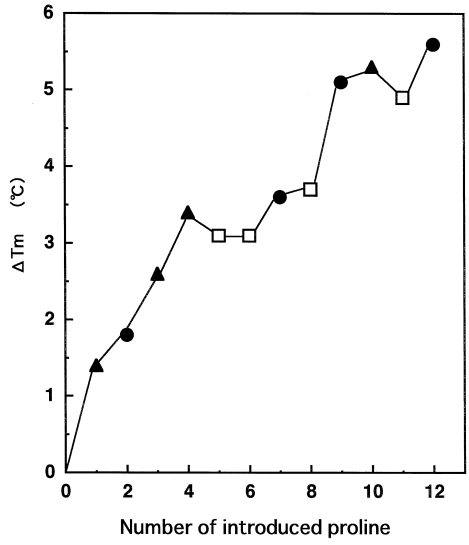

Fig. 8. Additive thermostabilization by cumulative proline introductions into the *B. cereus* oligo-1,6-glucosidase. The number of proline residues introduced into each mutant corresponds to the mutant number *N* in Mut-N. ▲, ● and □ indicate second positions of β-turns, N1 positions of α-helices and flexible loops, respectively, at which the last proline residues were introduced into the mutants.

positions of α-helices (Asn109, Glu175, Thr261, Ile403) [39]. The increase in ΔT_m caused from one introduced proline residue steadily occurred within the range of 0.8–1.4°C for second positions of β-turns and of 0.4–1.4°C for N1 positions of α-helices, respectively. The differences of enhancement in thermostability could be explained by pointing out how each substitution site has the advantages for proline introduction in secondary structure, conformation, accessible solvent accessibility, hydrophobicity and preceding residues. Structural and statistical analyses indicate a preference for proline residues at the second positions of β-turns and N1 positions of α-helices. The conformations of those residues at second positions of β-turns and N1 positions of α-helices are within the region adapted for proline residues and thus accommodate their backbones to the preferred conformation for proline residues without any conformational strain. Those residues at the second positions of β-turns and N1 positions of α-helices

are exposed to the solvent with a larger surface area. Since the residues replaced with proline are hydrophilic amino acids such as Glu, Lys and Thr, the hydrophobic environment should be enhanced in the limited region [43]. In addition, the residues preceding the 21 critical sites in the *B. cereus* enzyme are mostly conserved in more thermostable counterparts together with the proline residues [40,46]. Therefore, the cumulative proline substitutions could minimize the change of mutant enzymes in structure and function. Second positions of β-turns and N1 positions of α-helices have more prerequisites than sites in other secondary structures.

We have demonstrated that the increase in thermostability is essentially due to the entropy decrease of backbone unfolding caused by proline substitution on the basis of the proline rule [13,58]. The entropic effect can presumably work out most efficiently at second positions of β-turns and N1 positions of α-helices. Proline residues in flexible loops would give no effect on protein thermostability in the case of the oligo-1,6-glucosidase. Flexible loops have more interactions with water molecules on the protein surface and thus show rather high mobility. Although much attention has been paid to the stabilization of loops [2], the necessity of proline residues at flexible loops seems to vary depending on the structure and the environment around substitution sites.

The finding that proline residues introduced at the critical site in β-turns and α-helices can contribute most efficiently to the enhancement of protein thermostability is of significance from an evolutionary point of view. We demonstrated that the proline substitution in oligo-1,6-glucosidases supports the neutral theory of evolution, and that the proline substitution is an important factor in their selection by thermostability [60–62]. From the results of thermostabilization by cumulative proline substitutions, it could be predicted that nature introduces proline residues to the most reasonable positions of β-turns and/or of α-helices as one of the strategies for protein thermostabilization. Therefore, proline residues in β-turns and α-helices have more significance than those in other elements.

6. Thermostabilization of other proteins by proline substitutions

The thermostabilization of T4 lysozyme by proline substitution was reported by Matthews et al. [13]. The lysozyme mutant Ala82 → Pro showed the elevation of T_m by 0.7°C. Although the significance of prolines in protein thermostability and protein engineering was already demonstrated [38], this was the first report proving the thermostabilization by site-directed mutagenesis for proline introduction with the experimental evidence. In the paper, the enhanced thermostability was attributed to the decrease of the configurational entropy of unfolding of the polypeptide backbone by proline introduction and the resulting increase of free energy of unfolding, considering the consistent result by Gly → Ala substitution in the same protein. Thereafter, this entropic stabilization by proline substitutions has been applied to the stabilization of other proteins, hen egg lysozyme [63], human lysozyme [64], *E. coli* ribonuclease HI [65,66], *Bacillus stearothermophilus* neutral protease [67], *Bacillus* sp. protease [68] and murine immunoglobulin light-chain [69].

Recently, additional interpretations have been presented for protein thermostabilization. Vieille and Zeikus demonstrated that proline residues located in constrained loops or turns often strengthen the stabilizing interactions between the two adjacent thermostabilizing core elements, based on the idea that protein rigidity induced by the compactness of loops and turns is significant for protein thermostability [2]. This is in good agreement with the fact that thermostable enzymes from thermophilic microorganisms are often shorter than those in their counterparts from mesophiles [70,71]. Ohage et al. developed the idea that structural motifs with predominantly local interactions can serve as templates with which patterns of sequence pref-

erences can be extracted from the database of protein structures. This idea was derived from the finding that mutants with increased β-turn propensities due to proline substitution display increased folding cooperativities [69].

Protein thermostabilization by proline substitutions has the advantage that it can be applied at different sites in a protein. However, the selection of appropriate sites for proline introduction is very important. Some sites might not allow the proline-specific conformation as shown in the previous chapters, or proline introduction might disrupt hydrogen bonds or hydrophobic interactions necessary to maintain the thermostability and structure. If a proline residue is imprudently mis-substituted for the residues in such sites, it may destroy both secondary and tertiary structures as well as protein function and thermostability and, furthermore, result in undesirable kinks in the backbone [72,73]. To avoid the disadvantages and to utilize the advantages for thermostabilization which result from proline introduction, it is advisable to select the substituting sites with high proline preference based on the statistical database or structural motifs [69]. As shown in the case of the oligo-1,6-glucosidase, the most effective sites are the second positions of β-turns and the N1 positions of α-helices [58]. Residue preference is a valuable clue in engineering a protein for thermostabilization. In trials, it is recommendable to exclude the possibility that there is any steric hindrance in the substitution site by referring to knowledge of the three dimensional structure.

Another important matter for protein thermostabilization by site-directed mutagenesis is whether the stability increases must be additive [74,75]. Since the effect of individual proline substitutions on protein thermostability is expected to be small based upon the results reported so far, it is necessary that a number of the point mutations causing replacements with proline cumulatively increase the protein thermostability. The cumulative effects of a few replacements with various amino acid residues were proved in T4 lysozyme and alkaline protease [76,77]. Thereafter, through the utilization of the information on promising single mutations for thermostabilization, the combination of more replacements in a protein distinctively yielded an overall increase of thermostability [58,67,78–80]. This feature suggests that the additivity is derived from the sum of steady thermostabilization in many local regions of a protein. The cumulative mutations and their additive thermostabilization should take place during the protein evolution. Therefore, the effects of proline substitutions depend in part on how far each region around the substitution sites is separated and on how independent the interactions between the regions are. In case of the oligo-1,6-glucosidase, the sites for cumulative mutations are spread over the surface of the protein molecule in light of its crystallographic structure [46]. It is noted that the substitution sites must be apart from the active site to retain the protein functions. Thus, we must keep in mind that cumulative substitutions often cause the drastic loss of enzyme activities [78] though it is also possible for substitutions to improve the enzyme functions by chance [80].

There should be many appropriate sites for proline substitutions to accomplish thermostabilization in a protein. The substitution sites are more usual than the sites to construct a disulfide bridge between two cysteine residues for protein thermostabilization [81] because of the general abundance of α-helices and β-turns in proteins. Therefore, promising sites for proline substitutions should occur in a protein much more frequently than, for example, the sites for cysteine substitutions. The combination of multiple mutations following such promising strategies as the proline rule has been found to be a significant method improving protein thermostability.

Acknowledgements

The crystallographic work of the oligo-1,6-glucosidase from *B. cereus* ATCC7064 was

done in collaboration with Dr. Yukiteru Katsube, Institute for Protein Research, Osaka University, and Dr. Yasuo Hata, Institute for Chemical Research, Kyoto University. We are deeply indebted to Greg York for helpful comments and English corrections.

References

[1] M.J. Zoller, M. Smith, Methods Enzymol. 100 (1983) 468.
[2] C. Vieille, G. Zeikus, Trends Biotechnol. 14 (1996) 183.
[3] A. Shaw, R. Bott, Curr. Opin. Struct. Biol. 6 (1996) 546.
[4] M. Danson, D. Hough, R. Russel, G. Taylor, L. Pearl, Protein Eng. 9 (1996) 629.
[5] R. Jaenicke, FASEB J. 10 (1996) 84.
[6] R.M. Daniel, M. Dines, H. Petach, Biochem. J. 317 (1996) 1.
[7] C.N. Pace, J. Mol. Biol. 226 (1992) 226.
[8] G. Vriend, H.J.C. Berendsen, J.R. van der Zee, B. van den Burg, G. Venema, V.G.H. Eijsink, Protein Eng. 4 (1991) 941.
[9] H. Nicholson, B.J. Becktel, B.W. Matthews, Nature 336 (1988) 651.
[10] N.J. Mrabet, A. Van der Broeck, I. Van den Brande, P. Stansssens, Y. Laroche, A.M. Lameir, G. Matthijissens, J. Jenkins, M. Chiaclmi, H. Tilbeurgh, F. Rey, J. Jania, W.J. Quax, I. Lasters, M. De Maeyer, S.J. Wodak, Biochemistry 31 (1992) 2239.
[11] S. Kawamura, Y. Kakuta, I. Tanaka, K. Hikichi, S. Kuhara, N. Yamazaki, M. Kimura, Biochemistry 35 (1996) 1195.
[12] D.B. Wigley, A.R. Clarke, C.R. Dunn, D.A. Barstow, T. Atkinson, W.N. Chia, H. Muirhead, J.J. Holbrook, Biochim. Biophys. Acta 916 (1987) 145.
[13] B.W. Matthews, H. Nicholson, W.J. Becktel, Proc. Natl. Acad. Sci. USA 84 (1987) 6663.
[14] M.W. MacArthur, J.M. Thornton, J. Mol. Biol. 218 (1991) 397.
[15] P. Schimmel, P.J. Flory, J. Mol. Biol. 34 (1968) 105.
[16] F.C. Bernstein, T.F. Koetzle, G.J. B Williams, E.F. Meyer, M.D. Brice, J.R. Rogers, O. Kennard, T. Shimanouchi, M. Tatsumi, J. Mol. Biol. 112 (1977) 535.
[17] G.N. Ramachandran, V. Sasisekharan, Adv. Protein Chem. 23 (1968) 282.
[18] J.H. Hurley, D.A. Mason, B.W. Matthews, Biopolymers 32 (1992) 1443.
[19] P. Argos, J. Palau, Int. J. Pept. Protein Res. 19 (1982) 380.
[20] J.S. Richardson, Adv. Protein Chem. 34 (1981) 167.
[21] P.Y. Chou, G.D. Fasman, Adv. Enzymol. 47 (1978) 45.
[22] E.G. Hutchinson, J.M. Thornton, Protein Sci. 3 (1994) 2207.
[23] K. O'Neil, W. DeGrado, Science 250 (1990) 646.
[24] S. Dasgupta, J. Bell, Int. J. Pept. Protein Res. 41 (1993) 499.
[25] J.S. Richardson, D.C. Richardson, Science 240 (1988) 1648.
[26] R.H. Yun, A. Anderson, J. Hermans, Proteins 10 (1991) 219.
[27] G. von Heijne, J. Mol. Biol. 218 (1991) 499.
[28] K. Saksela, G. Cheng, D. Baltimore, EMBO J. 14 (1995) 484.
[29] S. Hong, K. Ryu, M. Oh, I. Ji, T.H. Ji, J. Biol. Chem. 272 (1997) 4166.
[30] J.M. O'Sullivan, R.D. Cannon, P.A. Sullivan, H.F. Jenkinson, Microbiology 143 (1997) 1341.
[31] A. Galat, Eur. J. Biochem. 216 (1993) 686.
[32] Y. Suzuki, R. Aoki, H. Hayashi, Biochim. Biophys. Acta 704 (1982) 476.
[33] Y. Suzuki, Y. Tomura, Eur. J. Biochem. 158 (1986) 77.
[34] Y. Suzuki, K. Oishi, H. Nakano, T. Nagayama, Appl. Microbiol. Biotechnol. 26 (1987) 546.
[35] Y. Suzuki, T. Yuki, T. Kishigami, S. Abe, Biochim. Biophys. Acta 445 (1976) 386.
[36] Y. Suzuki, N. Sugita, T. Kishimoto, Starch/Stärke 49 (1997) 148.
[37] Y. Suzuki, H. Fujii, H. Uemura, M. Suzuki, Starch/Stärke 39 (1987) 17.
[38] Y. Suzuki, Proc. Jpn. Acad. Ser. B Phys. Biol. Sci. 65 (1989) 146.
[39] K. Watanabe, K. Kitamura, H. Iha, Y. Suzuki, Eur. J. Biochem. 192 (1990) 609.
[40] K. Watanabe, K. Kitamura, Y. Suzuki, Appl. Environ. Microbiol. 62 (1996) 2066.
[41] K. Watanabe, H. Iha, A. Ohashi, Y. Suzuki, J. Bacteriol. 171 (1989) 1219.
[42] K. Watanabe, K. Chishiro, K. Kitamura, Y. Suzuki, J. Biol. Chem. 266 (1991) 24287.
[43] J.L. Fauchere, V. Pliska, Eur. J. Med Chem. Chim. Ther. 18 (1983) 369.
[44] Y. Suzuki, N. Ito, T. Yuuki, H. Yamagata, S. Udaka, J. Biol. Chem. 264 (1989) 18933.
[45] R. Kuroki, Y. Taniyama, C. Seko, H. Nakamura, K. Kikuchi, M. Ikehara, Proc. Natl. Acad. Sci. USA 86 (1989) 6903.
[46] K. Watanabe, Y. Hata, H. Kizaki, Y. Katsube, Y. Suzuki, J. Mol. Biol. 269 (1997) 142.
[47] K. Watanabe, T. Saito, Y. Suzuki, Abstr. of the Annual Meeting of Japanese Agricultural Chemical Society, 1990, 574 pp.
[48] K. Watanabe, K. Kitamura, Y. Hata, Y. Katsube, Y. Suzuki, FEBS Lett. 290 (1991) 221.
[49] H. Kizaki, Y. Hata, K. Watanabe, Y. Katsube, Y. Suzuki, J. Biochem. 113 (1993) 646.
[50] Y. Matsuura, M. Kusunoki, W. Harada, M. Kakudo, J. Biochem. 95 (1984) 697.
[51] S.B. Larson, A. Greenwood, D. Cascilo, J. Day, A. MacPherson, J. Mol. Biol. 235 (1994) 1560.
[52] M. Qian, R. Haser, F. Payan, J. Mol. Biol. 231 (1994) 785.
[53] M. Machius, G. Wiegand, R. Huber, J. Mol. Biol. 246 (1995) 545.
[54] H.E. Jespersen, E.A. MacGregor, M.R. Sierks, B. Svensson, Biochem. J. 280 (1991) 51.
[55] D.W. Banner, A.C. Bloomer, G.A. Ptsko, D.C. Phillips, C.I. Pogson, I.A. Wilson, P.H. Corran, A.J. Furth, J.D. Milman, R.E. Offord, J.D. Priddle, S.G. Waley, Nature 255 (1975) 609.
[56] B. Svensson, Plant Mol. Biol. 25 (1994) 141.
[57] K. Miyake, K. Kishimoto, K. Watanabe, Y. Suzuki, Abstr. of

the Annual Meeting of Japanese Agricultural Chemical Society, 1994, 243 pp.
[58] K. Watanabe, T. Masuda, H. Ohashi, H. Mihara, Y. Suzuki, Eur. J. Biol. 226 (1994) 277.
[59] T.A. Kunkel, J. Robert, R. Zakour, Methods Enzymol. 154 (1987) 367.
[60] M. Kimura, The Neutral Theory of Molecular Evolution, Cambridge Press, Cambridge, 1983.
[61] B.A. Malcolm, K.P. Wilson, B.W. Matthews, J.F. Kirsch, A.C. Wilson, Nature 345 (1990) 86.
[62] G.J. Pierak, D.S. Auld, J.R. Beasley, S.F. Betz, D.S. Cohen, D.F. Doyle, S.A. Finger, Z.L. Fredericks, S. Hilgen-Willis, A.J. Saunders, S.K. Trojak, Biochemistry 34 (1995) 3268.
[63] T. Ueda, T. Tamura, Y. Maeda, Y. Hashimoto, T. Miki, H. Yamada, T. Imoto, Protein Eng. 6 (1993) 183.
[64] T. Herning, K. Yutani, K. Inaka, R. Kuroki, M. Matsushima, M. Kikuchi, Biochemistry 31 (1992) 7077.
[65] K. Ishikawa, S. Kanaya, K. Morikawa, H. Nakamura, Protein Eng. 6 (1993) 85.
[66] S. Kimura, H. Nakamura, T. Hashimoto, M. Oobatake, S. Kanaya, J. Biol. Chem. 267 (1992) 21535.
[67] F.G. Hardy, G. Vriend, O.R. Veltman, B. van der Vinne, G. Venema, V.G.H. Eijsink, FEBS Lett. 317 (1993) 898.
[68] A. Masui, N. Fujiwara, T. Imanaka, Appl. Environ. Microbiol. 60 (1994) 3579.
[69] E.C. Ohage, W. Graml, M.M. Walter, S. Steinbacher, B. Steipe, Protein Sci. 6 (1997) 233.
[70] J.M. Muir, R.J.M. Russel, D.W. Hough, M.J. Danson, Protein Eng. 8 (1995) 583.
[71] G.L. Wallen, G.A. Petsko, Protein Eng. 8 (1995) 905.
[72] T. Imanaka, M. Nakae, T. Ohta, M. Takagi, J. Bacteriol. 174 (1992) 1423.
[73] G.X. Luo, P.M. Horowits, J. Biol. Chem. 268 (1993) 10246.
[74] J.A. Wells, Biochemistry 29 (1990) 8509.
[75] B.W. Matthews, Biochemistry 26 (1987) 6885.
[76] S. Dao-Pin, W.A. Baase, B.W. Matthews, Proteins 7 (1990) 198.
[77] T. Imanaka, M. Shibasaki, M. Takagi, Nature 324 (1986) 695.
[78] X.J. Zhang, W.A. Baase, B.K. Shoichet, K.P. Wilson, B.W. Matthews, Protein Eng. 8 (1995) 1017.
[79] P. Shih, J.F. Kirsch, Protein Sci. 4 (1995) 2063.
[80] S.L. Strausberg, P.A. Alexander, D.T. Gallagher, G.L. Gilliland, B.L. Barnett, P.N. Bryan, Biotechnology 13 (1995) 669.
[81] J. Perry, R. Wetzel, Science 226 (1984) 545.