OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Structural studies on thermostable proteins by using a novel X-ray diffraction measurement system
Author(s)	Okazaki, Nobuo
Citation	大阪大学, 2013, 博士論文
Version Type	VoR
URL	https://hdl.handle.net/11094/27532
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

工 8 9762

博士学位論文

Structural studies on thermostable proteins by using a novel X-ray diffraction measurement system

（新たなX線回折測定システムを利用した耐熱性タンパク質の構造学的研究）

岡 﨑 伸 生

2012年12月

大阪大学大学院工学研究科

763

博士学位論文

Structural studies on thermostable proteins by using a novel X-ray diffraction measurement system

（新たなX線回折測定システムを利用した耐熱性タンパク質の構造学的研究）

岡 﨑 伸 生

2012年12月

大阪大学大学院工学研究科

Preface

The studies presented here contain works at three institutes. At the period from 2003 to 2005, the studies have been carried out at High-throughput Factory, RIKEN Harima Institute (Hyogo). Those studies were supported RIKEN Structural Genomics/Proteomics Initiative and carried out as a part of the Protein 3000 project. At the period from 2005 to 2008, the studies have been carried out at Structural Biology Group, Japan Synchrotron Radiation Research Institute (JASRI; Hyogo). Those studies were achieved at SPring-8 Structural Biology Beamlines, which are BL26B1/B2 (RIKEN Structural Genomics I/II), BL38B1 (Structural Biology III), and BL41XU (Structural Biology I). At the period from 2008 to 2012, the studies were carried out at Molecular Structural Biology Group, Japan Atomic Energy Agency (JAEA; Ibaraki).

The object of this thesis contains X-ray structural analysis and discussion of proteins from thermophiles to reveal the properties of thermostability, reaction mechanism, and structure–function relationship.

December, 2012
Nobuo Okazaki

Table of Contents

Preface i

General introduction 1

Contents of this thesis 9

Chapter I 9

Chapter II 10

Chapter III 11

Chapter I Glycosyltrehalose synthase from *Sulfolobus shibatae* DSM5389 13

I-1. Introduction 13

I-2. Results 15

I-3. Discussion 22

I-4. Materials and methods 30

Chapter II Glycosyltrehalose trehalohydrolase from *Sulfolobus solfataricus* KM1 33

II-1. Introduction 33

II-2. Results 36

II-3. Discussion 48

II-4. Materials and methods 54

Chapter III Structure determination of other proteins from thermophiles for application to industrial use 59

III-1. Introduction 59

III-2. Results 61

III-3. Discussion 69

III-4. Materials and methods 78

General discussion and conclusion 87

Appendix 91

Appendix A. Mail-in data collection at SPring-8 protein crystallography beamlines . 91

Appendix B. Further structural investigation by using neutron crystallography 101

References 111

List of publications 131

List of supplementary publications 133

List of supplementary publications in Japanese 137

Acknowledgements 139

General introduction

Importance of protein structural information

Diverse biological functions are mediated by proteins, which are biochemical compounds composed of polypeptides. A polypeptide, in turn, is a polymer made from combinations of 20 main amino acids, the exact composition of which is genetically encoded. In the polymer chain, interactions such as hydrogen bonds and van der Waals interactions are responsible for the three-dimensional structure of proteins. Generally, this structure confers one or more functions on the protein, such as transport of other molecules or catalysis of biochemical reactions. Proteins that perform catalytic reactions are also referred to as enzymes. In recent years, the diverse and sophisticated functionalities of proteins have been utilized in industrial processes. Examples include Mitsubishi Rayon Co., Ltd. (Tokyo, Japan), which established processes related to the bio-method acrylamide process, and Hayashibara Co., Ltd. (Okayama, Japan), which produced trehalose (trade name: *TREHA*) enzymatically. In addition, knowledge of three-dimensional structural analysis is also advantageous for drug discovery.

Protein engineering is a rapidly developing field. It is based on the understanding of protein folding and recognition mechanism, which in turn are derived from three-dimensional structure. Protein engineering is particularly important for bio-industrial processes. For example, glycosidases (glycoside hydrolases) are industrially important enzymes that catalyze the hydrolysis of a glycosidic linkage to release smaller sugars. A well-known example is the α-amylases, which are used in the industrial trehalose production process. Glycosidases are also used for bioethanol production, which proceeds via the following steps: microbial (yeast) fermentation of sugars, distillation, dehydration, and denaturation. Prior to fermentation, some crops

require saccharification or hydrolysis of carbohydrates such as cellulose and starch into sugars; this process is carried out by glycosidases [1]. Thus, structure-based improvements to such activities of enzymes will facilitate the development of future energy resources and will find utility in green chemistry applications.

Proteins from thermophiles

Various organisms live in a range of extreme environments such as high temperature, high salt concentration, high pressure, and radioactivity. Proteins isolated from thermophiles, which live in high-temperature environments, have heat-resistant properties. The stabilization of thermophilic proteins is described with equilibrium theory, kinetics, or both in high temperature. Covalent bonds, electrostatic interaction, hydrogen bonds, van der Waals interaction (Lennard-Jones interaction), and entropy can enhance the stability of those proteins. In the Structural Biological Whole-Cell Project [2] (http://www.thermus.org/), *Thermus thermophilus* HB8 is used as a model organism. Proteins from *T. thermophilus* HB8 are thermostable and suitable for physiochemical studies, including X-ray crystallography. Many structural features are thought to contribute to thermostability, such as enhancing the stability of secondary structures and increasing various types of interactions, including hydrophobic interaction, hydrogen bonds, ion-pairing, oligomerization, and solvent-accessible surface areas (SASAs) [3,4].

Thermostability of proteins is also very important for industrial processes, where efficiency and selectivity are required. Generally, catalytic reactions are accelerated at high temperatures. Thermophilic proteins are not denatured under such conditions and are therefore advantageous in industrial processes. Recently, enzymes have also

been widely used to synthesize industrial biochemicals. Therefore, enhancing protein thermostability is an important aim of protein engineering, as underscored in several studies [5-7]. To find a key of thermostability of protein needs widely targeted structural analysis.

The comprehensive structure analysis

As described above, three-dimensional protein structure information is essential to advance protein engineering for industrial applications, including drug discovery and green chemistry. X-ray protein crystallography is one of the most powerful methods for three-dimensional protein structure determination. It enables us to visualize protein structures at the atomic level and enhances our understanding of protein function. Important knowledge can be obtained from the information at the atomic level. For example, the data can reveal how proteins interact with other molecules, how they undergo conformational changes, and how they act as catalysts in the case of enzymes. By using this structural information, researchers can design novel drugs or engineer enzymes for industrial processes. Among the structural genomics projects that pioneered in Japan is the "Protein 3000" project, supported by the RIKEN Structural Genomics/Proteomics Initiative (RSGI) [8]. The Protein 3000 project ran from 2002 to 2006, which founded by Ministry of Education, Culture, Sports, Science and Technology (MEXT). This major project aims to determine bacterial, mammalian, and plant protein structures by X-ray crystallography and NMR spectroscopy and to perform functional analyses with the target proteins by exhaustive structure analysis. Although there are likely hundred thousand of protein species in *Homo sapiens*, they can be classified based on the presence of structural motifs. The "3000" of the project means a

third of the number of the motifs in total. Those structures are applied as base motifs for protein three-dimensional structure prediction. Another project related to Protein 3000 is the Structural Biological Whole-Cell project. This is promoted by the RIKEN SPring-8 Center, with aims to increase the understanding of fundamental biological phenomena in cells at the atomic level on the basis of the three-dimensional structures of proteins [2]. The project selected an extreme thermophile, *T. thermophilus* HB8, as a model organism because its genome is amenable to genetic manipulation, the bacterium can grow at high temperatures (65–85°C) and its proteins are highly stable and thus suitable for structure–function studies. In these projects, a vast number of protein structures will be determined. Therefore, the development of efficient methods for obtaining information from these studies is critical. For this reason, synchrotron radiation facilities aim to develop more efficient measurement systems for X-ray protein crystallography and to verify the advantage of the system. Improvement of efficiency is primarily required for this approach to improve industrial processes and support green chemistry.

Recent synchrotron radiation facilities for X-ray protein crystallography

For more than two decades, researchers conducted measurements manually, which used synchrotron radiation as an X-ray source. Therefore, to characterize a large number of diverse protein structures, more effective analysis methods are required. Protein 3000 thus requires high-efficiency measurement methods, robotics, and automated procedures, such as those introduced in Photon Factory (PF; KEK, Tsukuba, Japan) and SPring-8 (Hyogo, Japan). An automated sample exchanger robot was developed and installed in the synchrotron facilities. *PAM* (PF Automated

Mounting system) [9] and *SPACE* (SPring-8 Precise Automatic Cryo-sample Exchanger) [10] were also installed to PF and SPring-8, respectively.

At SPring-8, the RIKEN structural genomics beamline I and II (BL26B1 and BL26B2) [11] have been constructed to contribute to structural genomics research, which includes projects such as Protein 3000. Similar projects include Structural Biology I and III (BL41XU and BL38B1); these beamlines are also set up for public research use. Recently, BL32XU (RIKEN targeted protein beamline) [12] has been constructed in the Targeted Protein Research Program founded by MEXT (http://www.tanpaku.org/). BL41XU and BL32XU are undulator beamlines capable of advanced protein structure determination. They both have a high-flux beam and BL32XU in addition is a micro-focus beamline. On the other hand, bending magnet beamlines, which are BL38B1, BL26B1 and BL26B2, can be used to measure with high-throughput efficiency. To achieve high-throughput protein crystallography in SPring-8, each of these two special components has been further developed and combined. One is *SPACE*, which is an automated sample exchanger robot (Fig. 1). It can change up to 100 crystals in a diffractometer. Another is *BSS* [13] which is *Beamline Scheduling Software* for management of measurement schedule (Fig. 2). *BSS* can perform successive data collection by controlling beamline devices, including *SPACE*. In addition to *BSS* and *SPACE*, a novel measurement system *D-Cha* [14],

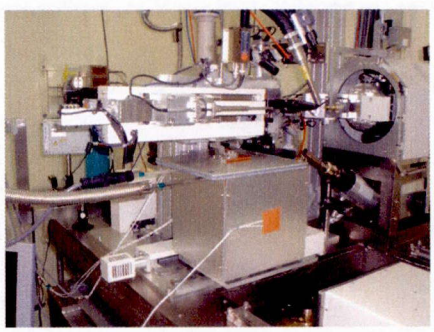


Figure 1. Automated sample exchanger robot SPACE

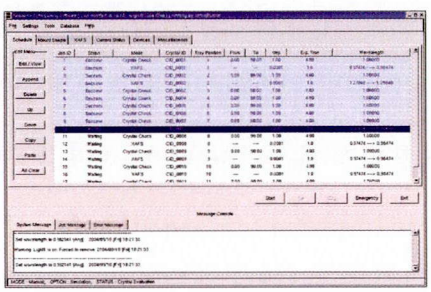


Figure 2. Beamline Scheduling Software

which enables distant users to operate those components, has been available by the author.

Neutron protein crystallography for further investigation of thermostability

For further investigation of enzyme catalytic activity and/or thermostability, structural studies of the positions of hydrogen atoms are essential. Hydrogen atoms widely contribute to the formation of protein structures and stabilization through hydrogen bonds, which play important roles in biological processes. In addition, the information of the hydrogen atoms makes effects on results of molecular dynamics calculation. For these reasons, structural information of hydrogen atoms is important in biochemical and computational studies for investigation on thermostable proteins. There are two methods for the determination of hydrogen atom arrangement, namely, ultra-high resolution X-ray protein crystallography and neutron crystallography. The determination of the positions of hydrogen atoms requires high resolution and high accuracy in X-ray protein crystallography. Since X-ray diffraction depends on the number of electrons in each atom, the observation of hydrogen atoms with this method is challenging. On the other hand, neutron crystallography overcomes these limitations, since it relies on the interaction of neutrons with the nucleus. This permits a direct and highly accurate determination of hydrogen atom distribution. Enzymes represent important targets for many pharmaceutical agents. Therefore, elucidation of the catalytic function and the molecular recognition mechanisms of enzymes provide important information for structure-based drug design. Obviously, to investigate thermostable properties also have wide availability on it. Neutron crystallography can provide accurate information on the locations of hydrogen atoms that are essential

for thermostability, enzymatic function, and molecular recognition mechanisms.

Despite the advantages described as above, the neutron diffraction technique has a number of bottlenecks. One of them is that acquisition of neutron diffraction data is extremely time-consuming. At BIX-3 and BIX-4 in the JRR-3 atomic reactor (Japan Atomic Energy Agency; Ibaraki, Japan), image acquisition requires four to six hours per image, normally. Therefore, several months are required to acquire a dataset for one protein crystal. Another bottleneck is that larger crystals are required for such experiments when compared with X-ray diffraction. Normally, a crystal is required with more than ≥ 2 mm^3 for a crystal volume and with more than 2 Å resolution for diffraction ability by using X-ray beam. Hence, a large crystal requires some trials of crystallization to obtain.

Contents of this thesis

In the introduction, the author described the aims of the studies. The aim of this thesis is to advance the use of enzymes in industrial processes by enhancing their thermostability. A selection of glycosidases and thermostable proteins from thermophiles are chosen as the targets. Especially, glycosidases are likely to contribute to more efficient industrial processes and, in turn, contribute to green chemistry for sustainable activity. In the industrial use of proteins as catalysts, protein thermostability is a very important factor. This thesis consists of cases of protein structure determination by using effective measurement system *D-Cha* (see Appendix A) for thermostable proteins at SPring-8, and neutron crystallography for further investigation on those proteins. In Chapter I and II, glycosidases from thermophiles that involve in trehalose production are described. Chapter III describes the other cases of structure determination and diffraction studies on thermostable proteins. The discussion includes glycosidases and thermostable proteins from thermophiles, which increase our understanding of the importance of thermostability. The contents of the thesis chapters are described below.

Chapter I

The crystal structure of glycosyltrehalose synthase (GTSase) from the hyperthermophilic archaeon *Sulfolobus shibatae* DSM5389 has been determined to 2.3 Å resolution by X-ray crystallography. GTSase converts the glucosidic bond between the last two glucose residues of amylose from an α-1,4 bond to an α-1,1 bond by an intramolecular transglycosylation mechanism. This produces a non-reducing glycosyl trehaloside in

the first step of the biosynthesis of trehalose. The reaction mechanism of GTSase was proposed from model structures of the enzyme–substrate complex (ES complex) and covalent intermediates based on the structural data of 5389-GTSase and other α-amylase enzymes. The author also combined these elements with the previous biochemical knowledge. The structural information helps us understand the cleavage of the α-1,4 glucosidic bond at the reducing end of substrate and subsequent intramolecular conversion of an α-1,4 bond to an α-1,1 bond. To investigate thermostable property, features of solvent-accessible surface area and ion interactions were compared with mesophilic α-amylase.

Chapter II

Glycosyltrehalose trehalohydrolase (GTHase) is an α-amylase that cleaves the α-1,4 bond adjacent to the α-1,1 bond of maltooligosyltrehalose to release trehalose. In order to investigate the catalytic and substrate recognition mechanisms of GTHase, two residues, Asp252 (nucleophile) and Glu283 (general acid/base), located at the catalytic site of GTHase were mutated (Asp252→Ser (D252S), Glu (D252E) and Glu283→Gln (E283Q)). The activity and structure of the enzyme were then investigated. The crystal structure of the E283Q mutant GTHase in complex with the substrate, maltotriosyltrehalose (G3-Tre), was determined to 2.6 Å resolution. The structure with G3-Tre indicated that GTHase has at least five substrate-binding subsites. The complex structure also revealed a scheme for substrate recognition by GTHase. Comparison of solvent-accessible area and ion networks with mesophilic ones were similar trend to GTSase in former chapter.

Chapter III

This chapter describes the other cases of structure determination in thermophiles. The author's first aim is to apply structural features to proteins for industrial use, which will help improving the today's industrial use of ones. For this aim, a part of the comprehensive structure determination was attempted by using X-ray and neutron diffraction experiments. Thus this chapter describes some cases of structure determination by using different two methods. One of methods is X-ray crystallography with *D-Cha* as described in Chapter I and II. The other is neutron crystallography (see Appendix B). The first is structure determination of UPF0150-family protein from *Thermus thermophilus* HB8 (TTHA0281). TTHA0281 is a hypothetical protein from *T. thermophilus* HB8 that belongs to an uncharacterized protein family, UPF0150. The X-ray crystal structure of the protein was determined by a multiple-wavelength anomalous dispersion technique and was refined at 1.9 Å. Another case in X-ray crystallography is crystallization and preliminary X-ray diffraction studies of chitinase from the moderately thermophilic bacterium *Ralstonia* sp. A-471. Chitin, a β-1,4 polymer of *N*-acetyl-D-glucosamine is the second most abundant biopolymer in nature after cellulose. The catalytic domain with part of an interdomain loop (Ra-ChiC 89–252) was crystallized under several different conditions using polyethylene glycol as a precipitant.

Thus, for further investigation of structural features of thermostable enzymes, the determination of structures including hydrogen atoms which involve in the reaction is essential. It is a challenging work to obtain a large crystal which is suitable for neutron diffraction measurement. Therefore, the author attempted that they are test cases with dealing large crystals in neutron crystallography, and conducted crystallization of

thermostable proteins. To obtain a large crystal, the author attempted a macroseeding method for ADP-ribose pyrophosphatase-I from *T. thermophilus* (Tt-ADPRase-I) [15]. Tt-ADPRase-I prevents the intracellular accumulation of ADP-ribose by hydrolyzing it to AMP and ribose 5′-phosphate. To understand the catalytic mechanism of Tt-ADPRase-I, it is necessary to investigate the role of glutamates and metal ions, as well as the coordination of water molecules located at the active site. Neutron and X-ray diffraction experiments were performed at room temperature using the same crystal. The crystal diffracted to 2.1 and 1.5 Å resolution in the neutron and X-ray diffraction experiments, respectively.

As previously mentioned, thermostable proteins are the main target for the comprehensive structural analysis. Such proteins are advantageous for handling and in industrial processes due to their thermostable nature. Therefore, elucidation of thermostable features will widely contribute to the improvement of industrial processes. In the last of this chapter, the thermostable features are discussed by using their structures in this chapter and Chapter I and II.

Chapter I
Glycosyltrehalose synthase from *Sulfolobus shibatae* DSM5389

I-1. Introduction

Trehalose (α-D-glucopyranosyl-1,1-α-D-glucopyranose), a non-reducing α-1,1 linked disaccharide, is an important osmolyte in insects, plants, fungi, and bacteria, and is one of most efficacious agents for the protection of proteins and lipid membranes from freezing, desiccation, high temperatures, and osmotic stress [16–18]. These preservative properties are particularly important for survival of organisms under extreme anhydrobiotic conditions: trehalose interacts directly with membranes to maintain their fluid state during desiccation and prevent liquid crystal → gel phase transitions that result in membrane fusion and permeability. Trehalose also has various useful industrial applications; for example, it has wide-spread utility as a preservative

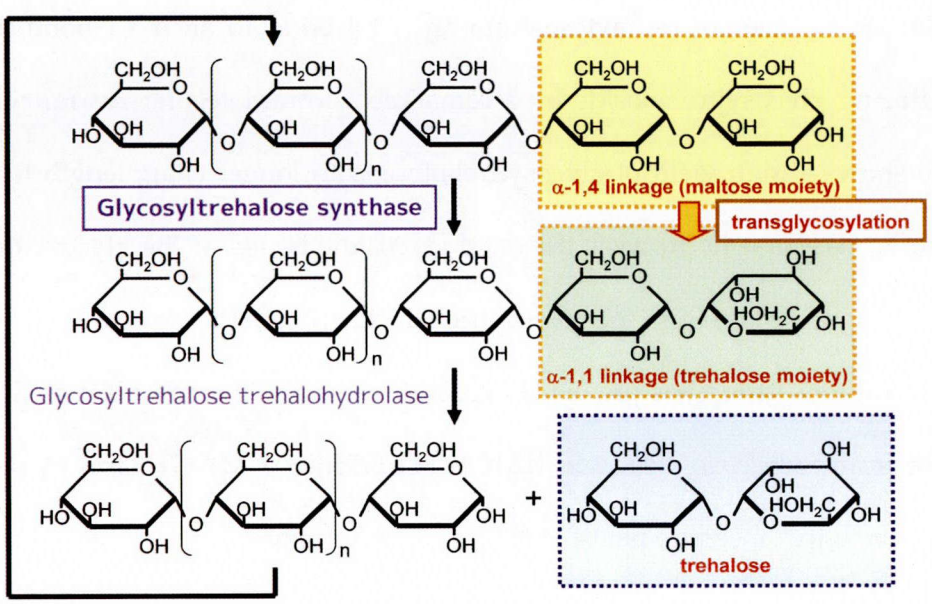


Figure I-1. Scheme of trehalose biosynthesis system.

for cosmetics, medicines, and foods, and exhibits notably low toxicity. However, the high cost of commercialized trehalose (US$700/kg), which is produced by extraction from plants and brewery-spent yeast, had limited industrial application until the early 1990s. The trehalose production by novel biosynthesis from starch brought to the dramatic decrease in trehalose price (US$700 to US$5–6) [19]. This biochemical synthesis of trehalose is accomplished by two coupled enzymes: glycosyltrehalose synthase (GTSase; EC 2.4.1.25; also called maltooligosyltrehalose synthase; MTSase) and glycosyltrehalose trehalohydrolase (GTHase; EC 3.3.1.1; also called maltooligosyltrehalose trehalohydrolase; MTHase), identified from several microorganisms, including *Arthrobacter* [20–22] and *Sulfolobus* [23–27]. These enzymes belong to α-amylase family (α-1,4-glucan 4-glucanohydrolase, EC 3.2.1.1) [28] which are classified into a glycoside hydrolase (GH) family [29] 13 (*CAZy* database [30]; http://www.cazy.org/). GTSase is a glycosyltransferase and GTHase is an α-amylase; an enzymatic pathway for trehalose production is shown in Fig. I-1. GTSase converts the glycosidic bond between the last two glucose residues of α-amylose from an α-1,4 bond to an α-1,1 bond, making a non-reducing glycosyltrehaloside by a remarkable intramolecular rearrangement [20,26]. GTSase shows a high yield of glycosyltrehalose with longer chain-length than maltotetraose [26]. GTHase hydrolyzes the α-1,4 glycosidic bonds of the glycosyltrehaloside and releasing trehalose from the glycosyltrehaloside (Fig. I-1).

The author's group has previously probed the catalytic mechanism of GTSase and GTHase from *Sulfolobus sofataricus* KM1 (KM1-GTSase, KM1-GTHase) by several enzymatic activity measurements [25,26,31], and determined the crystal structure of KM1-GTHase [32]. The structural analysis of the GTSase may also help elucidate details of the intramolecular transglycosylation mechanism. Here the author reports the

crystal structure of GTSase from *Sulfolobus shibatae* DSM5389 (5389-GTSase), which consists of 728 amino acids and is 97% identical to KM1-GTSase. In 5389-GTSase structure, the configuration of common three catalytic residues is highly conserved within α-amylase family structures. Intraglycosylation mechanism by GTSase was proposed from the model structures of enzyme–substrate complex (ES complex) and covalent intermediate based on the structural data of 5389-GTSase and other α-amylase enzymes, combined with the previous biochemical knowledge of KM1-GTSase.

I-2. Results

I-2-1. Overall and domain structures

The refinement data statistics are summarized in Table I-1. In the crystal structure of 5389-GTSase, there are two monomers in the asymmetric unit. The final model has 96.8% of the residues in the favored conformation of Ramachandran plot, 3.0% of allowed region, and 0.1% of outlier region. Two non-glycine residues, Asp610 in molecule A and B, are outside the normally allowed region in the plot. However, the conformation of these residues was unambiguous in the electron density map. The monomer structure of 5389-GTSase is similar to those of MTSase from *Sulfolobus acidocaldarius* (Sa-MTSase) [33] (PDB ID: **1IV8**) and from *Sulfolobus tokodaii* (St-MTSase) [34] (PDB ID: **3HJE**) with a root-mean-square deviation (RMSD) value of 1.25 Å and 1.22 Å for Cα atoms, respectively.

The monomer of 5389-GTSase consists of five domains, A, B, C, D, and E (Fig. I-2a). Domain A comprises a $(\beta/\alpha)_8$ barrel catalytic domain as is typical in the α-amylase family enzymes, and this domain consists of residues 2–88, 218–302, 434–476, and 571–658

Table I-1. Summary of data collection and refinement statistics

	Native
Data collection	
Detector	MarCCD 165 at SPring-8 BL41XU
Wavelength (Å)	1.0000
Space group	$P2_1$
Unit-cell parameters (Å, °)	a = 71.15, b = 84.56, c = 128.7, β = 103.7
Resolution range a (Å)	2.3 (2.38–2.30)
No. of measured reflections	132,869
No. of unique reflections	59,315
R_{merge} [a,b] (%)	10.4 (37.3)
Completeness of data [a] (%)	90.0 (88.0)
$\langle I \rangle / \langle \Sigma(I) \rangle$ [a]	8.7 (1.7)
Refinement	
Resolution range [c] (Å)	2.30 (2.36–2.30)
R_{work} [c] (%)	21.7 (27.0)
R_{free} [d] (%)	26.4 (35.3)
No. of protein atoms	12,116
No. of water molecules	457
No. of Mg^{2+}/glycerol atoms	2 / 36
R.m.s.d. bonds (Å)	0.022
R.m.s.d. angles (°)	1.969

[a] Values in the parentheses are for the highest resolution shell (2.38–2.30)

[b] $R_{merge} = \Sigma_h \Sigma_i |I(h)_i - \langle I(h) \rangle| / \Sigma_h \Sigma_i \langle I(h)_i \rangle$, where I is the observed intensity and $\langle I \rangle$ is the average intensity of multiple observations of symmetry-related reflections.

[c] Values in the parentheses are for the highest resolution shell (2.36–2.30)

[d] $R_{work} = \Sigma ||F_{obs}| - |F_{calc}|| / \Sigma |F_{obs}|$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively.

[e] R_{free} is calculated as R_{work} using 5% of all reflections randomly chosen and excluded from structure calculation and refinement.

(Figs. I-2b and -3a). This barrel is an incomplete $(\beta/\alpha)_8$ barrel structure which lacks one helix (α5) between Aβ5 and Aβ6. On the other hand, there are an additional two helices (Aα6′ and Aα7′) in the vicinity of Aβ7 and one helix (Aα8′) after Aβ8, respectively. Two β-strands (Aβ6 and Aβ7) are composed of only two residues as are observed in the structures of *Aspergillus* α-amylases, *A. oryzae* (TAKA) [35] and *A. niger* (An) [36]. Domains D (residues 303–433) and E (residues 477–570) insert between two helices (Aα6–Aα6′ and Aα7′–Aα7) in domain A, respectively. The helix Aα7 kinks at

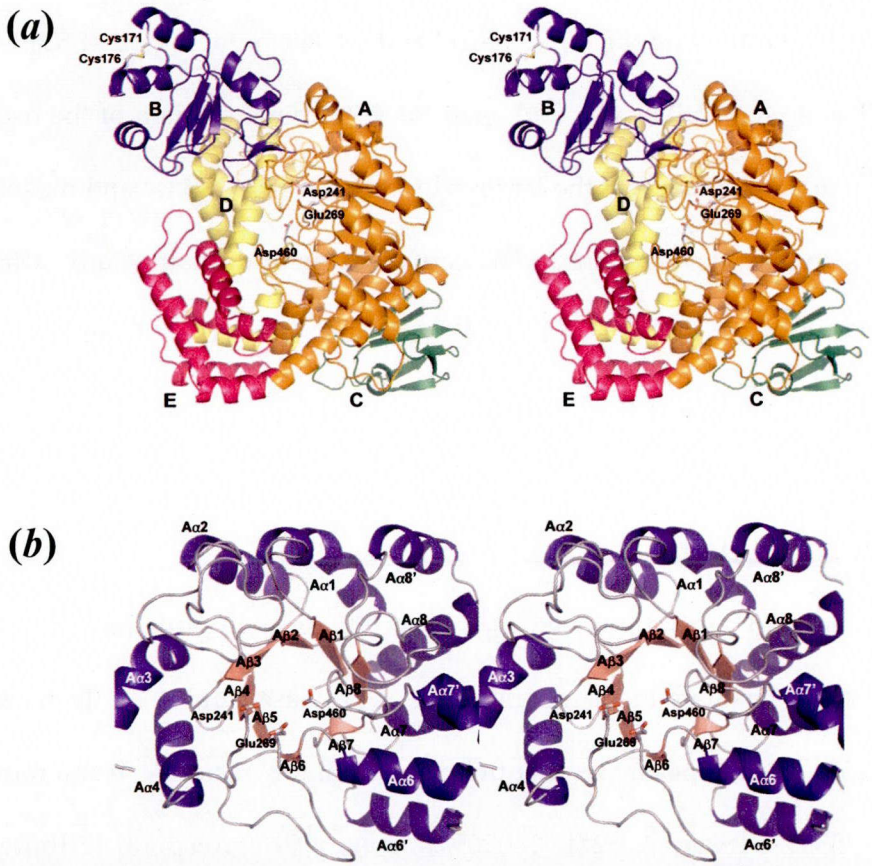


Figure I-2. (*a*) Stereo view of overall structure of 5389-GTSase. Domains A, B, C, D, and E are represented in orange, blue, green, yellow, and purple. The conserved catalytic residues are drawn in stick model, which are Asp241, Glu269, and Asp460. (*b*) Stereo view of structure of domain A. Only domain A is shown for $(\beta/\alpha)_8$ barrel description. The α-helices and β-strands are represented in blue and red. The catalytic residues (Asp241, Glu269, and Asp460) are shown in stick. These figures are drawn by using the program *PyMOL*.

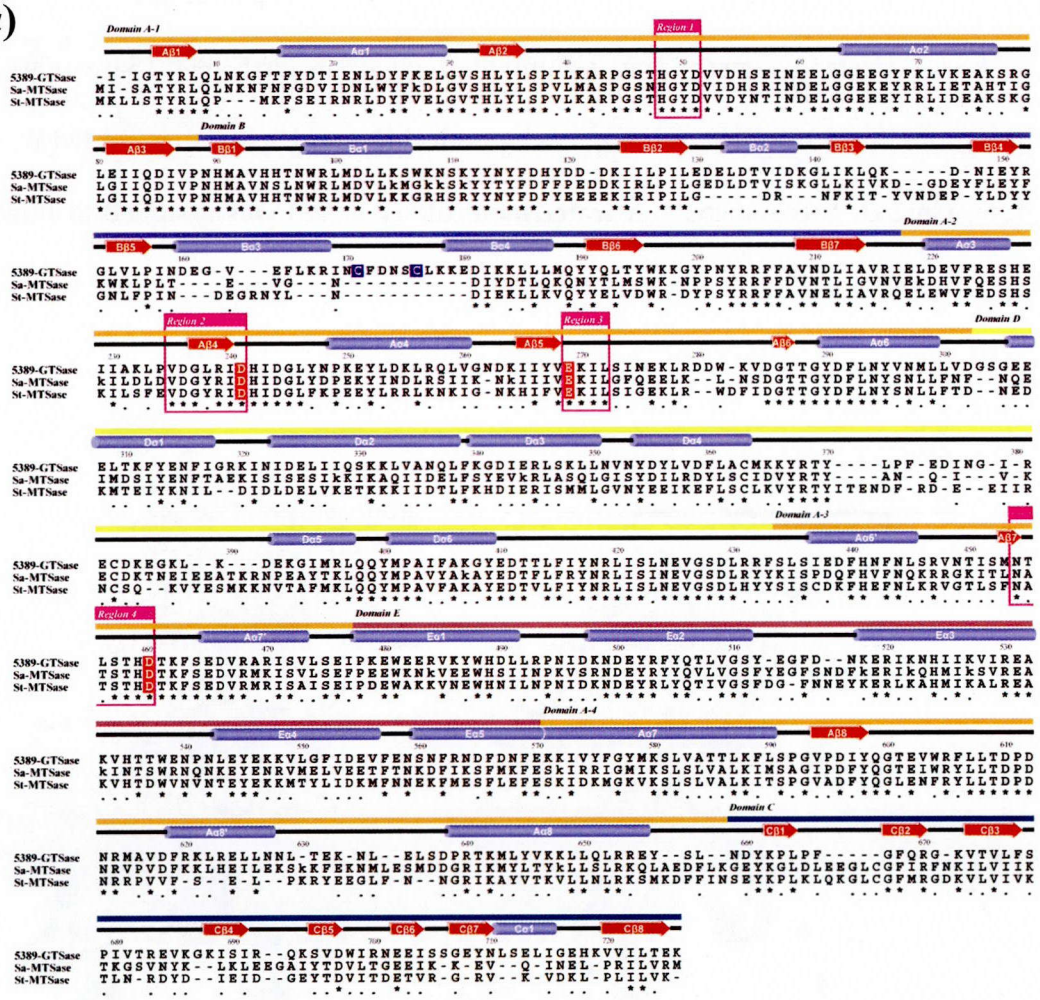
half way, and the N-terminal of Aα7 is included in domain E. Both domains D and E are alpha domain structures, and contain six and five helices, respectively. Domain B (residues 89–217) inserts between Aβ3 and Aα3 in domain A. This domain is composed of four α-helices and seven β-strands which construct two anti-parallel (three and two-stranded) and one two-stranded parallel sheet structures. Domain B has a unique disulfide bond (Cys171–Cys176) which is not found usually in α-amylase family enzymes (Fig. I-2a). Domain C (residues 659–726) forms a six-stranded anti paral-

lel β-sheet structure following the $(\beta/\alpha)_8$ barrel at the C-terminal region of 5389-GTSase. This domain has a γ-crystallin-type fold as in the C-terminal domain of the α-amylase family of enzymes. Domain A is the largest domain in 5389-GTSase and makes extensive contacts with all other domains. The contact areas of domain A are 1,046, 1,521, 1,624, and 1,352 Å^2 against domains B, C, D, and E, respectively. Domains D and E also pack against each other with a contact area of 1,031 Å^2.

I-2-2. Catalytic cleft

Three amino acids, including two aspartic acids and one glutamic acid, are conserved as catalytic residues in members of the α-amylase family [28]. In Sa-MTSase, Asp228, Glu255, and Asp443 are identified as catalytic residues from mutational analyses [37]; in 5389-GTSase, Asp241, Glu269, and Asp460 correspond to those amino acids. These three residues are located on the $(\beta/\alpha)_8$ barrel (Fig. I-2b). Figure I-3b shows the sequence alignment between α-amylase family enzymes at highly conserved four regions in domain A which are considered to be an identifying feature of α-amylases [38]. Three catalytic residues, Asp241, Glu269, and Asp460 in 5389-GTSase, locate in regions 2, 3, and 4, respectively. This cleft is large enough to incorporate the substrate (shadow area in Fig. I-4a). In addition to the three catalytic residues, six additional acidic amino acids, Asp51, Asp211, Asp244, Glu410, Asp411, and Asp610, participate in defining the catalytic cleft. Eleven basic (Arg43, His48, His90, Arg239, His242, Lys270, Lys407, His459, Lys462, Lys527, Arg614) and seven aromatic (Tyr50, Trp196, Phe206, Phe207, Tyr290, Tyr367, Trp538) amino acids also participate in defining the boundary of the catalytic cleft (Fig. I-4a).

(*a*)



(*b*)

		Region 1		Region 2		Region 3		Region 4	
5389	GTSase	HGYD	51	VDGLRID	241	EK-IL	272	NTLSTHD	460
Sa	MTSase	HGYD	51	VDGYRID	228	EK-IL	258	NATSTHD	443
St	MTSase	HGTD	49	VDGLRID	222	EK-IL	252	NATSTHD	441
KM1	GTHase	WGYD	153	VDGFRLD	252	ESDL-	286	VYIQNHD	377
TVAII		HKYD	205	IDGWRLD	325	EIWH-	357	NLLDSHD	421
Ps	Amylase	EGYF	79	AGGFRFD	193	ELWKG	223	TFVDNHD	294
TAKA	Amylase	HGYW	83	IDGLRID	206	EVLDG	234	TFVENHD	297
Bc	CGTase	HGYW	101	IDGIRMD	234	EWFLG	261	TFIDNHD	328

Figure I-3. (*a*) Secondary structure of 5389-GTSase and sequence alignment between 5389-GTSase, Sa-MTSase, and St-MTSase. (*b*) Sequence alignments of GTSases, GTHase, TVAII, α-amylases, and CGTase. 5389-GTSase, Sa-MTSase, St-MTSase, KM1-GTHase, TVAII, Ps-Amylase, TAKA-Amylase, and Bc-CGTase were aligned. Identical residues in all sequences, conserved substitutions and semi-conserved substitutions are represented by red with yellow background, brown, and green, respectively. These alignments and classification of identical residues are performed by the *MATRAS* web service [86]. Three conserved catalytic residues are represented by red background.

Three glycerol molecules (Gol1, 2, 3) derived from cryoprotectant were assigned into $|F_o-F_c|$ electron density map around the active site of 5389-GTSase (Fig. I-4b). Furthermore, the author can confirm one spherical shape density surrounded residues Asp241, Glu269, and Asp460. A water molecule (Wat729) was assigned at this position in St-MTSase [34] (Fig. I-4c). However, the electron density at this position in

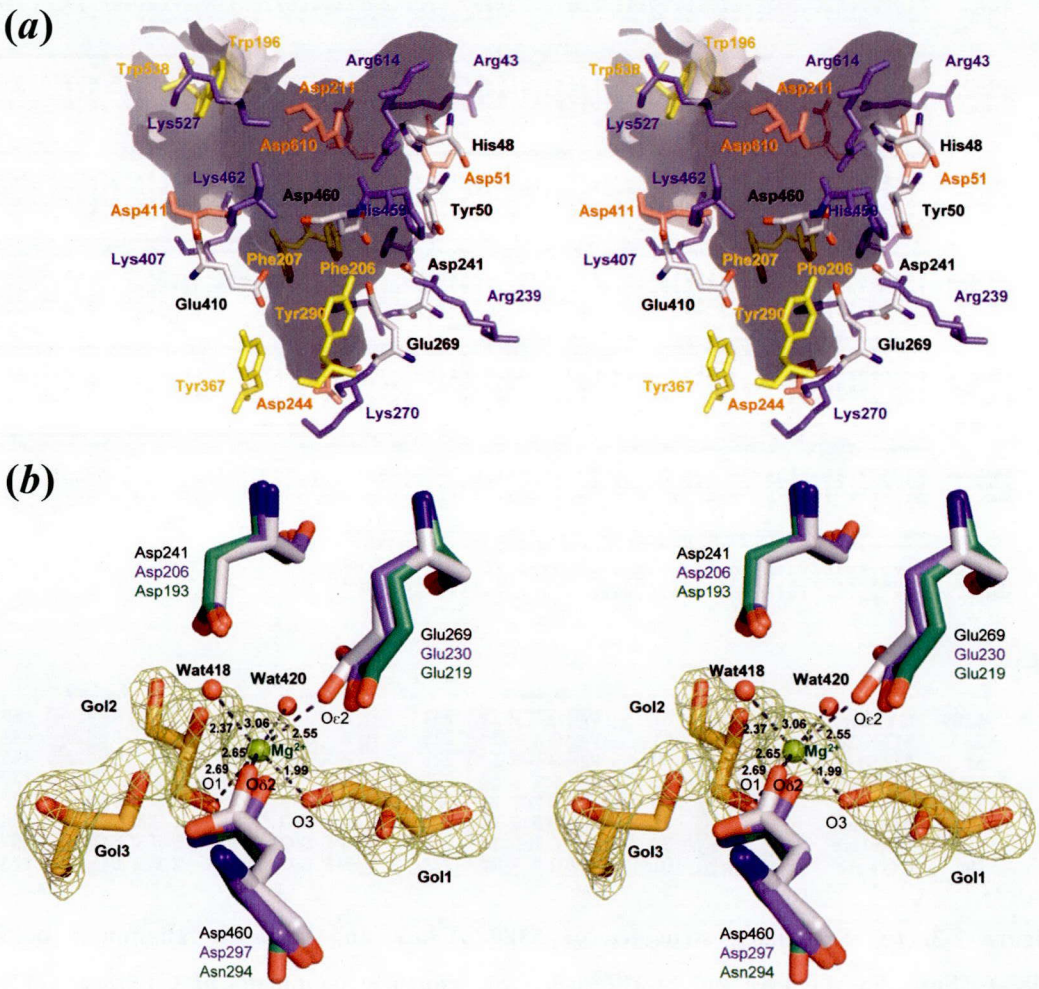


Figure I-4*ab*. (*a*) Catalytic cleft of 5389-GTSase. Acidic, basic, and aromatic residues are represented in red, blue, and yellow, respectively. **(*b*) Mg^{2+} binding site, catalytic residues, and glycerol molecules of 5389-GTSase.** Catalytic residues of 5389-GTSase, TAKA-Amylase, and Ps-Amylase are colored by white, blue, and green, respectively. Gold stick drawing, green, and red spheres are glycerol molecules, Mg^{2+} ion, and water molecules, respectively. Green mesh is 3.0 σ cutoff $|F_o - F_c|$ omit electron density map. Distances are shown near dashed lines (Å).

5389-GTSase is too strong to assign a water molecule because of appearance of excessive electron density. When one Mg^{2+} ion involved in culture medium assigned into the density, no excessive electron density was confirmed. The Mg^{2+} forms a characteristic octahedral coordination with six oxygens derived from two side chains of Glu269, Asp460, two glycerol molecules, and two water molecules (Fig. I-4b). The coordinated oxygens are 2.55 Å (Oε2 of Glu269), 2.65 Å (Oδ2 of Asp460), 1.99 Å (O3 of Gol1), and 2.69 Å (O1 of Gol2), 2.37 Å (Wat418) and 3.06 Å (Wat420) apart from the Mg^{2+} ion. In St-MTSase, three glycerol molecules are also observed in nearly same position in 5389-GTSase (Fig. I-4c). Hydrogen bonds which involve in the glycerol molecules of 5389-GTSase are Gol1 with Glu269 and Glu410, Gol2 with His90 and Asp241, and Gol3 with Asp460 and Arg614.

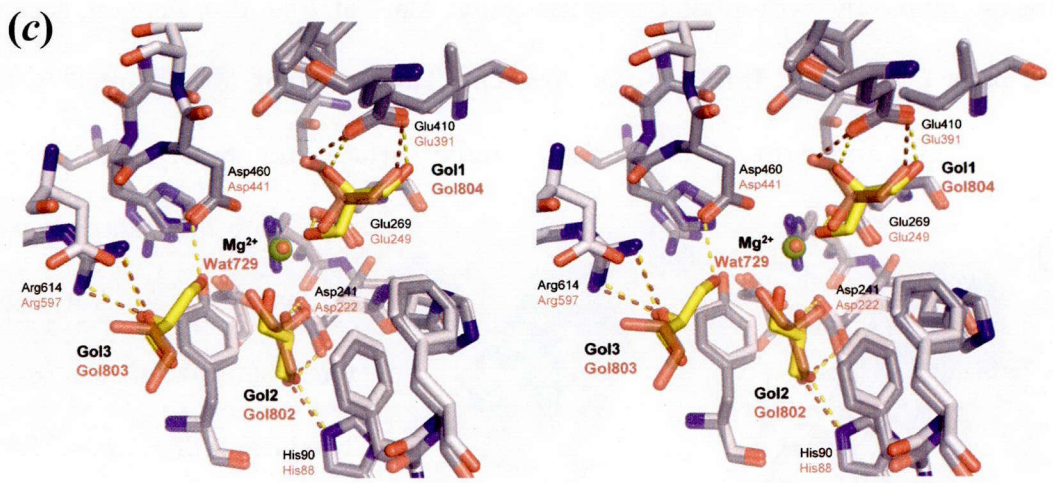


Figure I-4*c*. The comparison of glycerol positions between 5389-GTSase and St-MTSase. 5389-GTSase, St-MTSase, glycerols of 5389-GTSase, glycerols of St-MTSase are represented in white, gray, green, and blue, respectively. Green sphere is Mg^{2+} ion of 5389-GTSase. These are drawn by *PyMOL*.

I-3. Discussion

I-3-1. Thermostability

5389-GTSase is an α-amylase from thermophiles. Significantly, 5389-GTSase has unique disulfide bond between Cys171 and Cys176 (Fig. I-2a). It is not found usually in α-amylase family, which may contribute thermostability of 5389-GTSase. For discussion of overall structure about thermostability, comparison between 5389-GTSase and a mesophilic α-amylase was conducted. Selected structure as a mesophilic α-amylase for comparison was α-amylase from *Aspergillus niger* (An-Amylase; PDB ID: **2AAA**) [36]. This structure had moderate similarity, whose *Z*-score was 21.8 according to *DALI* server [39]. First, their solvent-accessible surface areas (SASA) were compared. Figure I-5 represents solvent-accessible surfaces of GTSase and An-Amylase. The surfaces colored by hydrophobic residues (gray; Ala, Val, Phe, Pro, Ile, Leu, Met, and Gly), polar (green; Ser, Thr, Tyr, Cys, Asn, Gln, His, and Trp), and charged (yellow; Asp, Glu, Lys, and Arg). Characteristic colored surfaces are showed on the figure.

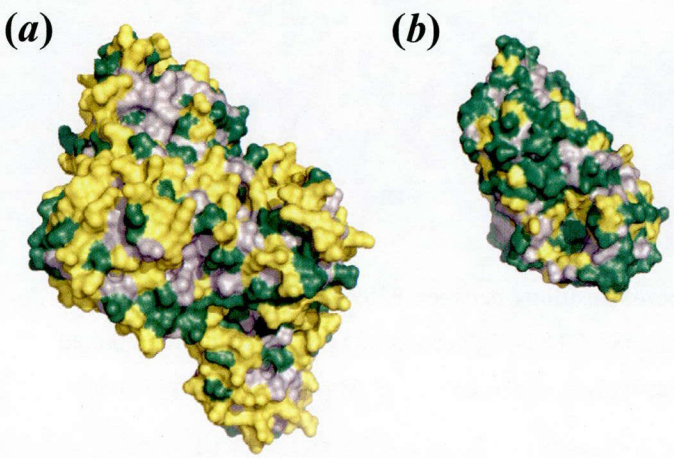


Figure I-5 SASAs of GTSase and An-Amylase. (*a*) Colored SASAs of 5389-GTSase. (*b*) Colored SASAs of An-Amylase. Gray, green, and yellow represents hydrophobic residues, polar, and charged, respectively.

There are differences between GTSase and An-Amylase in the proportion of each residue. The largest SASA of GTSase is charged residues. On the other hand, An-Amylase has polar residues as the largest surface area. Ratios of surface area separated by hydrophobic,

polar, and charged were 17.8%, 25.4%, and 56.8% in GTSase, are 23.1%, 51.0%, and 25.9%, respectively. On the basis of characteristic of SASA, charged-rich surface area contributes to thermostability in GTSase. Next, ion networks (5 Å cutoff) were compared between GTSase and An-Amylase. Those were compared in intra-helices, inter-helices, and networks more than 4 residues. There are characteristic differences in the comparison. Ratios of 5 Å cutoff ion interactions, which are intra-helices, inter-helices, and networks, contoured by more than 4 residues in GTSase were 1.5%, 5.9%, and 1.7%, respectively. In contrast, the ratios in An-Amylase were 0.2%, 0.8%, and 0.2%. Also in this comparison, GTSase had more ion interactions than An-Amylase. Especially, inter-helices networks were the major in ion interaction. Thus, in GTSase, thermostability can be caused by ion interactions which fix structure among helices. Comparison with other proteins will be discussed in Chapter III.

I–3–2. Structural features and reaction mechanism

The crystal structure of 5389-GTSase was determined by X-ray crystallography at 2.3 Å resolution. Structure information of 5389-GTSase elucidated the detail configuration of catalytic residues and the character of substrate binding cleft. Complex structure of the enzyme–substrate (and/or product) can clarify its catalytic mechanism at atomic level directly. However, no complex structures of any GTSase/MTSase have been reported. The author also could not determined the complex structure between 5389-GTSase and substrate/product, despite several attempts for preparation of complex crystals. Therefore, the author constructed the model structures at several states during catalysis between 5389-GTSase and substrate (maltotetraose) based on the structural data of 5389-GTSase and other α-amylase enzymes, combined with the pre-

vious biochemical knowledge of KM1-GTSase [31].

The enzymatic reaction by α-amylase is considered to proceed via general acid/base catalysis [40]. Three catalytic residues in α-amylase family are specified by previous mutagenesis and structural studies [28,41]. In addition, Brzozowski *et al.* describe the hydrolysis mechanism based on complex structure of TAKA-Amylase with the inhibitor acarbose [35]. Asp206 and Glu230 act as a nucleophile and a general acid/base catalyst in TAKA-Amylase, respectively. The role of remained aspartic acid of catalytic residues is describes in a structural and mutagenesis study of *Pseudomonas stutzeri* maltotetraose-forming α-amylase (Ps-Amylase) [42]. Another aspartic acid, Asp294 in Ps-Amylase, works for tightly binding the substrate to give a twisted and a deformed conformation of the glucose ring at position −1. Thus they named this aspartic acid "fixer". From sequence alignment, Asp241, Glu269, and Asp460 correspond to "nucleophile", "acid/base", and "fixer" in 5389-GTSase, respectively (Fig. I-3b). In addition to sequence alignment, the configurations of these three residues in 5389-GTSase are highly conserved (RMSD values of 0.72 Å and 1.02 Å for all atoms of three residues in TAKA-Amylase (PDB ID: **7TAA**) and Ps-Amylase (PDB ID: **1QI3**), respectively) (Fig. I-4b). Brzozowski *et al.* also confirmed using crystallography that transglycosylation event occurred on catalytic reaction by TAKA-amylase [35]. Although for hydrolytic enzymes is usually deglycosylated by an incoming water molecule nucleophile, it is possible that an alternate nucleophile such as an oligosaccharide species attacks, resulting in the formation of an extended sugar chain by transglycosylation. The configurations of catalytic residues and above previous knowledge support that the catalytic residues of GTSase on its transglycosylation mechanism have similar roles of those of other amylases.

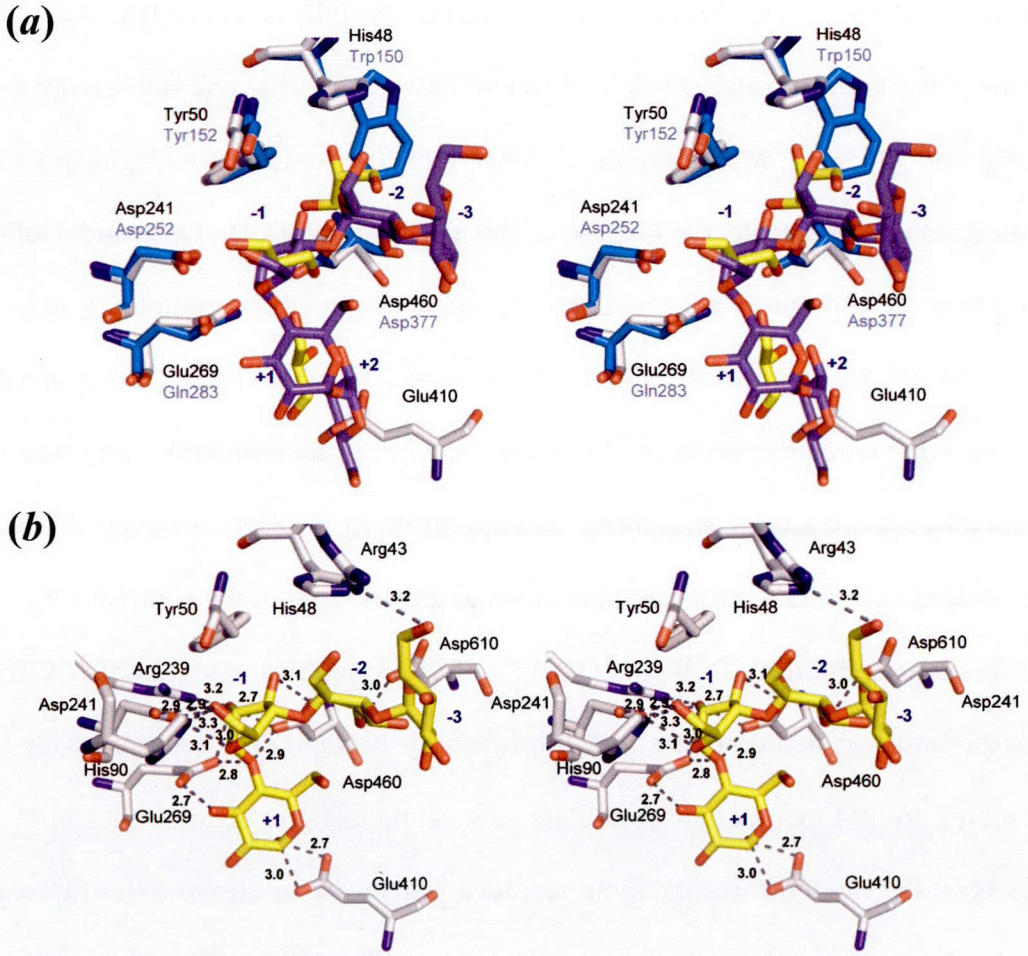


Figure I-6*ab*. (*a*) Superposed model of 5389-GTSase and KM1-GTHase around the catalytic residues. 5839-GTSase and KM1-GTHase are represented in white and cyan, respectively. Glycerol molecules of 5389-GTSase and maltotriosyltrehalose of KM1-GTHase are colored by yellow and blue, respectively. This superposition was using the program *LSQKAB* [134] with a RMSD value of 1.1 Å. **(*b*) Hydrogen bonds (cutoff 3.3 Å) of 5389-GTSase ES complex model.** 5389-GTSase and maltotetraose are white and yellow, respectively. The hydrogen bonds were calculated by using program *NCONT* [88].

For discuss interactions of between 5389-GTSase and oligosaccharide, the author constructed the enzyme–substrate complex (ES complex) model based on the complex structure between KM1-GTHase inactive mutant (E283Q) and its substrate, which is product for GTSase, maltotriosyltrehalose (MTT) (unpublished data). The KM1-GTHase structure with MTT molecule was superimposed to 5389-GTSase by

matching all atoms of each three catalytic residues (RMSD values of 0.91 Å). MTT molecule was placed in catalytic cleft of 5389-GTSase with little collision except for a reducing end glucose. Maltotetraose which removed a reducing end glucose from MTT is substrate molecule for GTSase. The positions of Gol1, Gol2 and Gol3 in 5389-GTSase correspond to the +1, −1 and −2 glucose rings of the maltotetraose in superposed structure, respectively (Fig. I-6a). Glycerol molecule used as a cryoprotectant can occupy similar position of oligosaccharide in crystal structure. For example, in cellobiohydrolase Cel6A structures, two glycerols in enzyme structure (PDB ID: **1OC6**) indeed occupied almost same positions of glucose ring in the complex structure with thio-oligosaccharides (PDB ID: **1OC5**) [43]. The ES complex model of 5389-GTSase with superimposed maltotetraose was improved by harmonically restrained for side chain atoms around maltotetraose simulated annealing using program *CNS* 1.21 [44]. In the ES complex model, three catalytic residues just locate to cleave α-1,4 glycosidic bond at between the last two glucose of maltotetraose (Fig. I-6b). The configuration of Asp241, Glu269, and Asp460 satisfy the role for "nucleophile", "acid/base", and "fixer", respectively. The C1 atom of the −1 glucose is positioned near the Oδ of Asp241 as possible to nucleophilic attack. The enzymatic acid/base, Glu269, makes a hydrogen bond with the glycosydic oxygen between −1 and +1 glucose. Asp460 fixes the −1 glucose via two hydrogen bonds. In addition, aromatic residues, His48 and Tyr50, make stacking interaction with substrate, −2 and −1 glucose, respectively. These two aromatic residues are conserved as substrate recognition sites in α-amylase family (region 1 in Fig. I-3b). Stacking interaction between two aromatic residues and −2, −1 glucose rings was also confirmed in the complex structures, α-amylase 2 / acarbose (inhibitor complex) [45], cyclodextrin glucosyltransferase (CGTase) mutant / substrate (co-

valent intermediate) [46] Glu410 interacts the +1 glucose ring and fix the position of the reducing end of maltooligosaccharide. Three catalytic residues can be located to cleave α-1,4 glycosidic bond at between the last two glucose by these substrate recognition interaction.

Catalysis by α-amylase is widely believed to take place via a double-displacement reaction, in which a covalent glycosyl-enzyme intermediate is formed after cleavage of α-1,4 glycosidic bond and subsequently hydrolyzed via oxocarbenium ion transition states [35,40]. The covalent intermediate is formed via covalent bonding of nucleophilic

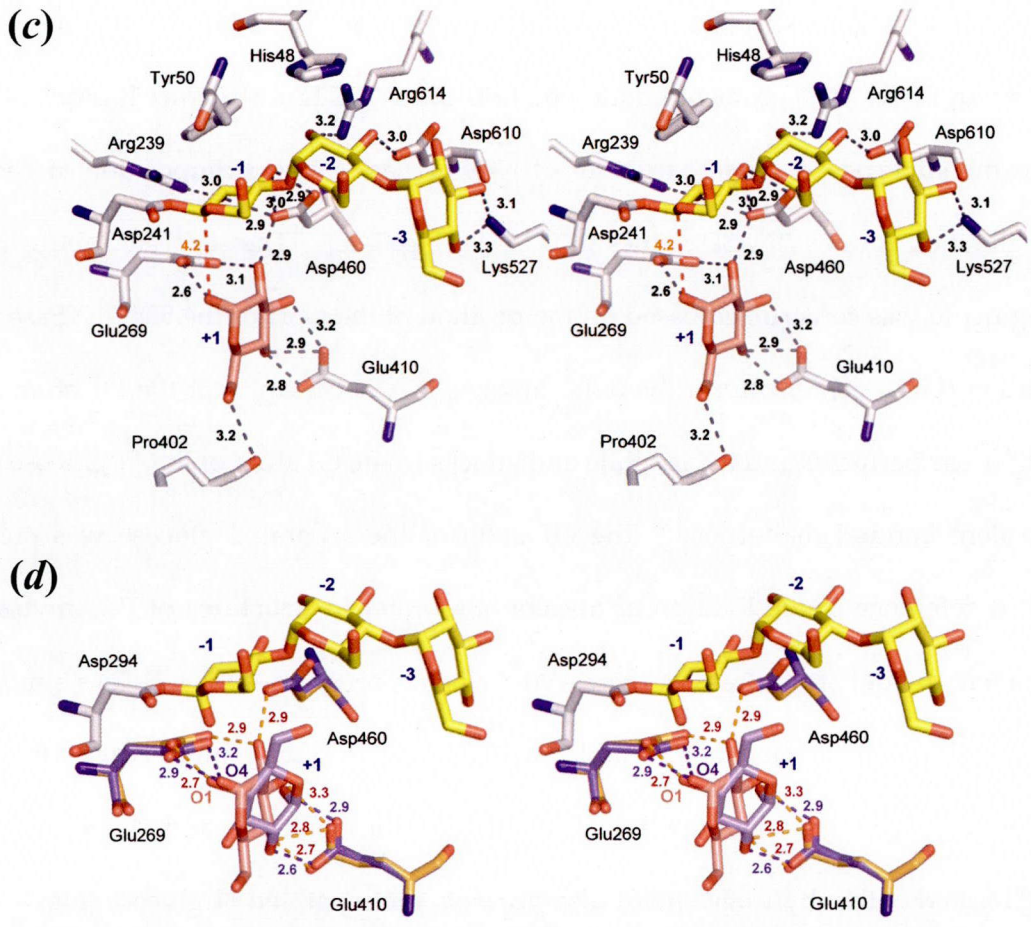


Figure I-6*cd*. (*c*) Hydrogen bonds (cutoff 3.3 Å) of 5389-GTSase intermediate covalent bonded model. Maltotriose and glucose are represented in yellow and red, respectively. (*d*) Comparison of hydrogen bonds (cutoff 3.3 Å) between α-1,1 position and α-1,4 position. α-1,1 position and α-1,4 are represented in red and blue, respectively. These are drawn by *PyMOL*.

aspartate (Asp241 in 5389-GTSase) to the maltooligosaccharide at C1 position. In intramolecular transglycosylation mechanism, the O1 atom of the separated +1 glucose attacks to the C1 atom of the −1 glucose of the covalent bonded maltotriose. The O1 atom performs as a nucleophile in this situation. The author also constructed the model structure of the covalent intermediate with maltotriose and a separated glucose. The covalent intermediate structure of the CGTase from *Bacillus circulans* E257Q inactive mutant (PDB ID: **1CXL**; [46]) was used for reference structure to construct intermediate model. This coordinate was selected by using *DALI* server [39] with *Z* score of 22.2 as highest homologous structure to 5389-GTSase from covalent intermediate complex structures in PDB. The covalent bonded maltotriose in CGTase structure has no collision on the superposed model structure of 5389-GTSase by superimposition at each three catalytic residues (RMSD values of 0.34 Å) (Fig. I-6c). The separated reducing end α-glucose was constructed based on the position of the Gol1 of the 5389-GTSase at +1 position (Fig. I-6c). To form the α-1,1 linkage, it is necessary that the O1 atom of the +1 glucose performs as a nucleophile and attacks to the C1 atom of the −1 glucose of the covalent bonded maltotriose. The O1 atom of the separated glucose was positioned by reference to the location of nucleophilic water in structures of Ps-Amylase. Intermediate model structure was improved by using program *CNS* 1.21 in a similar manner as the construction of ES model structure. In present intermediate model structure, side chain oxygens of two catalytic residues (Glu269 and Asp460) and Glu410 located within hydrogen bond distance from the separated +1 glucose (Fig. I-6c). There was an additional hydrogen bond between O5 atom of glucose and carbonyl oxygen of Pro402. On the assumption that O4 atom of the separated glucose immediately after cleavage was located at the O1 position in present intermediate model, hy-

drogen bonds concerned with Asp460 and Pro402 disappeared in the model structure immediately after cleavage (Fig. I-6d). From the comparison between both model structures, character of substrate binding cleft of 5389-GTSase may be advantage to rotation of the separated +1 glucose in order to form the α-1,1 linkage.

Catalytic mechanism by 5389-GTSase is summarized in Figure I-7 schematically. GTSase once cleaves α-1,4 glucosidic bond at the reducing end of the polysaccharide and releases the α-glucose in the same manner as the initial step of hydrolysis by α-amylase, and then, the released glucose rebinds to maltotriose via α-1,1 glucosidic bond by its rotation motion in the catalytic cleft.

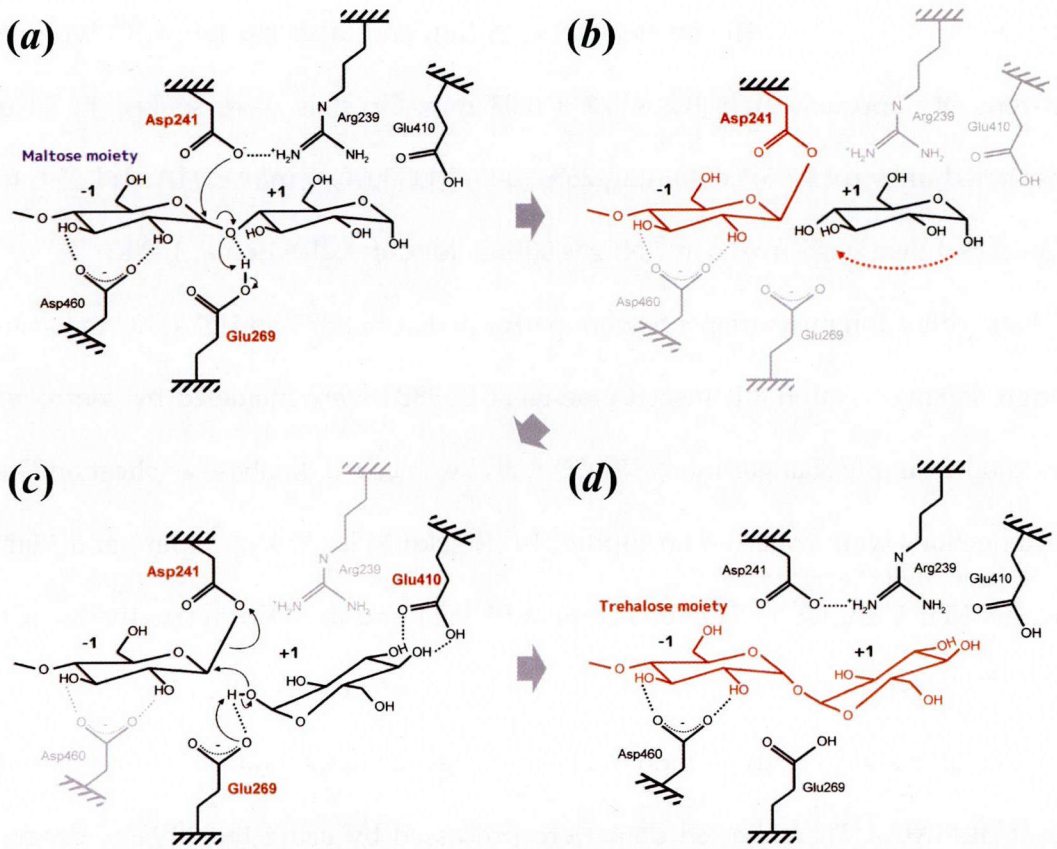


Figure I-7. Possible mechanism of catalysis involving three conserved residues, ES complex, and covalent intermediate state in 5389-GTSase. Residue numbers are as in 5389-GTSase. (*a*) 5389-GTSase complexed with maltose moiety as the ES complex. (*b*, *c*) The covalent intermediate state of reaction. (*d*) Generated the trehalose moiety.

I-4. Materials and methods

I-4-1. Crystallization and data collection

Both 5389-GTSases and KM1-GTHase were used for crystallization experiments. Expression and purification of KM1-GTSase were as described previously [27]. 5389-GTSase also expressed and purified by the same way. Co-crystallization experiments were performed using the mix solution of 10–20 mg/ml (0.13–0.26 m*M*) of GTSase and 1.3–2.6 m*M* of MTT which is the main product from maltopentaose. Crystals of both GTSase were obtained in 50 m*M* phosphate buffer (pH 7.5) containing 20% (*w*/*v*) polyethyleneglycol (PEG) 8000 and 1 m*M* EDTA by hanging drop vapor diffusion method at 20°C. The crystals grew as thin and layer clusters with typical dimensions of approximately 0.2 × 0.2 × 0.03 mm. Crystals were soaked in 50 m*M* phosphate buffer (pH 7.5) containing 25% (*w*/*v*) PEG8000, 1 m*M* EDTA and 15% (*v*/*v*) glycerol and then flash-frozen in a N_2 gas stream for data collection at 100 K.

Data collection measurements were performed at BL38B1 and BL41XU in SPring-8 (Hyogo, Japan). Automatic measurements at BL38B1 were managed by control software *BSS* [13], sample changer robot *SPACE* [10], and web based database application *D-Cha* [14]. Reflections were collected on Jupiter210 (Rigaku/MSC, Tokyo, Japan) and Mar165 (Rayonix/Mar USA, USA) CCD detectors at BL38B1 and BL41XU, respectively. Crystals of KM1-GTSase had poor diffraction power (below 4 Å resolution) for precise structure analysis. The diffraction data of 5389-GTSase were collected to 2.3 Å resolution at BL41XU. The collected data were processed by using the suite of programs *HKL2000* [47]. The crystal belonged to a space group of $P2_1$ with unit cell dimensions a = 71 Å, b = 85 Å, c = 129 Å, and β = 103.7°. The data collection statistics are shown in Table I-1.

I-4-2. Refinement

Initial phase information was obtained by the molecular replacement method using the program *MOLREP* [48] with Sa-MTSase [33] (PDB ID: **1IV8**) as the search model. In the result, two GTSase molecules were found in the asymmetric unit with result of solvent content of 47%. Model refinement was carried out by using the programs *CNS* 1.21 [44] and *REFMAC* 5.5 [49]. Manual model building, including identification and placement of solvent was performed with the program *Coot* [50]. Inter-domain protein contact areas were calculated using the *AREAIMOL* [51] software. RMSD for an overlay of atomic coordinates are calculating using the *MATRAS* web service [52]. The refinement data statistics are also given in Table I-1. The final dimer model contains 1450 residues, 464 waters, two magnesium ions and six glycerol molecules. Densities derived from MTT could not be observed in this crystal. According to *RAMPAGE* [53] in the CCP4 Program Suite [54], the final model has 96.5% of the residues in the favored conformation of Ramachandran plot, 3.3% of allowed region, and 0.2% of outlier region. Three non-glycine residues, Asp610 in molecule A, and Asp122, Asp610 in molecule B, are outside the normally allowed region in the plot. However, the conformation of these residues was unambiguous in the electron density map.

Chapter II
Glycosyltrehalose trehalohydrolase from *Sulfolobus solfataricus* KM1

II-1. Introduction

Trehalose (O-α-D-glucopyranosyl-(1→1)-α-D-glucopyranoside) is used for energy storage and for protecting proteins and cell membranes from extreme temperatures and osmotic shock in plants, insects, and microorganisms [16–18]. A coupled enzyme pathway for the production of trehalose from soluble starch has been discovered in the *Arthrobacter* sp. Q36 [20,21] as well as in the *Sulfolobaceae* family of thermoacidophilic archaebacteria [23–27]. This pathway comprises two enzymes: glycosyltrehalose synthase (GTSase; also called maltooligosyltrehalose synthase; MTSase), an intramolecular glycosyltransferase (1,4-α-glucan 4-α-glycosyltransferase, EC 2.4.1.25), and glycosyltrehalose trehalohydrolase (GTHase; also called maltooligosyltrehalose trehalohydrolase; MTHase), 4-α-D-((1→4)-α-D-glucano)trehalose trehalohydrolase, EC 3.2.1.141) which is

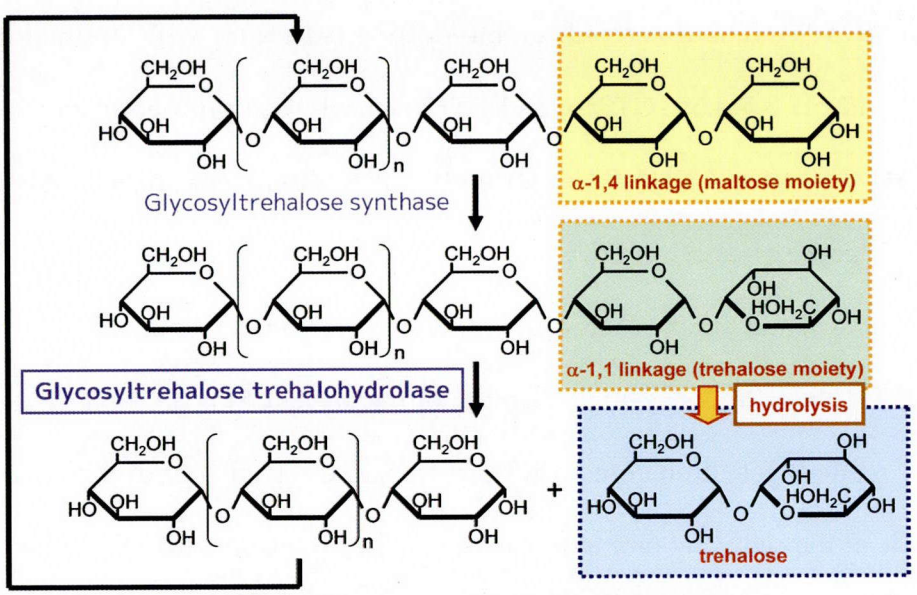


Figure II-1. Biosynthesis of trehalose production

one of α-amylases (α-1,4-D-glucanohydrolase, EC 3.2.1.1). These enzymes are classified within the glycoside hydrolase (GH) family 13 [29] (*CAZy* database [30]; http://www.cazy.org/). Furthermore, family GH13 is divided into subfamilies [55] and GTHase is classified into the subfamily GH13_10, which is described as 4-α-(1,4-α-glucano)trehalose -trehalohydrolase. GTSase converts the glycosidic bond between the last two glucose residues of a maltooligosaccharide from an α-1,4 bond to an α-1,1 bond, making a non-reducing maltooligosyltrehalose (Fig. II-1). GTHase then cleaves the α-1,4 bond adjacent to the α-1,1 bond, releasing a trehalose molecule and regenerating a substrate for GTSase. This pathway differs considerably from the previously characterized phosphate-dependent pathway [56] in which trehalose is synthesized from glucose-6-phosphate (a high-energy glycolytic intermediate) and UDP- or GDP-glucose by trehalose-6-phosphate synthase (EC 2.4.2.15) and trehalose-6-phosphate phosphatase (EC 3.1.3.12).

Sulfolobus solfataricus KM1, a member of the *Sulfolobaceae* family of hyperthermophilic, acidophilic archaea, was isolated from an acid hot spring in Gunma Prefecture, Japan in 1993 [27]. It is a coccoid, gram-negative bacterium with optimal growth at 75–85 °C and pH 3.5–4.5. GTSase and GTHase have been purified from *S. solfataricus* KM1 (KM1-GTSase and KM1-GTHase); each of the two enzymes have also been cloned, sequenced, and expressed in *Escherichia coli* (GTSase) [57] and *Candida utilis* (GTHase) [58]. The biochemical and kinetic characteristics of these two enzymes are as follows.

KM1-GTHase (Mr, 64,644 Da; number of residues, 558) is an exoamylase that hydrolyzes maltooligosyltrehalose to release trehalose. It also hydrolyzes maltooligosaccharide at the reducing end to release glucose or maltose with ~16-fold lower activity [25]; this side reaction decreases the purity of the trehalose, and the extra step of puri-

fication decreases the yield of industrial trehalose production. Therefore, the exact substrate recognition mechanism of this enzyme is of interest in order to increase its substrate specificity. The activity of KM1-GTHase places it in the α-amylase family, which includes α-amylases, cyclodextrin glucosyltransferases [46], *Thermoactinomyces vulgaris* R-47 α-amylase II (TVAII) [45], and neopullulanases [59]. The author's group has previously determined the crystal structure of wild-type KM1-GTHase by X-ray diffraction [32]. The structure of the wild-type KM1-GTHase comprises three major domains (A, C, and E) and two subdomains (B and D), and with an N-terminal extension forming a stably folded immunoglobulin type domain connected by an extended linker peptide to the $(\beta/\alpha)_8$ barrel catalytic domain. The enzyme exists as a symmetric dimer covalently linked by a disulfide bond at Cys298.

In order to investigate the catalytic and substrate recognition mechanisms, the author mutated two residues, Asp252 and Glu283, located at the active site of KM1-GTHase. Two single and one double amino acid mutant KM1-GTHase proteins were produced and crystal structures of the three mutant proteins in complex with their substrate were obtained. Here, the author presents the entire interaction scheme between KM1-GTHase and its substrate; the scheme includes a specific mechanism for recognition of the trehalose moiety in the +1 and +2 substrate binding subsites. The recognition mechanism to distinguish the substrate and product is discussed.

II-2. Results

II-2-1. Enzymatic activities of the E283Q, D252S, and D252E mutants of GTHase

The α-amylase reaction retains the α-anomeric configuration of the scissile glycosidic bond and is considered to proceed via general acid/base catalysis [40,46]. Based on the amino acid sequence similarities between the α-amylases, the two acidic residues Glu283 and Asp252 are predicted to be involved in the catalytic function of GTHase as a general acid/base and nucleophile, respectively. Glu283 was mutated to the corresponding amide (Gln) to abolish the general acid/base function and to inactivate the GTHase completely. Moreover, Asp252 was mutated to Ser to change its catalytic function from anomer retaining mechanism to inverting [7] and also mutated to Glu with the expectation of formation of an enzyme-substrate adduct, as has been successfully performed in a β-glycosidase (T4-phage lysozyme) [60–62].

The enzymatic activities of the wild-type and mutant GTHases were determined using maltotriosyltrehalose (G3-Tre) as a substrate (Table II-1). Under these conditions, native GTHase showed an enzymatic activity of 611 U/mg. The D252S mutant showed very little enzymatic activity (3.4 U/mg), and the D252E and E283Q mutants had almost no enzymatic activity. Detected enzymatic activities of the D252E and E283Q mutants were less than 0.3 U/mg (0.05% of the activity of wild-type GTHase), which is close to the limit of saccharide detection.

Table II-1. Enzymatic activities for the mutant GTHases.

	Enzymatic activity (U)	Protein amount (mg)	Relative activity (U/mg)
Wild type	501	0.82	611
E283Q	0.252	0.92	0.274
D252S	3.47	1.02	3.40
D252E	0.139	0.81	0.172

II-2-2. Structures of wild-type, D252S, D252E, and E283Q mutant GTHases with substrates

The structure of wild-type GTHase complexed with G3-Tre, where G3-Tre was soaked into the crystal, was determined at 2.65 Å resolution. The general structural features are the same as previously described [32]; these comprise of domains A, B, and C, which are common within the α-amylase family, domains D and E, which are unique in GTHase, the N-terminal extension of the immunoglobulin type domain connected by an extended linker peptide to the $(\beta/\alpha)_8$ barrel catalytic domain, and the covalently linked dimeric structures at Cys298. However, no additional electron density for G3-Tre was observed (Fig. II-2).

To further investigate the location of the substrate, the catalytic site mutants of GTHase were crystallized with G3-Tre. The E283Q mutant GTHase in complex with G3-Tre (E283Q-G3-Tre) and the E283Q mutant in complex with maltoheptaose (G7) (E283Q-G7) crystallized isomorphously to the wild-type GTHase (Table II-2), and their X-ray structures were determined at 2.60 and 2.65 Å resolutions, respectively. These structures revealed extra electron density at the substrate binding region (Figs. II-2, -3b, and -3c), which were assigned as G3-Tre in E238Q-G3-Tre and maltopentaose (G5) in E283Q-G7. The bound substrate was clearly observed in the E283Q-G3-Tre complex.

The D252S and D252E mutants in complex with G3-Tre (D252S-G3-Tre, D252E-G3-Tre) also crystallized isomorphously to the wild-type GTHase (Table II-2), and their X-ray structures were determined at 2.30 and 2.40 Å resolution, respectively. The crystal structure of D252S-G3-Tre revealed extra electron density at the substrate binding region (Fig. II-3d), which was assigned as maltotriose (G3), but no electron

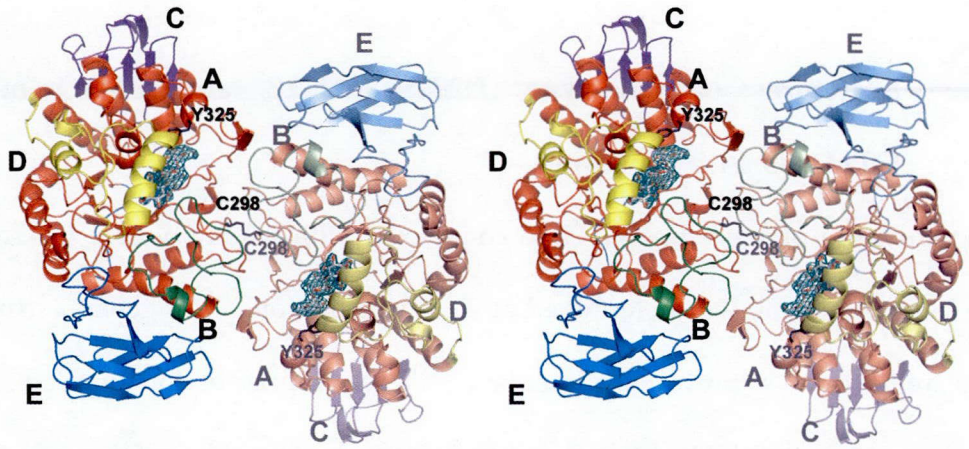


Figure II-2. Stereo view of the overall dimer structure of E283Q mutant GTHase. The dimer is crosslinked by an intermolecular disulfide bridge at Cys298. Domains (A-E) are colored by red, green, blue, yellow, and cyan, respectively. Green mesh represents $|F_o–F_c|$ omit map contoured 3.0 σ around G3-Tre. Tyr325 is stacking with the subsite -1 glucose. The figures were prepared by *PyMOL* (http://www.pymol.org/).

density corresponding to the substrate was observed in the structure of D252E-G3-Tre.

To investigate the location of the water molecule that may be activated by the newly introduced Ser252, the crystal structure of D252S mutant GTHase without substrate was determined at 2.70 Å resolution. The differences in the root mean square (RMS) distances for the Cα atoms in wild-type GTHase and in E283Q-G3-Tre, E283Q-G7, D252S-G3-Tre, and D252E-G3-Tre mutant GTHases were 0.26, 0.30, 0.32, and 0.40 Å, respectively. In the crystal structure of the D252S mutant, a density corresponding to the water molecule interacting with Ser252 was observed because of occupation by glycerol, which was used as a cryoprotectant.

Table II-2. Data processing and refinement statistics of the wild type and mutant KM1-amylases.

KM1-amylase Substrate	wild type G3-Tre	E283Q G3-Tre	E283Q G7	D252S none	D252S G3-Tre	D252E G3-Tre
Data Collection						
X-ray source	SPring-8 (BL40B2)	SPring-8 (BL41XU)	SPring-8 (BL41XU)	SPring-8 (BL41XU)	SPring-8 (BL40B2)	SPring-8 (BL41XU)
detector	ADSC Quantum 4R	Mar CCD 165	Mar CCD 165	Mar CCD 165	ADSC Quantum 4R	Mar CCD 165
resolution (Å)	2.65 (2.74-2.65)	2.60 (2.69-2.60)	2.65 (2.74-2.65)	2.70 (2.80-2.70)	2.30 (2.38-2.30)	2.40 (2.49-2.40)
total reflections	178,567	230,847	195,222	166,502	442,848	316,350
unique reflections	29,375	31,666	29,402	27,427	45,920	40,919
Redundancy	6.1 (3.2)	7.3 (4.7)	6.6 (4.5)	6.1 (3.3)	9.6 (7.1)	7.7 (6.8)
$\langle I/\sigma(I)\rangle$	10.2 (1.8)	13.2 (1.7)	9.3 (1.4)	18.1 (2.9)	40.5 (4.5)	30.5 (5.0)
R_{merge}[a]	0.097 (0.296)	0.099 (0.296)	0.086 (0.349)	0.096 (0.193)	0.087 (0.366)	0.060 (0.179)
Completeness (%)	97.3 (84.1)	97.9 (81.7)	97.3 (95.9)	95.7 (89.2)	99.9 (99.3)	99.9 (99.9)
Refinement						
resolution (Å)	2.65 (2.72-2.65)	2.60 (2.67-2.60)	2.66 (2.72-2.66)	2.70 (2.77-2.70)	2.30 (2.36-2.30)	2.40 (2.47-2.40)
used reflections	27,856	30,005	27,909	26,037	43,509	38,820
R_{work}[b]	0.201 (0.341)	0.197 (0.300)	0.175 (0.311)	0.159 (0.219)	0.177 (0.236)	0.158 (0.201)
R_{free}[b] (5% random)	0.243 (0.366)	0.244 (0.322)	0.227 (0.342)	0.212 (0.295)	0.218 (0.299)	0.198 (0.247)
No. of total atoms	4,721	4,800	4,769	4,773	4,888	4,973
Protein	4,513	4,550	4,550	4,553	4,553	4,538
Water/GOL*/FLC*	177/18/13	163/18/13	144/6/13	165/42/13	252/36/13	392/30/13
substrate	–	56 (G3-Tre)	56 (G5)	–	34 (G3)	–
mean B value (Å^2)	44.5	45.8	40.7	40.9	47.2	37.0
R.m.s.d. bonds (Å^2)	0.007	0.007	0.018	0.021	0.017	0.014
R.m.s.d. angles (°)	1.085	1.134	1.868	1.961	1.694	1.537

[a] $R_{merge} = \Sigma|I - \langle I\rangle| / \Sigma I$, where I is the intensity of a reflection and $\langle I\rangle$ is the average intensity.

[b] $R_{work(free)} = \Sigma||F_{obs}| - |F_{calc}|| / \Sigma|F_{obs}|$, where F_{obs} and F_{calc} are observed and calculated structure factor amplitude, respectively.

* GOL and FLC are three-letter-codes of glycerol and citrate anion in PDB, respectively.

II–2–3. Interaction of the substrate in the E283Q mutant GTHase

From the structures of inactive mutant GTHases in complex with G3-Tre, the entire substrate-binding site was determined. The electron density belonging to the substrate was detected at five subsites, from −3 to +2, which is shown most clearly in the structure of E283Q-G3-Tre (Fig. II-3b). The G3-Tre molecule bound GTHase through 21 hydrogen-bond interactions as shown in Table II-3a and Figure II-4. The trehalose end of maltooligosyltrehalose and the reducing end of maltooligosaccharide bind towards the C-terminal side of the $(\beta/\alpha)_8$ barrel. The non-reducing end of the substrate is expected to exit from the N-terminal side of the barrel at the interface with domain A.

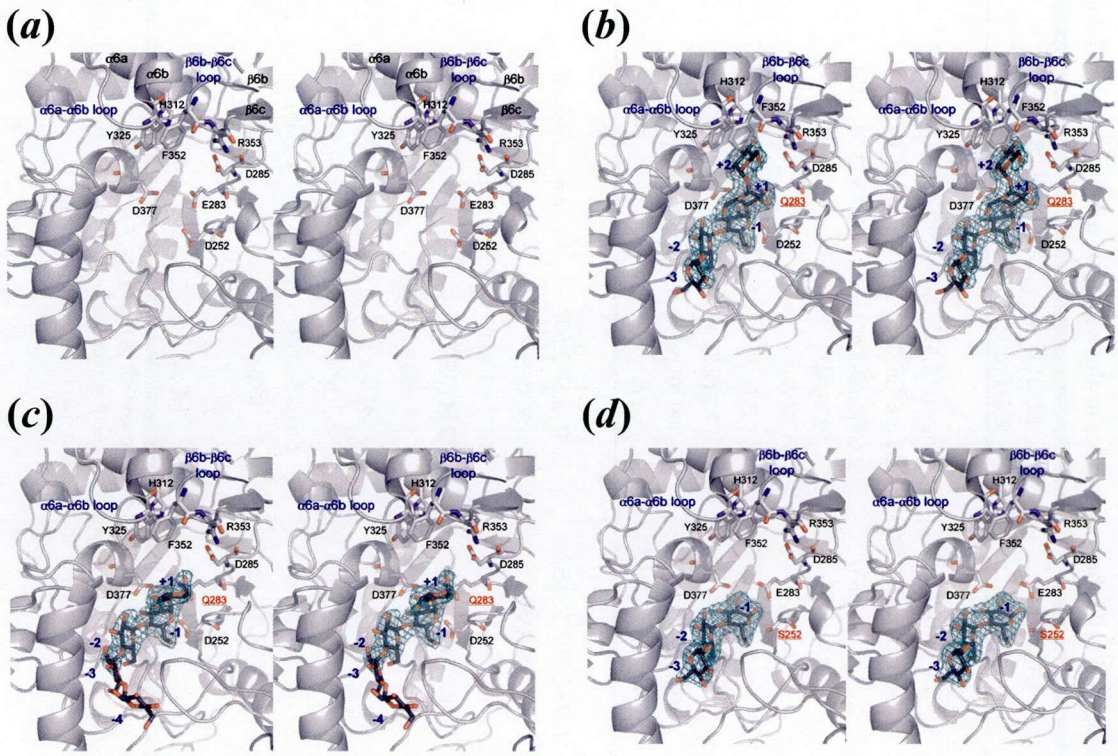


Figure II-3. Extra electron densities observed in the mutant GTHases. (*a*) Wild-type. (*b*) E283Q mutant GTHase in complex with maltotriosyltrehalose (G3-Tre). (*c*) E283Q mutant GTHase in complex with maltoheptaose (G7). (*d*) D252S mutant GTHase cocrystallized with G3-Tre. $|F_o-F_c|$ maps are colored in green mesh and sugar models fitted to the $|F_o-F_c|$ map are colored in black bonds.

The glucose molecule at subsite +2 of G3-Tre interacts with the distorted α-hairpin structure overhanging from the α6a–α6b and β6b–β6c loops (Fig. II-3b); in this interaction, Tyr325, Phe352, and Arg353 residues contribute to form a steric barrier at the +2 subsite (Fig. II-3b). The subsite +2 glucose and Tyr325 have potential stacking interactions in which the pyranose and hydroxyphenyl rings are nearly parallel and recognized by four hydrogen bonds; three between the O3 in the subsite +2 glucose moiety and Asp308, His312, and Gln378 and one between O4 in the subsite +2 glucose and His312 (Fig. II-5a).

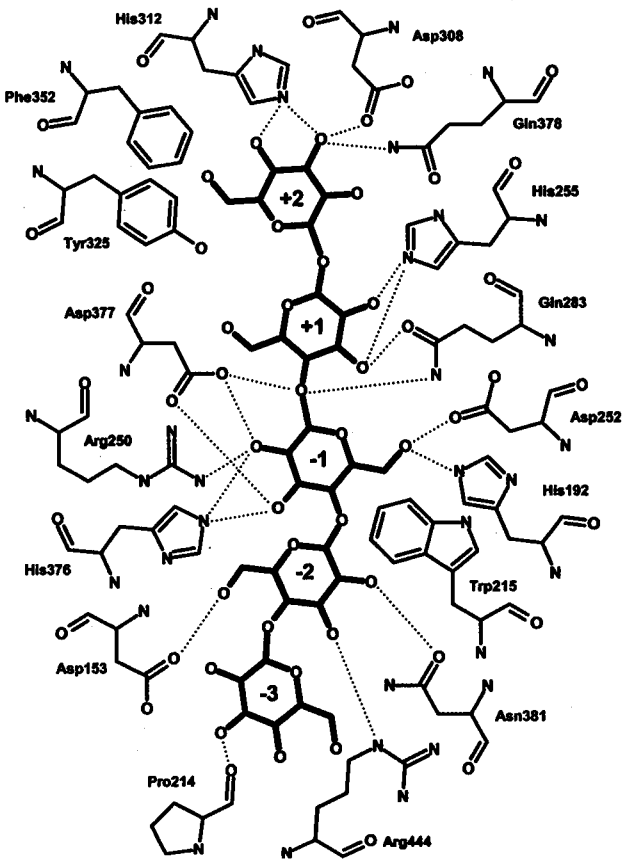


Figure II-4. Schematic view of the hydrogen bonding interactions (dotted) between the E283Q mutant GTHase and G3-Tre. Key hydrophobic residues for substrate recognition, Trp215, Phe325 and Phe352 are also included.

At the +1 subsite, the glucose residue exhibits five hydrogen bonds; involving O2 and Nε2 of His255, O3 and Nε2 of His255, O3 and Oε1 of Gln283, O4 and Nε2 of Gln283, and O4 and Oδ1 of Asp377. The subsite +1 glucose makes a van der Waals contact with Trp215 in which the planes of the indole and pyranose rings are almost perpendicular (Figs. II-5a and -6). The indole ring of Trp215 is fixed by adjacent Pro214 (Fig. II-5a).

The subsite −1 glucose appears distorted toward a half-chair form and is sur-

rounded by the invariant residues of Tyr152, His192, Trp215, and His376 (Fig. II-6). When a glucose with standard conformation was superposed using the C1, C2, and C3 atoms of the sugar ring (RMS deviation = 0.027 Å), the glucose ring distortion is hard to detect at this resolution, but was more obvious for the accompanying positional shift of O6 and C6 of the subsite −1 glucose (positional shift of C6 was 1.6 Å). This distortion is stabilized by forming seven hydrogen bonds, which is the largest number among the five subsites. Of these seven bonds, three are between O2 and Nη2 of Arg250, Nε2 of His376 and Oδ1 of Asp377, two are between O3 and Nε2 of His376 and Oδ2 of Asp377, and two are between O6 and Nε2 of His192 and Oδ2 of Asp252 (Table II-3a and Fig. II-6).

The subsite −2 glucose participates in three hydrogen bonds; one between O2 and Oδ1 of Asn381, one between O3 and Arg444, and one between O6 and Oδ1 or Oδ2 of Asp153. The subsite −3 glucose exhibited only one hydrogen bond between O3 of the subsite −3 glucose and O of Pro214. On the basis of all the structural information obtained for the inactive mutants of GTHase, at least five substrate recognition subsites (−3 to +2 subsites) were identified in GTHase.

To further investigate how GTHase distinguishes the correct substrate, G3-Tre and G7 (non-preferred substrates) were included in the crystallization of the E283Q mutant GTHase. Electron density belonging to a moiety of G5 lying within the −3 to +1 subsites was sufficient to build saccharide structures, although the electron density at the −3 and −2 subsites were relatively weak (Fig. II-3c). No electron density corresponding to a saccharide moiety bound at the +2 subsite was observed, and the location of the subsite +1 glucose was shifted outwards from the saccharide-binding cleft as shown in Fig. II-5c. Observed hydrogen bonds are listed in Table II-3b.

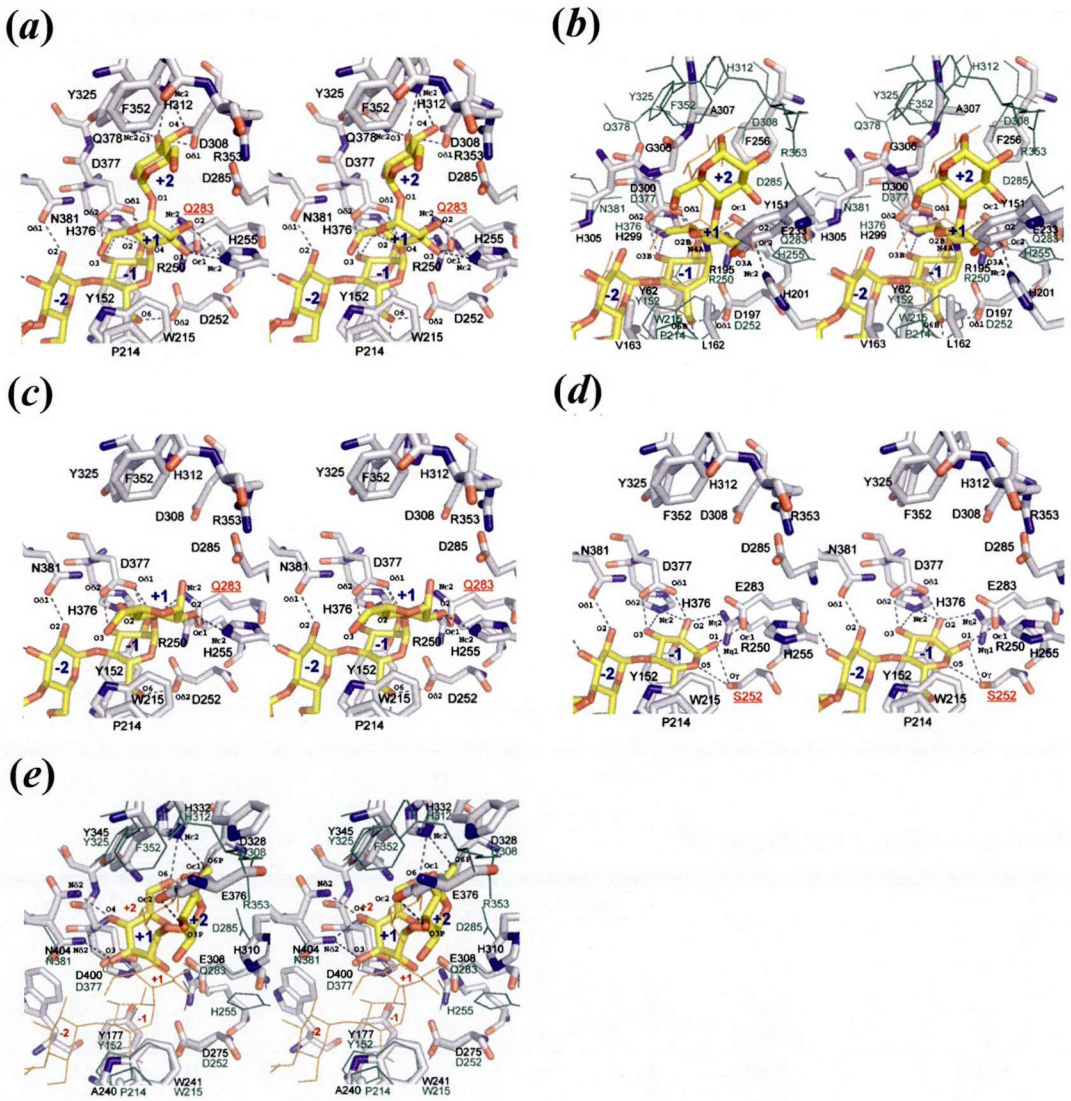


Figure II-5. Interaction of saccharide in each substrate-binding subsite of GTHase. (*a*) G3-Tre bound structure of E283Q mutant GTHase (E283Q-G3-Tre). (*b*) Structure of pig pancreatic α-amylase (PPA) with acarbose. E283Q-G3-Tre structure (thin bonds in green) is superposed as reference model. (*c*) Structure of E283Q mutant GTHase with G7 (E283Q-G7). (*d*) Structure of D252S mutant GTHase with G3-Tre (D252S-G3-Tre). (*e*) The trehalose bound structure of MTHase from *Deinococcus radiodurans*. E283Q-G3-Tre structure (thin bonds in green) is superposed as reference model.

Table II-3.

(*a*) Hydrophilic interaction between E283Q and G3-Tre

Protein		Subsite number		Distance
Asp308	Oδ2	+2	O3	2.7
His312	Nε2	+2	O3	3.0
His312	Nε2	+2	O4	3.2
Gln378	Nε2	+2	O3	2.8
His255	Nε2	+1	O2	3.0
His255	Nε2	+1	O3	3.1
Gln283	Oε1	+1	O3	2.9
Gln283	Nε2	+1	O4	3.0
Asp377	Oδ1	+1	O4	3.1
His192	Nε2	−1	O6	3.3
Arg250	Nη2	−1	O2	3.1
Asp252	Oδ2	−1	O6	2.7
His376	Nε2	−1	O2	2.8
His376	Nε2	−1	O3	2.8
Asp377	Oδ1	−1	O2	2.7
Asp377	Oδ2	−1	O3	2.9
Asp153	Oδ1	−2	O6	2.8
Asp153	Oδ2	−2	O6	3.2
Asn381	Oδ1	−2	O2	2.7
Arg444	Nε	−2	O3	3.2
Pro214	O	−3	O3	3.3

(*b*) Interaction between E283Q and G5

Protein		Subsite number		Distance
His255	Nε2	+1	O2	3.2
Gln283	Oε1	+1	O2	2.9
Gln283	Nε2	+1	O3	3.1
Asp377	Oδ1	+1	O3	3.0
Arg250	Nη2	−1	O2	2.9
Asp252	Oδ1	−1	O2	3.3
Asp252	Oδ1	−1	O5	3.3
Asp252	Oδ2	−1	O6	2.6
Gln283	Nε2	−1	O2	3.3
His376	Nε2	−1	O2	2.8
His376	Nε2	−1	O3	2.7
Asp377	Oδ1	−1	O2	2.8
Asp377	Oδ2	−1	O3	2.6
Asp153	Oδ1	−2	O6	2.8
Asn381	Oδ1	−2	O2	2.6
Arg444	Nε	−2	O3	2.9
Glu447	Oε1	−3	O2	3.0

(*c*) Interaction between D252S and G3

Protein		Subsite number		Distance
His192	Nε2	−1	O6	3.3
Ser252	Oγ	−1	O1	3.0
Ser252	Oγ	−1	O5	3.2
Arg250	Nη2	−1	O2	3.0
Glu283	Oε1	−1	O1	2.3
His376	Nε2	−1	O2	2.9
His376	Nε2	−1	O3	3.1
Asp377	Oδ1	−1	O2	2.5
Asp377	Oδ2	−1	O3	2.6
Asp153	Oδ1	−2	O6	2.8
Asp153	Oδ2	−2	O6	3.3
Asn381	Oδ1	−2	O2	2.8
Arg444	Nη2	−2	O3	3.1
Pro214	O	−3	O3	3.0

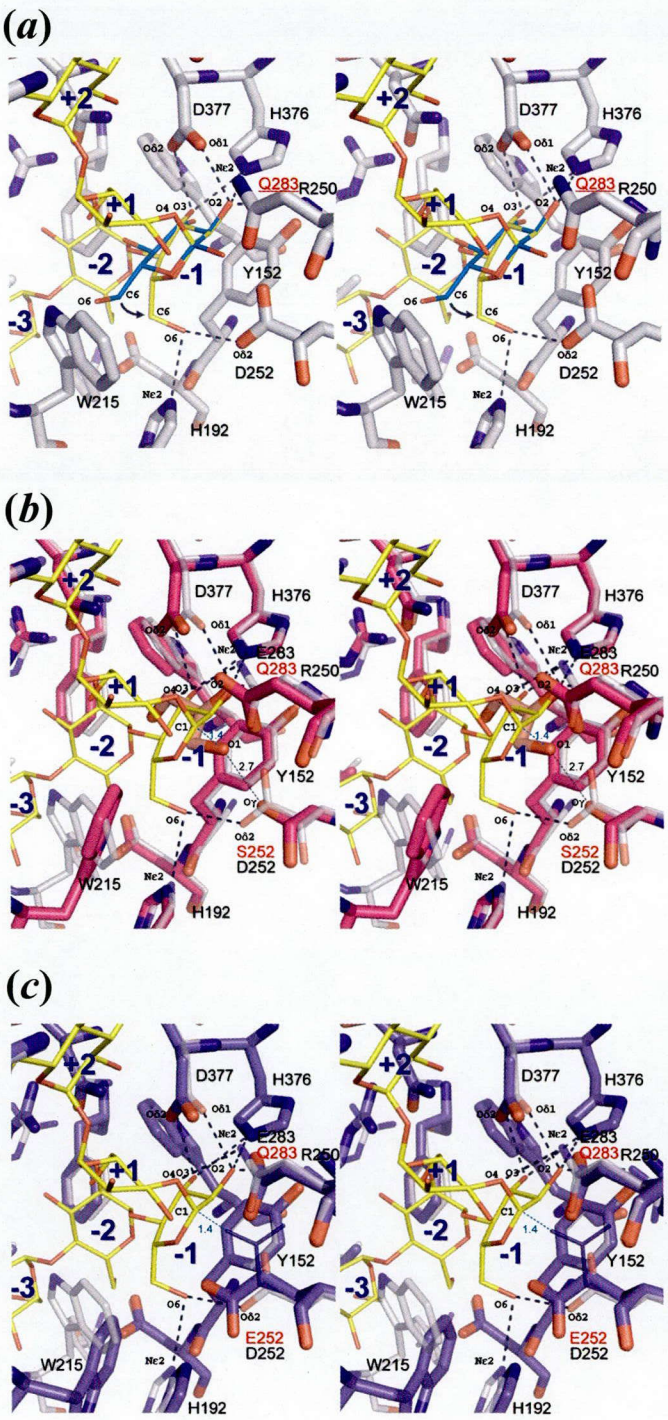


Figure II-6.

(*a*) Pyranose ring distortion observed at the −1 subsite. Thick bonds represent the structure of E283Q mutant GTHase, while thin bonds (colored in yellow) represent the bound G3-Tre. The glucose molecule in the standard conformation is represented by thin cyan bonds. (*b*) Superposition of D252S with glycerol and E283Q complexed with G3-Tre. D252S, glycerol, and E283Q-G3-Tre are represented in pink, red, and white, respectively. (*c*) Superposition of D252E and E283Q-G3-Tre. D252E and E283Q-G3-Tre are colored in blue and white, respectively. A rotamer of Glu252 side chain as a model of covalent enzyme-substrate adduct is shown in the thin stick presentation.

II-2-4. Saccharide interactions of the D252S and D252E mutant GTHases

Asp252 in GTHase, similar to Asp197 in pig pancreatic α-amylase (PPA) [63], participates as a nucleophile that attacks the C1 carbon of the subsite −1 glucose. Therefore, it was reasoned that the substitution of Asp252 with serine (shorter side chain) or glutamate (longer side chain) may result in a change in the anomer form of the product or formation of a covalent adduct.

Electron densities corresponding to an enzymatic product, G3, were observed over the −3 to −1 subsites in the crystal structure of the D252S mutant GTHase, whereas no clear density corresponding to a G3-Tre molecule was observed in the structure of the D252E mutant GTHase. The product bound structure of the D252S is shown in Figure II-3d. Hydrogen bonding interactions were observed between Oγ of Ser252 and O1 of the saccharide molecule at the reducing end. The interactions between G3 and D252S mutant GTHase are listed in Table II-3c. It was also found that the saccharide molecule at the reducing end showed the α-anomer as shown in Figure II-5d.

The D252E mutant was expected to form a covalent enzyme-substrate adduct in which the newly introduced glutamate residue would react directly with the G3-Tre; however, no electron density corresponding to the substrate was observed. Consistent with this interpretation, the location of the side chain of Trp215 was shifted inward, and the side chain occupied the substrate-binding cleft of GTHase accompanying the structural shift of the loop region from 211 to 218.

II-3. Discussion

II-3-1. Thermostability

It was a remarkable point that a thermostable protein GTHase formed dimer structure linked by disulfide bond. The disulfide bond formed at side chain of Cys298 to the other molecule of the dimer (Fig. II-2). Generally, oligomerization increases thermostability of protein. The disulfide bond of Cys298 contributes thermostability of GTHase also. As similar to discussion of GTSase in former chapter, SASA and ion networks were compared with α-amylase from *A. niger* (An-Amylase; PDB ID: **2AAA**)[36]. Colored surfaces of GTHase and An-Amylase are shown in Figure II-7. Ratios of hydrophobic, polar, and charged residues (in the same classification of GTSase in Chapter I) on surfaces were 18.4%, 26.0%, and 55.6% on GTHase, and 23.1%, 51.0%, and 25.9% on An-Amylase, respectively. These ratios of surfaces were calculated *AREAIMOL* [51]. In GTHase, surface area of charged residues was major of the SASA. In comparison of ion networks (5 Å cutoff), GTHase showed the similar trend of GTSase also. Ion networks of intra-helices, inter-helices, and more than 4 residues are 0.9%, 4.5%, and 1.1% in GTHase, respectively. Those results propose that ion interactions contribute thermostability of GTHase.

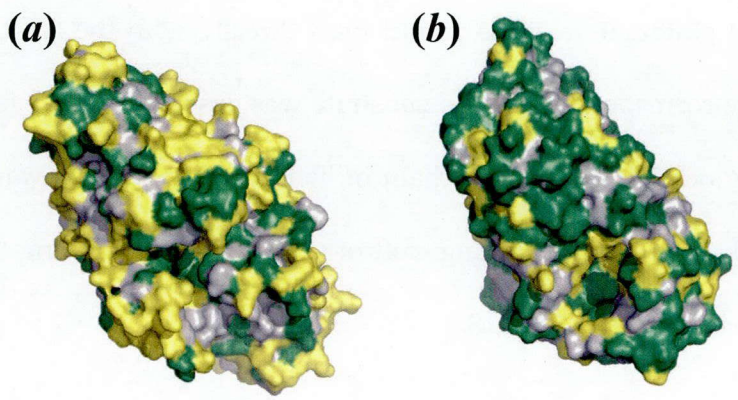


Figure II-7 SASAs of GTSase and An-Amylase. (*a*) Colored SASAs of KM1-GHSase. (*b*) Colored SASAs of An-Amylase. Gray, green, and yellow represents hydrophobic residues, polar, and charged, respectively.

II–3–2. Structure and reaction mechanism

It has been predicted that several substrate-binding subsites and some active site structures act as steric barrier to capture and recognize the substrate at the +2 subsite in GTHase [25]. To determine the substrate recognition mechanism of GTHase, which is important for improving the substrate specificity and thus increasing the enzymatic yield for industrial trehalose production, the author undertook two mutagenic approaches. One was complete inactivation by substitution of catalytic acid/base Glu283 with the corresponding amide, and the other was to trap the enzyme-substrate adduct by substituting Asp252 with two other nucleophiles, namely, serine and glutamate. Although the author's attempt to obtain an enzyme–substrate adduct was unsuccessful, the structure analysis of E283Q-G3-Tre revealed that at least 5 subsites from −3 to +2, including the +1 and +2 subsites, directly recognize the trehalose moiety of the substrate. This observation was consistent with the result that a substrate longer than G3-Tre shows a similar reaction rate and only one or two glucose residues can be cleaved from maltooligosaccharides.

The structure of bound trehalose reported for maltooligosyltrehalose trehalohydrolase (MTHase; another name of GTHase) derived from *Deinococcus radiodurans* [64] (Fig. II-5e) (PDB ID: **2BHY**) is quite different from the structure for E283Q-G3-Tre (Fig. II-5a). The glucose moiety of trehalose interacts with His310 and Glu376 in *Deinococcus radiodurans* MTHase (Fig. II-5e) whereas the glucose moiety of G3-Tre at the +2 subsite was surrounded by several residues, Asp308, Tyr325, Phe352, and Arg353 in GTHase. The stacking interaction between the pyranose and hydroxyphenyl ring of Tyr325 observed at the +2 subsite in GTHase was nearly parallel, indicating a stacking interaction for trehalose recognition. In contrast, the stacking interaction with Tyr345,

which corresponds to Tyr325 of GTHase, was not observed in the trehalose complex of *Deinococcus radiodurans* MTHase. Thus, the steric barrier in GTHase is attributed to the residues Tyr325, Phe352, and Arg353 located at the +2 subsite (Fig. II-5a).

The structural analysis of E283Q-G7 also provided important information about the substrate specificities of KM1-GTHase. It was known that GTHase weakly binds G7 that has α-1,4 linked glucose residues [25]. The structure analysis of E283Q-G7 revealed that the +2 subsite is vacant and G7 is bound through the +1 to −4 subsites (Figs. II-3c and -5c), indicating the lack of affinity of the +2 subsite to glucose with α-1,4 linkage. Since the α-1,1 linkage of the trehalose moiety gives it a very different shape than the α-1,4 linkage gives to maltose, GTHase has a different stacking interaction of the aromatic side chains (Tyr325 in KM1-GTHase is similar to Tyr151 in PPA [63] [PDB ID: **1HX0**]) to recognize the α-1,1 linked glucose moiety at the +2 subsite (Figs. II-5a and -5b). To compare recognition mechanisms between trehalose and maltose moieties, the author chose PPA because, there is a crystal structure of PPA complexed with acarbose in high-resolution and PPA has high structure similarity against GTHase (Z = 47.5 with 3D library search using *MATRAS* [52]).

In order for this observation to be more useful, the author further investigated the conformation of the trehalose moiety in the GTHase-G3-Tre complex and compared it to the calculated minimum energy conformation for trehalose and maltose as shown in Fig. II-8. The angle of α-1,1 bond ($\phi = -26.8°$ and $\psi = -46.9°$) was similar to that of the lowest energy conformation of trehalose ($\phi = -48.8°$, $\psi = -48.8°$) [65]. Furthermore, the low-energy conformations of maltose ($\phi = -2.31°$, $\psi = -23.6°$) [65] are very different from those of trehalose. Tyr325 provides specific binding with a low energy conformation of the subsite +2 glucose in the trehalose moiety. Only very high-energy conforma-

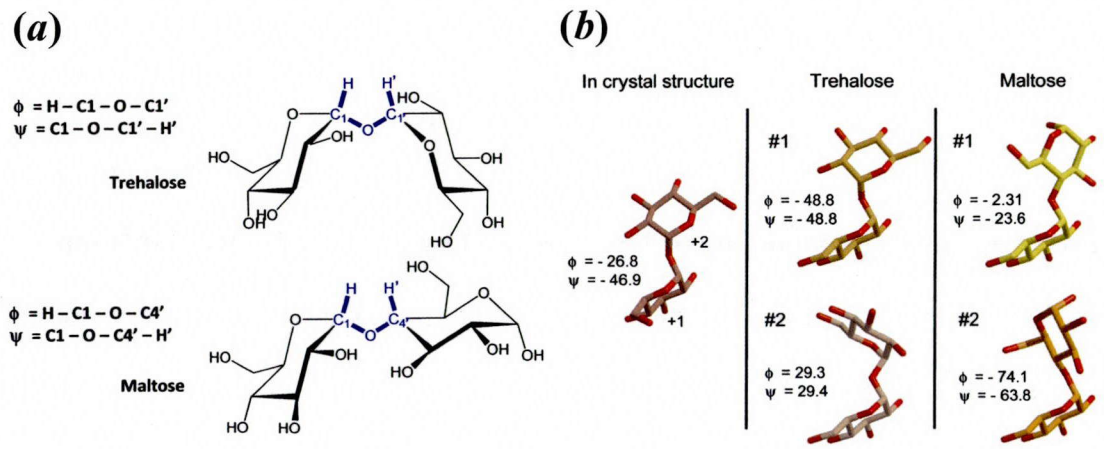


Figure II-8 (*a*) φ and ψ angle definitions of trehalose and maltose, respectively. (*b*) Comparison of the +1 and +2 conformation and minimum energy conformations of trehalose and maltose. The lowest (#1) and second lowest (#2) energy conformations of trehalose and maltose are shown. The definitions of the φ and ψ angles are provided at the left.

tions of maltooligosaccharides could have the same stacking interaction with Tyr325; this is consistent with the finding of the co-crystallization experiment of E283Q-G7 in which the +2 subsite was empty. This result indicates that the +2 subsite is disfavored for α-1,4 linked glucose moiety relative to α-1,1 linked glucose moiety. It seems that an important aspect of the enzyme's specificity for maltooligosyltrehalose arises from the ability of the α-1,1 linked glucose residue at the +2 subsite to make a stacking interaction with Tyr325, which is then not available to α-1,4 linked glucose residues.

The mechanism of binding of the natural substrate (G3-Tre) to the inactive E283Q mutant KM1-GTHase, especially the +1 and −1 subsites, was successfully determined. The glucose moiety of G3-Tre bound to the +1 subsite was essentially in the same position and orientation as seen in PPA (Figs. II-5a and -5b). The glucose molecule appears to form hydrogen bonds with His192, Gln283, and Asp377 via its 2-, 3-, 4-, and 6-hydroxyls and has a van der Waals contact with Trp215. Trp215 interacts with both the glucose molecules at the −1 and +1 subsites; in this interaction, the indole ring of

Trp215 is located almost perpendicular to the pyranose ring of the +1 subsite and also has contact with the C6 of the subsite −1 glucose (Fig. II-6). This hydrophobic contribution of Trp215 to the substrate binding at the −1 and +1 subsites appears unique to GTHase because the indole ring of Trp215 is fixed by adjacent Pro214 and is replaced with 2 hydrophobic residues (Leu162 and Val163) in PPA, which may conserve the role of Trp215 (Fig. II-5b). The subsite −1 glucose of G3-Tre is also stabilized by the invariant residues Tyr152, His192, and His376 (Figs. II-5a and -6). Tyr152 makes a stacking interaction with the saccharide ring of the glucose at the −1 subsite (Fig. II-6). His192 and His376 form hydrogen bonds with the 6-hydroxyl and the 2- or 3-hydroxyl of the glucose at the −1 subsite via their Nε atoms. These interactions are commonly conserved in α-amylases [66] and also shown by His101 and His299 in PPA complexed with acarbose. Together with the collision between the O6 atom and the indole ring of Trp215, these histidine residues seems to stabilize the distorted half-chair conformation proposed to be adopted by this glucose residue in the transition state. Indeed, the mutants of MTHase obtained from *S. solfataricus* ATCC 35092 that have substitution at Trp218 (W218A and W218F, Trp218 corresponds to Trp215 of KM1-GTHase) show a decrease in specific activities and catalytic efficiencies [67].

The structure of E283Q with G3-Tre will provide clues as to how to convert the catalytic mechanism of this enzyme. Previously, the author's group found that the location of nucleophile (Asp52) in hen egg white lysozyme is structurally equivalent to Thr26 in T4 phage lysozyme, and the replacement of these positions with a glutamate resulted in forming an enzyme-substrate adduct in both α-glycosidase (T4-phage lysozyme) [7,60–62] and β-glycosidase (hen egg white lysozyme) [68]. The author's group also found from the structural analysis of a series of mutant T4 phage lysozymes [60] that the

size of Ser/Thr with a bound water molecule is similar to that of a glutamate [68]. Therefore, the conversion of a nucleophile at Asp252 to shorter side chains (Thr, Ser) may position the water molecule between the enzyme and substrate and result in changing a double displacement mechanism to a single displacement mechanism. The conversion to a longer side chain (Glu252) forms a stable covalent enzyme-substrate adduct.

Although the author could not identify the location of the water molecule interacting with Ser252, the location of O1 of the bound glycerol (suggesting the oxygen position of water) was hydrogen bonded to OG of Ser252 and was located at a distance of 1.4 Å from the C1 of the subsite −1 glucose (Fig. II-6b). This suggests that the side chain of Ser252 may provide a catalytic water molecule to cleave the glycosidic bond between subsites −1 and +1 glucose through a single displacement mechanism. In this case, the enzymatic product of the D252S mutant GTHase should be a β-anomer, which is inconsistent with the α-anomer seen in the crystal structure of D252S-G3-Tre. Due to the relatively long period for crystallization, the β-anomer formed after hydrolysis may have spontaneously mutarotated to the more stable α-anomer.

The crystal structure of E283Q-G3-Tre also provided an explanation for the inactivation of the D252E mutant. The substrate-binding site of D252E was empty, despite the presence of G3-Tre in the crystallization solution. When the structure of D252E is superposed on the structure of E283Q-G3-Tre, the side chain of Glu252 forms an ion-pair with adjacent His192. The author originally expected that the side chain of Glu252 would be located near the C1 of the subsite −1 glucose (relative rotamer position of Glu252 is represented by thin sticks in Fig. II-6c). The shift of the Glu252 side chain towards His192 observed in the crystal structure of D252E mutant appears to

block the substrate-binding site at the −1 subsite, which may also lead to loss of the enzymatic activity in the D252E mutant (Table II-2).

Together with the crystal structures of the above-mentioned catalytic site mutants, the structural analysis of E283Q-G3-Tre suggests the specific substrate recognition mechanism, including the role of the structural features of the substrate binding site in the recognition of the trehalose moiety and stabilization of the distorted substrate. These findings are important for improving the catalytic activity and substrate specificity of GTHase, especially for the industrial applications of this enzyme.

II-4. Materials and methods

II-4-1. Materials

G3-Tre and G7 were prepared as reported previously [27]. Restriction endonucleases and T4 DNA ligase were obtained from Takara Shuzo (Japan). KOD DNA polymerase was obtained from TOYOBO (Japan). The concentration of purified protein was spectrophotometrically estimated at 280 nm using the previously reported extinction coefficient [32].

II-4-2. Mutagenesis

D252S, D252E, and E283Q variants were prepared by substituting the KM1-GTHase open reading frame of pGUSS2 [58] with the appropriate synthetic oligonucleotide. Replacement was performed using a mutagenesis kit (LA PCR in vitro Mutagenesis Kit, Takara, Japan) according to the protocol provided by the manufacturer. The *Xba*I-*Bgl*II fragment of pSS was ligated to the *Xba*I and *Bam*HI sites of

pGAPURA1 to construct pGUSS1. The pGAPURA1 construct was prepared by inserting a *Not*I-*Pst*I fragment containing the promoter and terminator sequences of the GAP gene isolated from the plasmid pGAPPT10 [69] into the *Not*I and *Pst*I sites of pURAL10, which contains the CYHr gene as well as the 5′ and 3′ fragments of the URA3 gene from *C. utilis* [69]. The plasmid pGUSS2 for high-level production of α-amylase was constructed in the same manner using pURAL11 [69]. The plasmid was used for transformation after cutting at the *Bgl*II sites to stimulate integration.

II-4-3. Cultivation

For production of α-amylase, seed cultures, grown in 100 mL of yeast extract peptone dextrose (YPD) medium in a 500 mL conical flask for 72 hr at 30 °C, were inoculated in 1.5 L of modified SD medium (3% glucose, 0.5% $(NH_4)_2SO_4$, vitamin, and rare metals) in a 2.5 liter jar for fed-batch fermentation (Tokyo Rika Co., Ltd. Japan). The fermentor was operated for 60 hrs at 30 °C and the pH was maintained at 5.0 by adding NH_4OH. A concentrated medium containing glucose, $(NH_4)_2SO_4$, rare metals, and vitamins was successively fed to the growing cells to obtain a high cell density culture. The glucose concentration was kept between 1% and 5%.

II-4-4. Protein purification

Cells cultured in YPD medium were resuspended in 50 m*M* sodium acetate (pH 5.5) containing 0.3 mg/mL Zymolyase 100T (Seikagakukogyo Co., Ltd. Japan) and incubated at 37 °C for 30 min and then at 70 °C for 1 hr. Cells in suspension were disrupted by vortexing for 5 min with glass beads (diameter, 400-600 μm; Sigma Chemical Co., Ltd.). Cell debris was removed by centrifugation at 12,000 *g* for 10 min. The

soluble fractions were recovered for further purification.

The mutant GTHases were purified as previously described [70]. In brief, the culture supernatant was loaded on the Shodex AsahiPak ES502C anion exchange column (7.5 × 100 mm, Showa Denko K.K, Japan). The protein was eluted by forming a linear gradient of 50 m*M* sodium acetate buffer (pH 5.0) in the same buffer containing 1.0 *M* sodium chloride at a flow rate of 0.5 mL/min for 40 min.

II-4-5. Crystallization and data collection

Co-crystallization experiments were performed using a mixture of 15 mg/mL (0.25 m*M*) of GTHase and 5 m*M* of G3-Tre or G7. The mutant GTHases and G3-Tre/G7 were crystallized isomorphously with the wild-type enzyme at 4 °C by the hanging drop vapor diffusion method by using 0.6–1.1 *M* sodium citrate, 0.1 *M* HEPES at pH 7.5.

Diffraction data of wild-type and mutant KM1-GTHases were measured at beamlines BL40B2 and BL41XU by using SPring-8 (Hyogo, Japan). Reflections of wild-type and D252S and of E283Q and D252E were collected on Quantum 4R (Area Detector Systems Corporation, USA) and Mar165 (Rayonix/Mar, USA) CCD detectors at BL40B2 and BL41XU, respectively. Crystals were soaked in a solution containing 1.35 *M* sodium citrate, 0.1 *M* HEPES (pH 7.5) and 10% (*v*/*v*) glycerol and then flash-frozen in a nitrogen gas stream for data collection at 100 K. Datasets of the wild-type, E283Q with G3-Tre, E283Q with G7, D252S with G3-Tre, and D252E with G3-Tre were collected. The diffraction data were processed using the *HKL2000* suite of programs [47]. All of the crystals belonged to space group $P3_221$ with unit cell dimensions of $a = b = 78$ Å, $c = 282$ Å, and $\gamma = 120°$. Details of data collection statistics are shown in Table II-2.

II-4-6. Structure determination

Initial phase information was obtained by molecular replacement methods using the program *MOLREP* [48] with the previous structure, KM1-GTHase [32] (PDB ID: **1EH9**), as the search model. Each KM1-GTHase molecule was found in the asymmetric unit with a solvent content of 69%. The refinement process for each model was carried out by using the programs *CNS* 1.21 [44] and *REFMAC* 5.5 [49]. Modification of the model, the initial selection, and manual verification of water molecules were performed with the program *Coot* [50]. The refinement statistics are also given in Table II-2.

II-4-7. Enzymatic activity

Enzymatic activities of the wild-type and mutant GTHases were assayed as reported previously [25,31]. In brief, the KM-1 GTHase was incubated in 10 m*M* maltotriosyltrehalose in 50 m*M* sodium acetate buffer (pH 5.5) at 60 °C. After the reaction was stopped by heating at 100 °C for 5 min, the trehalose produced was analyzed by HPLC using a TSKgel Amide-80 column (4.6 × 250 mm, Tosoh, Japan). Trehalose was eluted by 72.5% acetonitrile-deionized water solution and detected with a differential refractometer (RID-6A, Shimadzu, Kyoto, Japan). One unit (U) of GTHase was defined as the amount of enzyme that would liberate 1 μmol of trehalose from maltotriosyltrehalose per min at 60 °C.

Chapter III
Structure determination of other proteins from thermophiles for application to industrial use

III-1. Introduction

In Chapter I and II, structure of α-amylases related to trehalose synthesis were described. For investigating structural features of thermostable proteins, more structures of thermostable proteins are required comprehensively. Such comprehensive structural analysis to accumulate knowledge is one of approach to apply thermostable structural features to processes of industrial biosynthesis. Especially, it is thought that the knowledge of structural features provides solutions to biomass in energy issue, bioreactor with proteins, and production of expensive chemical compounds. On the other hand, the properties of thermostability and the details of reaction mechanisms including hydrogen atoms can not been enough recognized. Therefore, structural information containing hydrogen atoms is very important for further investigation. There is another way which is an analysis combined structural information and computer simulation, such as molecular dynamics calculation (MD). The analysis will be helpful to design an engineered protein which has superior ability such as thermostability. For the purpose, accurate MD result is necessary, but it needs structural information of hydrogen atoms. However, X-ray protein crystallography is difficult to determine the positions of hydrogen atoms. If researchers want to determine them, ultrahigh-resolution X-ray result is required. Or, there is another approach, which is neutron crystallography. X-ray diffraction depends on electron number of atom; by

contrast, neutron diffraction depends on atomic nucleus. The feature of neutron crystallography enables to determine positions of hydrogen atoms. For obtaining structural information including hydrogen atoms, the author attempted structure determination with X-ray crystallography by using *D-Cha* (see Appendix A) and neutron crystallography. In addition, neutron crystallography requires large high-quality crystals, which needs different trials from crystallization for X-ray diffraction measurement. Therefore, in neutron diffraction experiments, trial of structure determination by using neutron did not conduct only on thermostable protein, but also on some proteins except thermostable ones, which are human immunodeficiency virus (HIV-1) protease [71] and porcine pancreatic elastase (PPE) [72].

In this chapter, the author will describe cases of structure determination and preliminary diffraction study by using X-ray and neutron diffraction measurements. One is the structure of the TTHA0281 protein which is a UPF0150-family protein from *Thermus thermophilus* HB8 [73]. The UPF0150 family (http://pfam.janelia.org/family/PF03681) in the Pfam database [74] is an uncharacterized protein family containing 281 bacterial, 44 archaeal and five virus proteins. In the *T. thermophilus* HB8 genome, five ORFs encoding the UPF0150-family proteins have been identified. Next, the preliminary diffraction studies are a novel chitinase from *Ralstonia* sp. A471 (Ra-ChiC) and ADP-ribose pyrophosphatase-I from *T. thermophilus* HB8 (Tt-ADPRase-I) complexed with a non-hydrolyzable ADP-ribose (ADPR) analogue α,β-methyleneadenosine diphosphoribose (AMPCPR) by using X-ray and neutron, respectively. Ra-ChiC hydrolyzes *N*-acetylglucosamine hexasaccharide [$(GlcNAc)_6$] and produces $(GlcNAc)_2$ + $(GlcNAc)_4$ and $(GlcNAc)_3$ + $(GlcNAc)_3$, resulting from the second and third glycosidic linkage being split from the non-reducing end, at almost the same concentrations.

Tt-ADPRase-I is one of the main Nudix proteins [75,76]. The ADPRases prevent the intracellular accumulation of ADPR by hydrolyzing it to AMP and 5'-phosphate, and they are widely distributed in many organisms. ADPRases can be classified into the categories of ADPRase-I and II by their specificity in binding to ADPR. After that, the author will discuss about structural features of thermostability with those proteins including described in the former chapters. The discussion will mention application of structural features for industrial use.

III-2. Results

III-2-1. Overall structure of the thermostable protein TTHA0281

TTHA0281 exists as a single-domain structure of 87 residues with two α-helices, α1 and α2, and three β-strands, β1, β2 and β3 (Fig. III-1a). This arrangement forms a three-stranded twisted antiparallel β-sheet flanked by α2, known as an α-β-β-β-α fold, which is also present in several double-helical nucleic acid-binding proteins, such as ribosomal protein S5 and RNase III [77-79]. The TTHA0281 protein formed a tetramer in the asymmetric unit of the crystal (Fig. III-1b). This is consistent with the fact that the TTHA0281 protein forms an oligomer in solution, although the molecular weight in solution is estimated to be slightly higher than that of the tetrameric state of this protein, as described above. The interactions between subunits *A* and *B* and between subunits *C* and *D* are stabilized by hydrogen-bonding interactions between the N atom of Thr8 in α1 and the side-chain O atom of Glu58 in α2. The amido O atoms of Lys44 in the loop between β3 and α2, the side-chain N atom of Lys47 in α2 and the side-chain hydroxyl group of Tyr29 in β2 of subunit *A* form hydrogen bonds to the side-chain

amido group of Arg19 in β1, the amido O atom of Glu12 in α1 and the N atom of Tyr20 in β1 of subunit *D*, respectively; furthermore, the side-chain N atom of Lys47 in α2, the side-chain hydroxyl group of Tyr29 in β2 and the N atom of Tyr20 in β1 of subunit B form hydrogen bonds to the amido O atom of Ala18 in the loop between α1 and β1, the side-chain carboxyl group of Glu21 in β1 and the side-chain hydroxyl group of Tyr29 in β2 of subunit *C*, respectively. The interaction surface areas at the interfaces of subunits *AB* and *AD*, calculated with the *AREAIMOL* program (Collaborative Computational Project, Number 4, 1994), are ~800 and ~600 Å^2, which are 15% and 11% of the total surface area of subunit *A*, respectively (Fig. III-1b). In the crystal, CHES molecules, which may be derived from the crystallization reagent, were found around

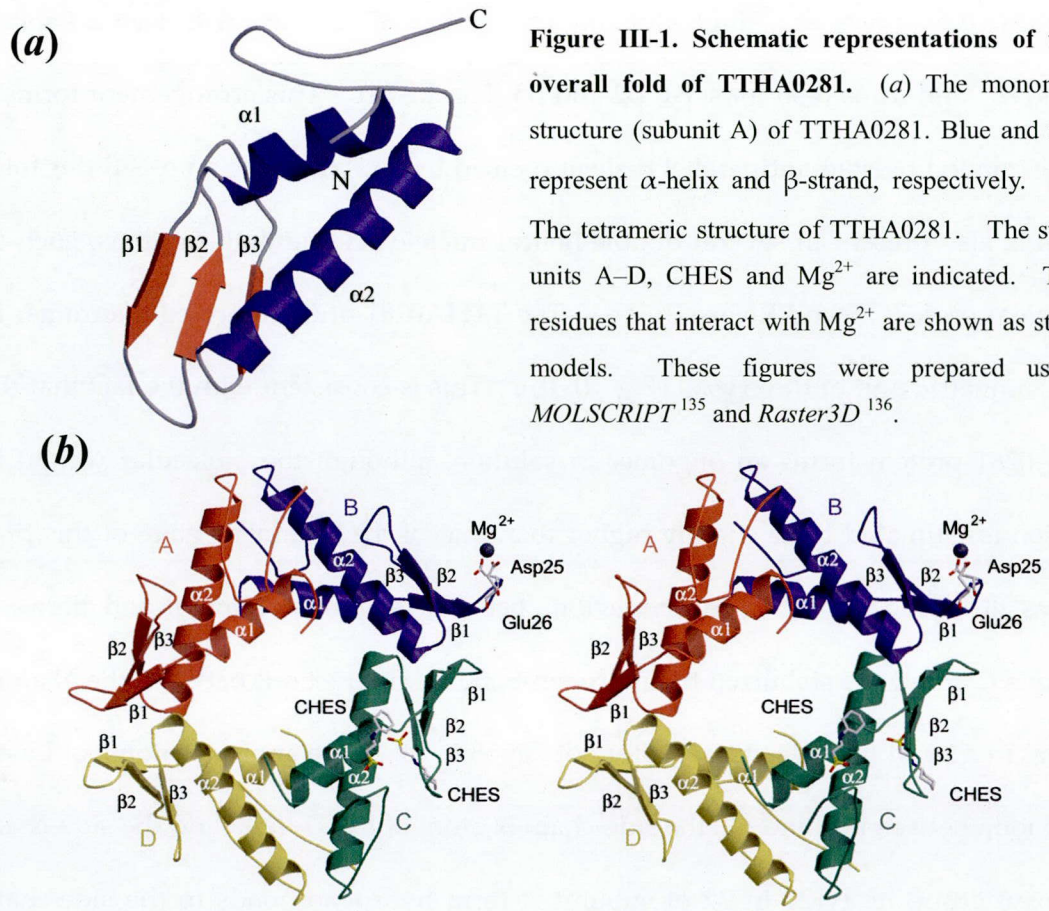


Figure III-1. Schematic representations of the overall fold of TTHA0281. (*a*) The monomer structure (subunit A) of TTHA0281. Blue and red represent α-helix and β-strand, respectively. (*b*) The tetrameric structure of TTHA0281. The subunits A–D, CHES and Mg^{2+} are indicated. The residues that interact with Mg^{2+} are shown as stick models. These figures were prepared using *MOLSCRIPT* [135] and *Raster3D* [136].

Glu13 and Trp40 of subunit C in the difference Fourier map, at 2σ and 4σ (Fig. III-1b). A magnesium ion was located around the side chains of Asp25 and Glu26 in subunit *B* (Fig. III-1b).

III−2−2. Crystallization and preliminary X-ray diffraction studies of the thermostable protein Ra-ChiC for structure determination

Crystallization conditions for Ra-ChiC$_{mature}$ (Fig. III-2b) were screened using the screening kits at 100 and 50% (CS, CS2, WZ I and WZ II) or 100, 67 and 33% (PS) of the initial concentrations of the precipitating agents. In addition, the protein was prepared at four different concentrations (20, 10, 6.5 and 3.3 mg ml^{-1}) for each crystallization screening. However, Ra-ChiC$_{mature}$ did not crystallize despite the screening of over 2000 crystallization conditions. The difficulty in crystallization of Ra-ChiC$_{mature}$ might derive from the flexibility of the protein, the linker of which includes seven consecutive glycines between the two domains. Therefore, the author constructed a deletion mutant, Ra-ChiC$_{89-252}$ (Fig. III-2c), from which the chitin-binding domain and con-

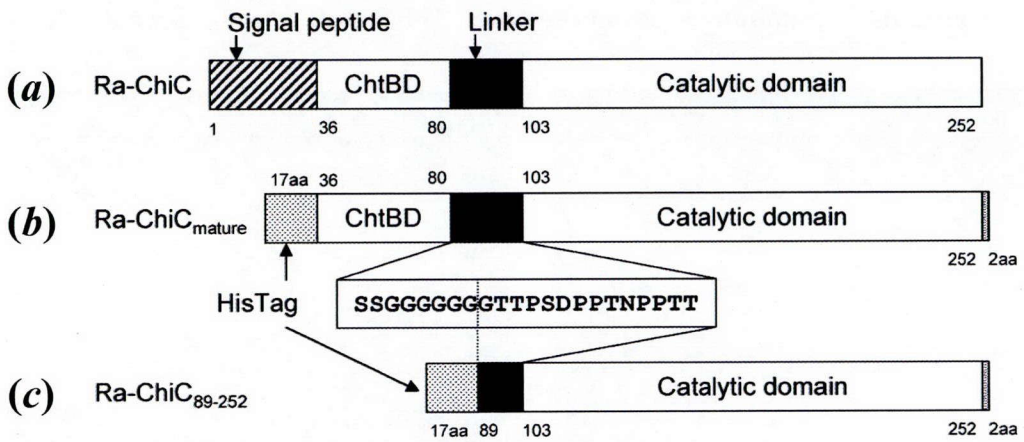


Figure III-2. Schematic representation of the Ra-ChiC domain structure. (*b*) and (*c*) are the recombinant constructs used in this study. The pCold I vector-derived amino acids correspond to the grey regions in (*b*) and (*c*).

secutive glycines were removed, for crystallization. Ra-ChiC$_{89-252}$ eluted as two independent peaks in cation-exchange chromatography. The two eluates were concentrated to 3.6 mg mL^{-1} (sample 1) and 2.8 mg mL^{-1} (sample 2). A total of 384 conditions were screened for crystallization of samples 1 and 2 using CS, CS2, WZ I, WZ II and PS. When using PS the original solution was also diluted to 33% for crystallization screening. After 10 d, small crystals had formed in 22 conditions. There was no significant difference in crystallization between samples 1 and 2. Two crystallization conditions, PS condition No. 44 at 33% concentration and PS condition No. 49 at 33% concentration, were optimized to produce crystals suitable for X-ray analysis. From the results of optimization, bipyramidal crystals with dimensions of over 0.1 mm were obtained using both precipitant solutions (Fig. III-3). Crystals A and B were obtained in 6% (*v*/*v*) PEG 400, 9% (*w*/*v*) PEG 1500, 30 m*M* HEPES pH 7.5 (PS condition No. 44 at 30% concentration) and 1.2% (*v*/*v*) 2-propanol, 7.5% (*w*/*v*) PEG 3350, 30 m*M* calcium chloride, 30 m*M* HEPES pH 7.5 (PS condition No. 49 at 30% concentration), respectively.

For X-ray diffraction measurements at 100 K, crystals A and B were soaked in reservoir solutions containing cryoprotectant [PEG 400 for crystal A and

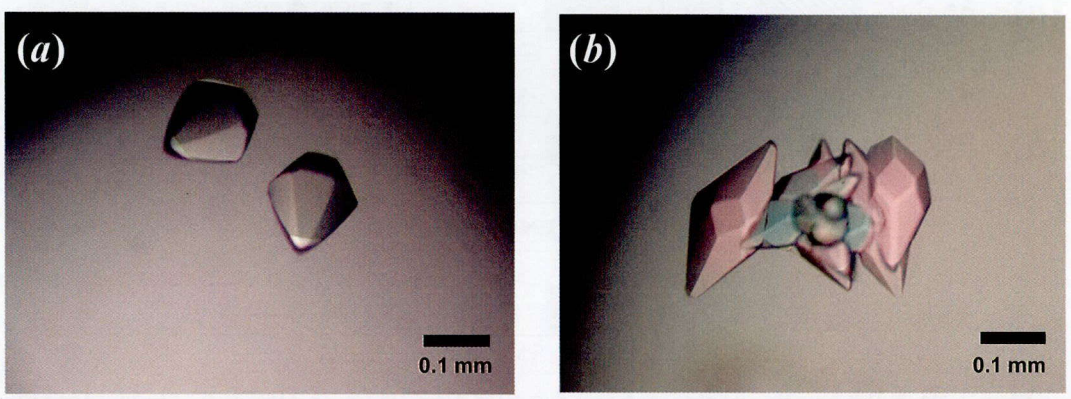


Figure III-3. Crystals of Ra-ChiC$_{89-252}$. (*a*) Crystal A from 6% (*v*/*v*) PEG 400, 9% (*w*/*v*) PEG 1500, 30 m*M* HEPES pH 7.5. (*b*) Crystal B from 1.5% (*v*/*v*) 2-propanol, 7.5% (*w*/v) PEG 3350, 30 m*M* calcium chloride, 30 m*M* HEPES pH 7.5.

2-methyl-2,4-pentanediol (MPD) for crystal B]. Crystal A was soaked in 28% (*v*/*v*) PEG 400, 9% (*w*/*v*) PEG 1500, 30 m*M* HEPES pH 7.5 and crystal B was soaked in 20% (*v*/*v*) MPD, 7.5% (*w*/*v*) PEG 3350, 30 m*M* calcium chloride and 30 m*M* HEPES pH 7.5. Both soaking experiments were carried out by increasing the concentration of the cryoprotectant in a stepwise fashion. Crystals A and B belonged to space group $P6_122$ or $P6_522$, with unit-cell parameters $a = b = 100$, $c = 243$ Å. The Matthews coefficient (V_M) was 3.2, 2.4 or 1.9 Å^3 Da^{-1}, assuming the presence of three, four or five molecules in the asymmetric unit, respectively. The diffraction data sets from crystals A and B were integrated and scaled to maximum resolutions of 1.90 and 1.85 Å, respectively. The data-collection statistics are summarized in Table III-1. Both crystals had sufficient diffraction quality for continued structure determination to elucidate the structure–function relationship of Ra-ChiC in detail. MR analysis was attempted using several sets of coordinates; however, no search model gave a significant solution for phase determination. Therefore, a search for heavy-atom derivatives for phase determination is under way using the isomorphous replacement and multiple/single anomalous dispersion (MAD/SAD) methods. In addition, selenomethionyl

Table III-1. Data-collection and reduction statistics for Ra-ChiC$_{89–252}$ crystals.

	Crystal A	Crystal B
Space group	$P6_122$ or $P6_522$	$P6_122$ or $P6_522$
Unit-cell parameters (Å)	$a = b = 100.0$, $c = 242.8$	$a = b = 99.7$, $c = 242.5$
Resolution (Å)	1.90	1.85
No. of measured reflections	653,979	898,217
No. of unique reflections	56,315	59,568
Redundancy	11.7 (6.7)	15.1 (6.0)
$I/\sigma(I)$	32.6 (2.7)	47.3 (3.2)
R_{merge} [†](%)	8.0 (43.8)	7.7 (40.5)
Completeness of data	97.8 (89.1)	96.7 (88.0)
Wilson plot B factor (Å^2)	24.9	26.3

Values in parentheses are for the highest resolution shell.

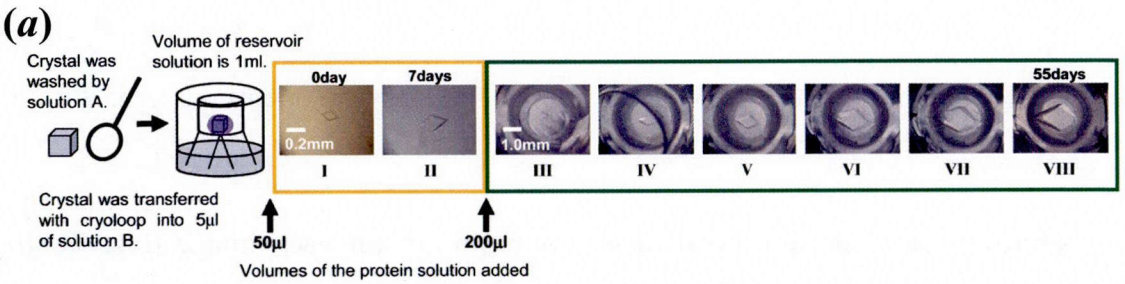
[†] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$

Ra-ChiC$_{89-252}$ is being prepared for SAD/MAD analysis.

III–2–3. Crystallization and preliminary neutron diffraction studies of the thermostable protein Tt-ADPRase-I for further investigation with X-ray/neutron combined structure determination

A neutron diffraction experiment requires a relatively larger crystal than an X-ray diffraction experiment because of the weak neutron diffraction signal. Although the initial crystallization attempts using a 5 μL crystallization volume and the hanging drop vapour diffusion method yielded a Tt-ADPRase-I crystal with a size of 1.0 × 1.0 × 0.4 mm, the crystal was still too small to obtain neutron diffraction signals in the BIX-3 diffractometer in JRR-3 research reactor at Japan Atomic Energy Agency (Ibaraki, Japan). Therefore, the author attempted to further grow the Tt-ADPRase-I crystal by increasing the crystallization volume in combination with the macro-seeding procedure. The initial seed crystals were prepared by mixing a 20 μL reservoir of solution A with a 20 μL protein solution containing 70 mg mL^{-1} of Tt-ADPRase-I. The resulting drops were equilibrated by vapour diffusion against the reservoir of solution A. After 30 days, the initial crystals, 0.2 × 0.2 × 0.2 mm in size, were obtained. The crystals with the clearest edges were selected and used as seeds.

After seeding, the growth of the crystal was monitored by measuring its length, as shown in Fig. III-4b. The labels I–VIII indicate the points at which the crystal length was observed, and the corresponding crystal images are shown in Fig. III-4a. After the initial seeding into 50 μL of solution (the period from point I to II), a crystallization drop that did not contain a cracked crystal or a side crystal was chosen, and 200 μL of protein solution was added. After a slow growth phase (four days from point II to III),



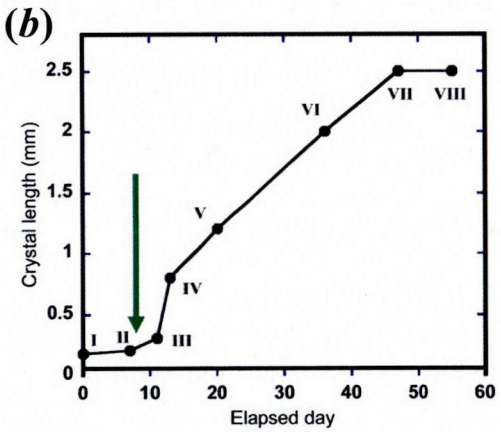


Figure III-4. Macro-seeding procedure and plot of crystal length vs. elapsed days. (*a*) Macro-seeding procedure. The orange and the green boxes indicate the pictures of the crystals after adding 50 μL and 200 μL of solution A containing protein, respectively. Numbers below pictures correspond to numbers in (*b*). (*b*) Plot of crystal length vs. elapsed days. The day when the 50 μL of solution A containing protein (28 mg mL^{-1}) was added is shown as the 0 day. The 200 μL of the same protein solution was added at the green arrow (elapsed 7 days).

followed by a short rapid phase (two days from point III to IV), the crystal consistently grew to 2.5 × 2.5 × 1.5 mm (from point IV to VII). The crystal growth seemed to stop after 50 days, that is, at point VII. After 55 days, the crystal was harvested at point VIII; by then, the crystal was 2.5 × 2.5 × 1.5 mm (Fig. III-5a), which was appropriate for the neutron diffraction experiment. Then, the crystal was transferred into deuterated mother liquor before the neutron diffraction experiment in order to reduce the back-

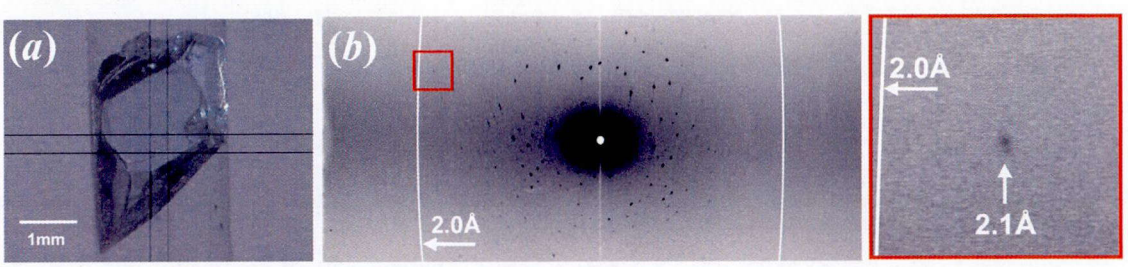


Figure III-5. Tt-ADPRase-I crystal and diffraction images. (*a*) The Tt-ADPRase-I crystal, approximate dimensions 2.5 × 2.5 × 1.5 mm (~10 mm^3), sealed into a quartz capillary (diameter of 2.5 mm). (*b*) An image of neutron diffraction spots taken from the Tt-ADPRase-I crystal. The area in the small red box on the left image is enlarged to the right.

ground noise caused by the hydrogen atoms. The crystal was soaked in a solution (final pD 5.3) containing both AMPCPR and $MgCl_2$ for 2 days to obtain the ternary complex.

Neutron data collection entailed 4 h of exposure for each image (Fig. III-5b). These diffraction images were integrated and scaled into 8082 unique reflections with an R_{merge} of 9.7%. The crystal belongs to the space group $P3_221$ with unit-cell parameters $a = b = 50.7$ and $c = 119$ Å. The data collection statistics are summarized in Table III-2. The X-ray diffraction dataset was collected with a Quantum 4R (Area Detector Systems Corporation, USA) charge coupled device (CCD) installed at the BL-6A beamline of the Photon Factory. The crystal diffracted to 1.5 Å resolution with a space group of $P3_221$ and unit-cell parameters of $a = b = 50.7$ and $c = 119$ Å at room temperature, as shown in Table III-2. The neutron dataset has sufficient quality to determine

Table III-2. Data collection statistics

	Neutron	X-ray
Collected at	JRR-3 BIX-3	KEK-PF BL-6A
Detector	Neutron Imaging Plate	Quantum 4R
Temperature (K)	293	293
Space group	$P3_221$	$P3_221$
Unit-cell parameters (Å)	$a = b = 50.7$, $c = 119$	$a = b = 50.7$, $c = 119$
Resolution range (Å)	43.9–2.10 (2.21–2.10)	29.5–1.50 (1.58–1.50)
No. of measured reflections	24,909	303,532
No. of unique reflections	8,023	28,885
Multiplicity	3.1 (2.4)	10.5 (10.7)
Mean $I / \sigma(I)$	18.6 (6.7)	41.0 (9.8)
R_{merge} †(%)	9.5 (30.4)	3.7 (35.3)
$R_{p.i.m.}$ ‡(%)	5.7 (21.2)	1.2 (11.2)
Average mosaicity (°)	1.11	0.15
Completeness	75.5 (62.8)	98.8 (99.8)
Wilson plot B factor (Å^2)	9.6	18.8

Values in parentheses are for the highest resolution shell.

†$R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where I is the intensity of a reflection and $\langle I \rangle$ is the average of the intensity.

‡$R_{p.i.m.} = \Sigma_{hkl} [1/(N-1)]^{1/2} \Sigma_j |I_j(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_j I_j(hkl)$

the location of hydrogen and deuterium, considering their scattering lengths [80].

The author has also attempted to reduce the incoherent back ground noise originating from residual hydrogen atoms in Tt-ADPRase-I. Because of the high yield of Tt-ADPRase-I in this *E. coli* expression system, perdeuteration of Tt-ADPRase-I using Bioexpress cell growth media (U-D, 98%, Cambridge Isotope Labs) was performed and yielded 20 mg of purified protein from 1 liter cultivation (data not shown). The perdeuteration of Tt-ADPRase-I as well as the reproducible preparation of large crystals will help us to examine the hydration structure of the active site of Tt-ADPRase-I as a structure before substrate recognition, and to elucidate the substrate or inhibitor interaction with Tt-ADPRase-I after soaking into the crystal. These neutron diffraction data will elucidate the catalytic mechanism of Tt-ADPRase-I by permitting observation of the important water molecules and hydrogen atoms involved in the catalysis.

III-3. Discussion

The author described crystal structures from thermophiles. Many structural features have been suggested for thermostability, such as hydrophobic interaction, hydrogen bonds, ion-pairs and solvent-accessible surface areas (SASAs). The author will discuss structural features from the solved structures and known structures by comparing SASAs and numbers of ion-pairs. Generally, decrease of SASA and increase of the fraction of buried hydrophobic atoms have been discussed as the stabilizing principles for thermostable proteins [3]. In this section, differences between thermophilic proteins which have been described in this thesis will be discussed. The discussion about thermostability will bring forward to the application of improving industrial use

of enzymes with thermostable features. In this discussion, comparison among α-amylase family proteins including GTSase (Chapter I) and GTHase (Chapter II) will be discussed. And the author will extend the insights of structural features from the knowledge of α-amylases to other protein (TTHA0281; described in Chapter III) and discuss about their thermostable features. After that, the trial of structural research of the thermostable proteins by using neutron diffraction study will be described. Neutron crystallography has some differences about crystallization and structural information from X-ray crystallography. The features of the method help improving the knowledge of thermostable features.

III–3–1. Comparison among α-amylase family proteins

The author described about enzymes from thermophiles, which involves in biosynthesis of trehalose, 5389-GTSase in Chapter I and KM1-GTHase in Chapter II. Those enzymes belong to α-amylase family. In this section, thermostability derived from the structural features of α-amylases will be discussed about structural comparison of 5389-GTSase and KM1-GTHase with other structures which are *S. acidocaldarius* MTSase [33] (Sa-MTSase; PDB ID: **1IV8**; 720 residues), *Thermoactinomyces vulgaris* R-47 α-amylase II [81] (TVAII; PDB ID: **1JIB**; 585 residues), *Deinococcus radiodurans* MTHase [64] (Dr-MTHase; PDB ID: **2BHY**; 602 residues), and *Aspergillus niger* α-amylase [36] (An-Amylase; PDB ID: **2AAA**; 484 residues). The comparison about SASA and ion-pair network will be discussed in below.

Table III-3 shows comparison of the SASAs of α-amylases from thermophilic organisms (KM1-GTHase, 5389-GTSase, Sa-MTSase, and TVAII), radioresistant bacterium (Dr-MTHase), and *Aspergillus* (An-Amylase). SASAs per residue of these α-amylase

proteins are about 35–40 Å^2, which are broadly similar to each other. On the other hand, SASAs classified by properties of residues show remarkable differences. Thermostable proteins, KM1-GTHase, 5389-GTSase, Sa-MTSase, and TVAII, have 20–25% of hydrophobic SASA, 25–30% of polar, and 50–55% of charged, approximately. Dr-MTHase has nearly fraction, which have 30% of hydrophobic SASA, 26% of polar, and 44% of charged. In contrast, An-Amylase is significantly different from the other proteins. It has 23% of hydrophobic SASA, 51% of polar, and 26% of charged. These proteins from the extremophilic bacteria (thermophilic and radioresistant) has a significantly high SASA ratio of charged residues in spite of no remarkably difference of

Table III-3. Comparison of solvent-accessible surface areas (SASAs)

	KM1-GTHase*		5389-GTSase*		Sa-MTSase	
Identity (%)	–		22		19	
R.m.s.d. (Å)	–		3.0		3.0	
Total SASA (Å^2 per residue)	20,918	(37.6)	28,550	(39.4)	26,635	(38.0)
SASA of hydrophobic residues (Å^2) (% of total)	3,858	(18.4)	5,088	(17.8)	5,304	(19.9)
SASA of polar residues (Å^2) (% of total)	5,433	(26.0)	7,244	(25.4)	7,593	(28.5)
SASA of charged residues (Å^2) (% of total)	11,627	(55.6)	16,217	(56.8)	13,737	(51.6)
No. of residues in crystal structure	557		725		701	
No. (% of total) of hydrophobic residues	244	(43.8)	306	(42.2)	300	(42.8)
No. (% of total) of polar residues	154	(27.7)	195	(26.9)	206	(29.4)
No. (% of total) of charged residues	159	(28.5)	224	(30.9)	195	(27.8)

	TVAII		Dr-MTHase		An-Amylase	
Identity (%)	22		36		18	
R.m.s.d. (Å)	2.5		1.7		2.7	
Total SASA (Å^2 per residue)	21,272	(36.4)	20,337	(35.1)	16,564	(34.8)
SASA of hydrophobic residues (Å^2) (% of total)	5,197	(24.4)	6,145	(30.2)	3,823	(23.1)
SASA of polar residues (Å^2) (% of total)	5,668	(26.7)	5,249	(25.8)	8,456	(51.0)
SASA of charged residues (Å^2) (% of total)	10,406	(48.9)	8,942	(44.0)	4,284	(25.9)
No. of residues in crystal structure	585		580		476	
No. (% of total) of hydrophobic residues	277	(47.4)	290	(50.0)	209	(43.9)
No. (% of total) of polar residues	153	(26.1)	158	(27.2)	188	(39.5)
No. (% of total) of charged residues	155	(26.5)	132	(22.8)	79	(16.6)

SASAs are calculated by program *AREAIMOL*. The author defined that hydrophobic residues are Ala, Val, Phe, Pro, Ile, Leu, Met and Gly, polar residues are Ser, Thr, Tyr, Cys, Asn, Gln, His and Trp, and charged residues are Asp, Glu, Lys and Arg. Values in parentheses indicates ratio to number of residues.

*) Structures of these enzymes have been analyzed in this thesis.

fractions of hydrophobic, polar, and charged residues. This indicates that these proteins expose less hydrophobic and more charged residues to the solvent region.

Recent structural studies on hyper-thermophilic proteins reveal an increase the number of ion-pairs and the ion-pair networks from mesophilic ones. Following that, Table III-4 shows comparison of ion-pairs among α-amylase family proteins. The total numbers of ion-pairs and the numbers of intra-, inter-helical ion-pairs, and numbers of ion-pair networks formed by over four residues [4] are shown in Table III-4. The criteria of counting ion-pairs and networks are following. 1) Secondary structures are specified by using the program *DSSP* [82]. 2) Helices are defined α-helix in *DSSP* criteria. 3) The distances of ion-pairs are calculated by using the program *NCONT* [54]. And those are identified using a cutoff of distance between oppositely charged residues of

Table III-4. Comparison of the number of ion-pairs

	KM1-GTHase*		5389-GTSase*		Sa-MTSase	
5 Å cutoff total (ion-pairs per residue)	58	(10.4)	87	(12.0)	71	(10.1)
Intra-helix	5	(0.9)	11	(1.5)	10	(1.4)
Inter-helix	25	(4.5)	43	(5.9)	32	(4.6)
Network ≥ 4 residues (networks per residue)	6	(1.1)	12	(1.7)	9	(1.3)
4 Å cutoff total (ion-pairs per residue)	43	(7.8)	58	(8.0)	53	(7.5)
Intra-helix	5	(0.9)	5	(0.7)	7	(1.0)
Inter-helix	18	(3.2)	32	(4.4)	24	(3.4)
Network ≥ 4 residues (networks per residue)	3	(0.5)	2	(0.3)	4	(0.6)

	TVAII		Dr-MTHase		An-Amylase	
5 Å cutoff total (ion-pairs per residue)	50	(8.5)	37	(6.4)	17	(3.6)
Intra-helix	13	(2.2)	4	(0.7)	1	(0.2)
Inter-helix	18	(3.1)	16	(2.8)	4	(0.8)
Network ≥ 4 residues (networks per residue)	6	(1.0)	4	(0.7)	1	(0.2)
4 Å cutoff total (ion-pairs per residue)	37	(6.3)	29	(4.9)	14	(3.0)
Intra-helix	9	(1.5)	1	(0.2)	1	(0.2)
Inter-helix	15	(2.6)	12	(2.0)	4	(0.8)
Network ≥ 4 residues (networks per residue)	5	(0.9)	3	(0.5)	1	(0.2)

This table values are calculated by *DSSP* and *NCONT*. The author defined that positive residues are Arg and Lys, and negative residues are Glu and Asp. Values in parentheses indicates ratio to number of residues.

4.0 [83] and 5.0 Å. 4) The networks over four residues means that number of residues involved in the network are four residues or more. The extremophilic proteins (proteins excluded An-Amylase) have high ratios of ion-pairs or networks per residue. Especially, proteins from *Sulfolobus* have remarkably high ratios of inter-helical ion-pairs than TVAII and Dr-MTHase. This may indicate that inter-helical ion-pairs contribute to higher thermostability. The result of inter- and intra-helical ion-pairs follows the result of *Pyrococcus kodakaraensis* KOD1 [84].

III-3-2. Expanding structural features to other thermostable protein

As described above, there are distinguished features in SASAs and ion-pair networks in the α-amylase family proteins. Those discussions are about proteins which are members of a same family, α-amylases. Here, the author makes an attempt to expand the discussed features to a protein which belongs to other family. The protein is a UPF0150-family protein TTHA0281 from *T. thermophilus* HB8, which has been described in this chapter (PDB ID **2DSY**; 87 residues [73]). Features, SASA and ion-pair networks, of the thermostable protein TTHA0281, will be discussed.

In SASA of the TTHA0281, the ratios of hydrophobic, polar, and charged residues are 53.8%, 18.7%, and 27.5%, respectively (Table III-5). This property is similar to the former results in the discussion. In the TTHA0281 protein, the largest SASA is the area of the charged residues (50.1% of total SASA). This result follows the comparison results of thermostable proteins in discussion as above. Moreover, the numbers of ion-pairs of TTHA0281 are shown in Table III-6.

Table III-5. Comparison of solvent-accessible surface area (SASA)

	TTHA0281*	
Total SASA (Å^2 per residue)	4,004	(50.1)
SASA of hydrophobic residues (Å^2) (% of total)	1,416	(35.4)
SASA of polar residues (Å^2) (% of total)	581	(14.5)
SASA of charged residues (Å^2) (% of total)	2,005	(50.1)
No. of residues in crystal structure	80	
No. (% of total) of hydrophobic residues	43	(53.8)
No. (% of total) of polar residues	15	(18.7)
No. (% of total) of charged residues	22	(27.5)

Table III-6. Comparison of the number of ion-pairs

	TTHA0281*	
5 Å cutoff total (ion-pairs per residue)	4	(4.6)
Intra-helix	0	(0.0)
Inter-helix	2	(2.3)
Network ≥ 4 residues (networks per residue)	0	(0.0)
4 Å cutoff total (ion-pairs per residue)	3	(3.4)
Intra-helix	0	(0.0)
Inter-helix	1	(1.1)
Network ≥ 4 residues (networks per residue)	0	(0.0)

These results discussed as above show the features of ratio of SASA and ion-pairs in thermostable proteins. It is essential for introducing the features to enzymes in industrial use to expand the knowledge of structural features of the proteins (i.e. α-amylase family) to other thermostable proteins. The ratio of hydrophobic surface is the major area in proteins discussed as above. This comparison demonstrates that SASAs of the thermostable proteins are very different from mesophilic one. Charged ratio of SASA among each thermophilic protein is the largest area, which are 50% to 60% of total SASA. This trend is similar to each other even though there are not so high identities (~30%). It is remarkable that thermostable proteins expose less hydrophobic and more charged residues to its surface. The effects of the electrostatic contribution of the charged residues were reported [85]. According to the report, the charged residues contribute to protect their protein from temperature changes. The comparison results as above follow the feature, and charged-residue-rich surface seems

to make its insensitive property higher than non-thermostable one. In contrast, the ratios of the hydrophobic surface areas are decreased against the ratios of the hydrophobic residues in thermostable proteins. The ratios of the hydrophobic SASA in thermostable proteins are smaller than the charged one. This shows that the hydrophobic residues located inside of the protein. About hydrophobic residue, there is a report which describes that cavity-filling mutation of hydrophobic core increases thermostability of itself [86]. The report shows that filling internal cavities of protein have an advantage in higher temperature. According to the comparison results of the proteins, thermostable proteins have smaller hydrophobic SASA than charged one. In contrast, hydrophobic SASA and charged one are similar size on non-thermostable An-Amylase. In other words, hydrophobic residues are packed in the internal of protein in thermostable proteins for increasing thermostability. The hydrophobic residues which are located inside of the proteins interact to each other and can contribute to forming the structure rigidly. This rigid structure increases its thermostability. As just described, the charged residues and the hydrophobic ones contribute thermostability of the protein.

In addition, the numbers of ion-pairs in thermophilic proteins are more than non-thermophilic one. These ion-pairs are considered to contribute increasing thermostability. It has been reported that the helical conformation is stabilized by [i+4] or [i+3] glutamate–lysine intra-helical ion-pairs in a short model peptides system [87]. However, the thermostable proteins do not always have larger numbers of intra-helical ion-pairs than non-thermostable one. In the comparison results (Table III-4), Dr-MTHase has only one intra-helical ion-pair (4Å cutoff), which is same number to non-thermostable An-Amylase. Especially, the TTHA0281 protein has no intra-helical

ion-pair (Table III-6). On the other hand, inter-helical ion-pairs are seen among the thermostable proteins, which are larger numbers than non-thermostable An-Amylase. From the things, it is considerable that inter-helical interactions have grater impact on thermostability of protein than intra-helical ones. Considerable reasons are following. 1) Intra-helical ion-pairs maintain mainly secondary structure. This may affect in shorter range than overall structure. 2) Inter-helical ion-pairs involve in forming overall structures of proteins. Therefore, it is possible that forming inter-helical ion-pairs has higher priority to obtain thermostability than intra-helical ones. This hypothesis indicates that putting inter-helical ion-pairs in can be more effective to increase thermostability.

III–3–3. Further investigation on thermostable proteins

For further investigation, more cases of structure determination on thermostable proteins. The author attempted crystallization and preliminary X-ray diffraction studies of thermostable protein Ra-ChiC [88]. The Ra-ChiC involves in degrading of chitin, which is useful for potential applications as a utilization of biomass. This result can contribute to the accumulation of structural information has been proceeding.

Structural information obtained by using neutron diffraction has also advantage about determination of hydrogen atom position [89]. Those structural information including hydrogen atoms makes computer simulation, such as MD, accurate and it is possible to apply on investigation of thermostable features. The author aims to apply thermostable features to enzymes in industrial use (for example, biosynthesis). Simulation results based on structural information can lead such protein engineering to

improving productivity. The author has carried out structure determination by using combination of X-ray and neutron diffraction on some proteins [71,72,89]. However, obtaining crystals for neutron diffraction measurements are more difficult work for X-ray ones. The author attempted crystallization on standard samples, which are HIV-1 protease and elastase. In consequence, although the proteins have no endurance ability at high temperature, which were determined those structures with neutron crystallography. Structural information by using neutron crystallography which combined with computer simulation has been used for molecular design of inhibitor in the case of the HIV-1 protease [71]. According to the HIV-1 protease, using neutron crystallography and structural information including hydrogen atoms is able to design efficient molecules related to reactions. This is a model case of demonstrating advantage of neutron crystallography. Then, the author attempted structure determination on thermostable protein. Using neutron crystallography, obtaining crystals suitable for neutron diffraction measurement is tough work. However, the author has carried out crystallization and preliminary neutron diffraction study of a thermostable protein [15]. This result will lead structural analysis with neutron crystallography to elucidation of details of structural features on thermostable proteins. Therefore, structural information with neutron crystallography will realize application of thermostability to proteins which involves in industrial convenient reaction by using protein engineering in future.

The accumulation of the knowledge described above will make a leap of applying biosynthesis to industrial productions. It is thought that the use of glycosidases, as described in Chapter I and II, is clearly effective. Trehalose production enzymes belong to α-amylase family. Variety substrates are targets of α-amylases, which are

mainly sugars as useful products for our lives. In addition, α-amylases have small differences of catalytic centers each other, which are derived from similar catalytic mechanisms. Hence, it can be relatively realistic to apply protein engineering in the family. Actually, the author introduced the mutation for changing anomeric configuration of the reaction in KM1-GTHase [90]. In this way, the accumulation of the results of the comprehensive structure analysis will be expected to expand for protein engineering in industrial use.

III-4. Materials and methods

III–4–1. Preparation, crystallization, and structure determination of the TTHA0281 protein

III-4-1-1. Preparation and gel-filtration analysis

The TTHA0281 gene was cloned under the control of the T7 promoter on the pET-11a expression vector (Novagen, Madison, WI, USA). In order to obtain the selenomethionine-labeled TTHA0281 protein (Se-TTHA0281), the *E. coli* methionine-auxotroph strain Rosetta834 (DE3), which the author obtained by introducing the pRARE plasmid carrying the tRNA genes for the codons AUA, AGG, AGA, CUA, CCC and GGA (Novagen) into the B834 (DE3) strain (Novagen), was used as the host. The recombinant strain was cultured in minimal medium containing 25 μg ml^{-1} L-selenomethionine, 30 μg mL^{-1} chloramphenicol and 50 μg mL^{-1} ampicillin. When an A_{600} of 0.7 was attained, isopropyl β-D-thiogalactopyranoside was added to a final concentration of 1 m*M* and the culture was incubated at 310 K for a further 5 h. The cells were collected by centrifugation and were disrupted by sonication in 20 m*M* Tris–HCl

buffer pH 8.0 containing 50 m*M* NaCl. The soluble fraction obtained after centrifugation at 15 000*g* for 20 min was heated at 343 K for 30 min. After the denatured proteins had been removed by centrifugation at 15 000*g* for 20 min, the sample was applied onto a Resource ISO column (GE Healthcare Bioscience Corp., Piscataway, NJ, USA) pre-equilibrated with 50 m*M* sodium phosphate buffer pH 7.0 containing 1.5 *M* $(NH_4)_2SO_4$. The bound proteins were eluted with a linear gradient of 1.5–0 *M* $(NH_4)_2SO_4$. The fractions containing Se-TTHA0281 were collected and applied onto a Resource Q column (GE Healthcare Bioscience Corp.), which was eluted with a linear gradient of 0–1 *M* NaCl. The fractions containing Se-TTHA0281 were collected and applied onto a BioScale CHT5-I hydroxyapatite column (Bio-Rad Laboratories Inc., Hercules, CA, USA) pre-equilibrated with 10 m*M* sodium phosphate buffer pH 7.0. The flowthrough fraction was collected, concentrated and then applied onto a HiLoad 16/60 Superdex 75 pg column (GE Healthcare Bioscience Corp.) pre-equilibrated with 20 m*M* Tris–HCl buffer pH 8.0 containing 0.15 *M* NaCl. Finally, the purified protein was applied onto a HiPrep 26/10 Desalting column (GE Healthcare Bioscience Corp.) pre-equilibrated with 20 m*M* Tris–HCl buffer pH 8.0. The sample was concentrated to 7.9 mg mL^{-1} with a Vivaspin 6 concentrator (5 kDa molecular-weight cutoff, Sartorius AG, Goettingen, Germany).

The TTHA0281 protein and molecular-weight standards were applied onto a Superdex 200 3.2/30 (GE Healthcare Bioscience Corp.) column, using HPLC (model 1100, Agilent Technologies, Palo Alto, CA) at a flow rate of 50 µL min^{-1}. The elution buffer was 20 m*M* Tris–HCl buffer pH 8.0 containing 1 m*M* DTT.

III-4-1-2. Crystallization

Crystals of the Se-TTHA0281 protein were obtained by the microbatch method [91] at 291 K using a *TERA* crystallization robot (Takeda Rika Kogyo Co. Ltd, Tokyo, Japan). A 0.5 μL aliquot of crystallization reagent, consisting of 17.5%(*w*/*w*) PEG4000, 0.1 *M* CHES buffer pH 9.4 containing 50 m*M* $MgCl_2$, was mixed with 0.5 μL 7.9 mg ml^{-1} protein solution and the mixture was then covered with 15 μL silicone and paraffin oil. Octahedral crystals grew in ten weeks to approximate dimensions of 0.2 × 0.15 × 0.15 mm.

III-4-1-3. X-ray diffraction and structure determination

Data were collected at the RIKEN Structural Genomics Beamline II (BL26B2) [11] at SPring-8 (Hyogo, Japan). The crystals were mounted in a nylon loop and cooled in an N_2-gas stream at 100 K using the *SPring-8 Precise Automatic Cryo-sample Exchanger* (*SPACE*), which was controlled using the *beamline-scheduling software* (*BSS*) [10,13]. Multiple-wavelength anomalous dispersion (MAD) data were collected utilizing the anomalous scattering from Se atoms. The data sets were collected to 1.9 Å resolution using a Jupiter210 CCD detector (Rigaku MSC Co., Tokyo, Japan). The data were processed using the *HKL2000* program suite [47]. The atomic positions of the Se atoms in the unit cell were determined using the program *SOLVE* [92] and density modification was then performed using the program *RESOLVE* [93]. The automatic tracing procedure in the program *ARP/wARP* [94] was utilized to build the initial model. Model refinement was carried out using the programs *REFMAC* [49] and *Coot* [50] from the *CCP4* program suite [54]. Initial picking and manual verifying of water molecules were performed with the program *Coot*. According to *PROCHECK* [95], the final model has

93.8% of the residues in the most favored conformation of the Ramachandran plot and no residues in the disallowed regions. Data-collection statistics and processed data statistics are shown in Table III-7.

Table III-7. Summary of data-collection and refinement statistics.

	Peak	Edge	Remote
Wavelength (Å)	0.9791	0.9794	0.9815
Space group	$P2_1$		
Unit cell parameters	a = 38.00 Å, b = 62.74 Å, c = 65.72 Å, β = 104.3°		
Resolution[a] (Å)	1.9 (1.97-1.90)	1.9 (1.97-1.90)	1.9 (1.97-1.90)
No. of measured reflections	89,118	88,350	86,333
No. of unique reflections	24,539	24,536	24,497
R_{merge}[†] (%)	6.9 (26.3)	7.0 (33.5)	5.7 (40.4)
Completeness of data (%)	100 (99.9)	99.9 (99.5)	99.2 (93.1)
I / σ(I)	10.9 (5.2)	10.8 (4.0)	12.1 (2.8)
R-factor[‡] (%)	18.5		
R_{free}[‡] (%)	23.2		
No. of protein atoms	2,495		
No. of water molecules	174		
RMSD bonds (Å)	0.015		
RMSD angles (°)	1.487		

Values in parentheses are for the highest resolution shell.

[†] $R_{merge} = \Sigma|I - \langle I \rangle| / \Sigma I$.

[‡] R-factor and $R_{free} = \Sigma||F_o| - |F_c|| / \Sigma|F_o|$, where the free reflections (5% of total used) were held aside for calculation of R_{free} throughout refinement.

III–4–2. Preparation, crystallization, and X-ray diffraction measurement of the Ra-ChiC

III-4-2-1. Expression and purification

Mature Ra-ChiC with an N-terminal hexahistidine tag (Ra-ChiC$_{mature}$; Fig. III-2b) was expressed in *Escherichia coli* as a soluble protein and purified using a HisTrap HP column [96]. Ra-ChiC mature was further purified for crystallization. The eluted fraction containing protein from the HisTrap HP column was desalted by gel filtration using a PD10 column (10 × 50 mm; GE Healthcare) equilibrated with 20 m*M* Tris–HCl

buffer pH 8.0. The desalted solution containing Ra-ChiC$_{mature}$ was applied onto a RESOURCE S cation-exchange column (1.0 ml; GE Healthcare). After the column had been washed with 5 ml 20 m*M* Tris–HCl buffer pH 8.0, Ra-ChiC$_{mature}$ was eluted with a linear gradient from 20 m*M* Tris–HCl buffer pH 8.0 to the same buffer containing 250 m*M* sodium chloride at a flow rate of 1.0 ml min^{-1} over 30 min. The purity of the recombinant Ra-ChiC$_{mature}$ was verified by SDS–PAGE analysis.

The expression vector for the catalytic domain with Pro/Thr-rich repeats in the interdomain linker, which includes residues 89–252 (Ra-ChiC$_{89-252}$; Fig. III-2c), was constructed by PCR using a Ra-ChiC DNA clone as template. The forward (5'-CGC **CAT ATG** GGA ACT ACA CCA TCC GAT-3'; the *Nde*I site is shown in bold) and reverse primers (5'-TGC **TCT AGA** ATA TCT GCC GCC ATA GCT CTT-3'; the *Xba*I site is shown in bold) for PCR were synthesized for the regions corresponding to amino-acid residues 89–94 and 246–252 of Ra-ChiC$_{89-252}$, respectively. PCR was conducted according to the method described by Ueda *et al.* [96]. A 0.5 kb DNA fragment obtained by PCR was cloned into pCR II-Blunt-Topo vector (Invitrogen). The nucleotides of the amplified fragments were confirmed by sequencing after ligation. The pCR II-Blunt-Topo vector containing the DNA fragment was treated with *Nde*I and *Xba*I and the fragment was recovered by agarose-gel electrophoresis. The DNA fragment and pCold I vector, which had been digested by *Nde*I and *Xba*I, were mixed and ligated with T4 DNA ligase. Ra-ChiC$_{89-252}$ was expressed and purified in *E. coli* in the same manner as used for Ra-ChiC$_{mature}$.

Crystallization

Initial crystallization conditions were screened by the sitting-drop va-

pour-diffusion method using Crystal Screen (CS) and Crystal Screen 2 (CS2) from Hampton Research (California, USA) and Wizard I (WZ I), Wizard II (WZ II) and Precipitant Synergy (PS) from Emerald BioSystems (Boston, USA) at 293 K. The screens were performed using 96-well Intelli-Plates (Art Robbins Instruments, USA) and a Hydra II Plus One (Thermo Fisher Scientific, USA). Drops were prepared by mixing 0.3 μL reservoir solution with 0.3 μL protein solution and were equilibrated by vapour-diffusion against the reservoir solution. The crystallization conditions for diffraction studies were optimized by the hanging-drop vapour-diffusion method, mixing 2.0 μL reservoir solution with 2.0 μL protein solution.

X-ray diffraction and molecular replacement

X-ray diffraction data from a crystal of the catalytic domain of Ra-ChiC were collected in a cold nitrogen-gas stream at 100 K using an MX-225HE system (Rayonix, USA) and synchrotron radiation (1.000 Å wavelength) on beamline BL41XU at SPring-8 (Hyogo, Japan). The oscillation angle was 1.0° and the exposure time was 1.0 s per frame. A total of 180 diffraction images were recorded at a camera distance of 170 mm and were processed using *HKL2000* [47]. Molecular-replacement (MR) analysis was performed using the program *Phaser* [97] using the coordinates of goose lysozyme (PDB entry **154L**) [98], which has 16% sequence identity to Ra-ChiC$_{89-252}$. The partial structure of bovine annexin VI (PDB ID: **1AVC**) [99] and the monozinc carbapenemase CphA from *Aeromonas hydrophila* (PDB ID: **1X8G**) [100] were also selected as search models for MR analysis from a protein *BLAST* search [101].

III–4–3. Preparation, crystallization, and neutron/X-ray diffraction measurement of the Tt-ADPRase-I

III-4-3-1. Expression and purification

Tt-ADPRase-I was expressed and purified as described previously [102], with minor modifications. Briefly, the expression plasmid (pET-11a) containing the gene encoding Tt-ADPRase-I was used to transform the *Escherichia coli* strain BL21 (DE3) (Merck). The transformant was cultured at 310 K in Luria-Bertani medium containing 50 mg ml^{-1} ampicillin, and Tt-ADPRase-I was overproduced without adding IPTG.

The complete purification procedure was performed at room temperature. After the cultured cells were suspended in 20 m*M* Tris-HCl (pH 8.0) containing 5 m*M* 2-mercaptoethanol and 50 m*M* NaCl, the cells were disrupted by sonication. The cell lysate was incubated at 343 K for 20 min, kept on ice, and then centrifuged (12 000 *g*) for 20 min at 277 K. Ammonium sulfate was added to the resulting supernatant to a final concentration of 1.35 *M*. The solution was applied to a Toyopearl Phenyl-650M column (25 × 100 mm, TOSOH, Japan), equilibrated with 50 m*M* sodium phosphate (pH 7.0) containing 1.35 *M* ammonium sulfate. The proteins were eluted with a linear gradient from 1.35 to 0 *M* ammonium sulfate in 50 m*M* sodium phosphate (pH 7.0). Fractions containing the target protein were collected, desalted by dialysis against 20 m*M* Tris-HCl (pH 8.0), and applied to a HiLoad 26/10 Q Sepharose HP column (GE Healthcare, UK). Proteins were eluted with a linear gradient from 0 to 500 m*M* NaCl in 20 m*M* Tris-HCl (pH 8.0). Fractions containing the Tt-ADPRase-I protein were then dialyzed against the 20 m*M* Tris-HCl buffer (pH 8.0). The protein solution was concentrated to 75 mg mL^{-1} and stored at 277 K. From 1 liter cultivation, 120 mg of Tt-ADPRase-I was purified, finally.

III-4-3-2. Crystallization

The initial seed crystals were obtained by sitting drop vapour diffusion method. The drops were prepared by mixing a 20 μL reservoir with 20 μL of 35 mg mL^{-1} Tt-ADPRase-I in 20 m*M* Tris–HCl buffer (pH 8.0), and were equilibrated against the reservoir solution containing 18% (*w*/*v*) PEG4000, 0.1 *M* sodium acetate buffer (pH 5.3), 20% (*w*/*v*) glycerol, and 0.2 *M* ammonium sulfate at 293 K.

Macro seeding was performed using the seed crystals washed with the solution A containing 18% (*w*/*v*) PEG4000, 0.05 *M* sodium acetate buffer (pH 5.3), 20% (*w*/*v*) glycerol, and 0.2 *M* ammonium sulfate and moved it to 5 μL of solution B (26% (*w*/*v*) PEG4000 instead of 18% in solution A) with a cryoloop. After keeping the washed seed crystals for a day in a 50 μL drop of solution B, 50 μL of solution A containing 28 mg mL^{-1} protein was added to the drop. After 7 days, another 200 μL of the same protein solution were added to the drop and kept for 55 days to obtain a large crystal (2.5 × 2.5 × 1.5 mm). This process is shown in Figure III-4a.

The solvent exchange of the crystal prior to neutron diffraction experiment was conducted by performing the vapour diffusion method. The large crystal in 250 μL of crystallization solution was vapour diffused against 1000 μL of the 50 m*M* D-substituted sodium acetate buffer (pD 5.3) containing 20% (*v*/*v*) D-substituted glycerol, and 0.2 *M* $(ND_4)_2SO_4$ in D_2O for 7 days. Then, the crystal was transferred into 250 μL of the same buffer (50 m*M* D-substituted sodium acetate buffer (pD 5.3) containing 20% (*v*/*v*) D-substituted glycerol, and 0.2 *M* $(ND_4)_2SO_4$) and soaked for 21 days.

The crystal of a ternary complex comprising *Tt*ADPRase-I, AMPCPR, and Mg^{2+} was obtained by soaking the crystal into 50 m*M* D-substituted sodium acetate buffer

(pD 5.3) containing 20% (*v*/*v*) D-substituted glycerol, 0.2 *M* $(ND_4)_2SO_4$, 71 m*M* AMPCPR (10 times excess amount of Tt-ADPRase-I) and 80 m*M* magnesium chloride.

III-4-3-3. Neutron and X-ray diffraction

The crystal thus obtained was sealed in a quartz capillary (Fig. III-5a) and used in both the neutron and X-ray diffraction experiments. The neutron diffraction data set was collected at room temperature with a monochromatic neutron beam (λ = 2.9 Å) on the BIX-3 diffractometer, which was installed at the 1G-A port of the JRR-3 research reactor of the Japan Atomic Energy Agency [103], and diffraction signals were recorded on a Neutron Imaging Plate. The diffraction data were collected to 2.1 Å resolution (Fig. III-5b) using the step-scanning method, and they consisted of 285 still images covering 85.5° with an interval angle of 0.3°. The exposure was 4 h per frame. The neutron diffraction data from the ternary complex of Tt-ADPRase-I with AMPCPR and Mg^{2+} were processed by the *DENZO* and *SCALEPACK* programs [47]. Additional X-ray diffraction data were collected with the same crystal at room temperature at the BL-6A beamline of the Photon Factory (KEK, Japan) for joint refinement [104]. These data were processed using the *HKL2000* suite of programs [47].

General discussion and conclusion

In this thesis, crystal structures of thermostable proteins, and further investigation with the newly developed system *D-Cha* and neutron crystallography were discussed. Protein structural information is more important, today. The author approached the features of thermostable proteins with structural analysis. To determine more numbers of protein structures, the author used the newly developed measurement system *D-Cha*. And to recognize more details of thermostability and reaction mechanism, the author attempted structure determination including hydrogen atoms of thermostable protein with neutron crystallography.

In Chapter I and II, the author described two proteins from thermophilic archaeon *Sulfolobus*, which are GTSase (glycosyltrehalose synthase) and GTHase (glycosyltrehalose trehalohydrolase). Those proteins involve in synthesis of trehalose from starch. Trehalose is widely used in industrial use, whose cost of production has been reduced with biosynthesis by the thermophilic enzymes. This is a major example of utilization of thermostable proteins for industrial production. The author figures out the structural features of thermostability, recognition, and reaction mechanisms with crystal structures. GTSase and GTHase have structures that are thought as stabilizing them, respectively. Additionally, comparisons of them with other proteins from thermophiles will be described in the latter of this chapter. In Chapter III, the author described the other thermostable proteins and obtaining large crystal for neutron diffraction measurement on thermostable protein. In this chapter, the author discussed the structural features of thermostability on proteins with structural analysis. First, with GTSase and GTHase in the previous chapters, the comparisons of structural features such as SASA and ion-pairs were discussed, which are in same classification of enzyme,

α-amylase family. The comparison revealed differences of structural features on SASA and ion-pairs. After that, an expanding the insight of the comparison among the α-amylase family proteins to other protein which was described this chapter was discussed. In the discussion, charged-residue-rich surface is the major area in solvent-accessible surface. Those charged residues have a role of easing up temperature changes, which can increase thermostability of the proteins. The comparison result mentioned one more important factor which was that the thermostable proteins obtained thermostability by interaction of hydrophobic residues. Packing the hydrophobic residues into the inside of the proteins was also increase thermostability. The thermostability is affected from various factors. One of the factors is effect of ion-pair networks. The discussion about comparison showed suggestions on the ion-pairs. Intra- and/or inter-helical ion-pairs can affect thermostability of the protein. According to the comparison, intra-helical ion-pairs did not always form in the compared proteins, even thermostable proteins. It suggested that inter-helical ion-pairs were more effective to maintain the folding of the proteins than intra-helical ones.

The discussions about differences between thermostable and un-thermostable proteins are useful for industrial use of proteins. As discussed above, proteins from thermophiles have remarkable features in solvent-accessible areas and ion-pair networks. Those features have been brought out by structural analysis. They will contribute to more effective and manageable bioreactor systems. With these trends, results like these cases are required higher efficiency and quality in these days. As previously described, *D-Cha* improved the solution of the requirements and verified the potential by itself. However, the elucidation of structural features of thermostability requires more and more structures. The author had developed a new effective system,

D-Cha. In such situation, more effective methods of structure determination are desired, which will be conducted by a system beyond *D-Cha*. And neutron diffraction method becomes more important for desire of understanding roles of hydrogen atoms and bonds. In this thesis, the preliminary neutron diffraction study on thermostable protein was carried out. However, overall structure of a thermostable protein by using neutron crystallography has not been determined yet. The structure including hydrogen atoms is needed for use of computer simulation. These results have important roles not only for further investigation of thermostability, but also for structure based drug discovery. From the things, use of neutron crystallography will be more important technique for application of structural information to industrial use. And, the author strongly desires using those methods for increasing thermostability of enzymes in industrial use. The author believes that the combination of these techniques will improve the world of structural analysis and accelerate practical use of structural information.

Appendix

Appendix A.
Mail-in data collection at SPring-8 protein crystallography beamlines

Introduction

At SPring-8, the RIKEN Structural Genomics Beamlines I and II (BL26B1 and B2) [11] are constructed to contribute to structural genomics research. To achieve high-throughput protein crystallography, the author's group has developed two special components: an automated sample changer robot, *SPring-8 Precise Automatic Cryo-sample Exchanger* (*SPACE*) [10], which can change up to 100 crystals in a diffractometer, and the beamline control software *BSS* (*Beamline Scheduling Software*) [13], which can perform successive data collection by controlling beamline devices, including *SPACE*, and managing the data collection schedule. The combination of *SPACE* and *BSS* enables unmanned overnight data collection and allows the beamline to operate with high efficiency.

To better accommodate the new system to the needs of distant users and a laboratory information management system (LIMS) for beamline operation and experiments, the author developed a new operation system that enables researchers to use the beamlines from their own laboratories. In general, the methods of conducting experiments using the synchrotron facilities from remote locations fall into two categories: (1) mail-in data collection and (2) remote-controlled data collection. The first method is used by the Swiss Light Source at the Paul Scherrer Institut (SLS/PSI) and the Advanced Photon Source (APS) (Advanced Photon Source, 2006). These users send their samples to a beamline, and datasets collected by beamline staff are returned to the us-

ers. The users, by the remote-controlled method, can actually control their experiments from remote locations via a network program after sending their samples to a beamline. For example, at the Stanford Synchrotron Radiation Laboratory (SSRL) remote data can be collected by using the beamline control application *Blu-Ice* and the *Distributed Control System* (*DCS*) [105]. At the European Synchrotron Radiation Facility (ESRF), this kind of service is available via their remote access control system (European Synchrotron Radiation Facility, 2007). The author has developed a mail-in data collection system, which is a combination of the mail-in and remote access methods. In our system, users can decide their own measurement conditions and benefit from the beamline operator's assistance. Here, the author discusses in detail the mail-in data collection at SPring-8 using the web database application *D-Cha* (*Database for Crystallography with Home-lab Arrangements*), which enables remote operation and mail-in data collection.

Automated operation

The beamline automation system at BL26B2 [11] has been in operation since 2003. There are two modes of operation: mode-1 is the evaluation phase and mode-2 is the data collection phase (Fig. A-1). These operations are made possible by the reproducibility of the *SPACE* mounting. In mode-1, the user or beamline operator interactively centers the crystal on the X-ray beam path and a few diffraction images and optionally the XAFS spectrum of the crystal are automatically measured by *BSS*. The centering position of each crystal is recorded in addition to the diffraction images and XAFS spectrum. Before the using mode-2, the user selects crystals from which to collect datasets, based on the mode-1 evaluation, and enters the measurement conditions

into *BSS*. In mode-2, each crystal is automatically restored to the centering position recorded in mode-1, and measurement is performed automatically and continuously. Thus, mode-1 requires the user to inspect the crystals and diffraction images, whereas mode-2 is an unmanned operation and the user simply waits for completion of data collection. Therefore, modes-1 and -2 are usually conducted during the day and night, respectively.

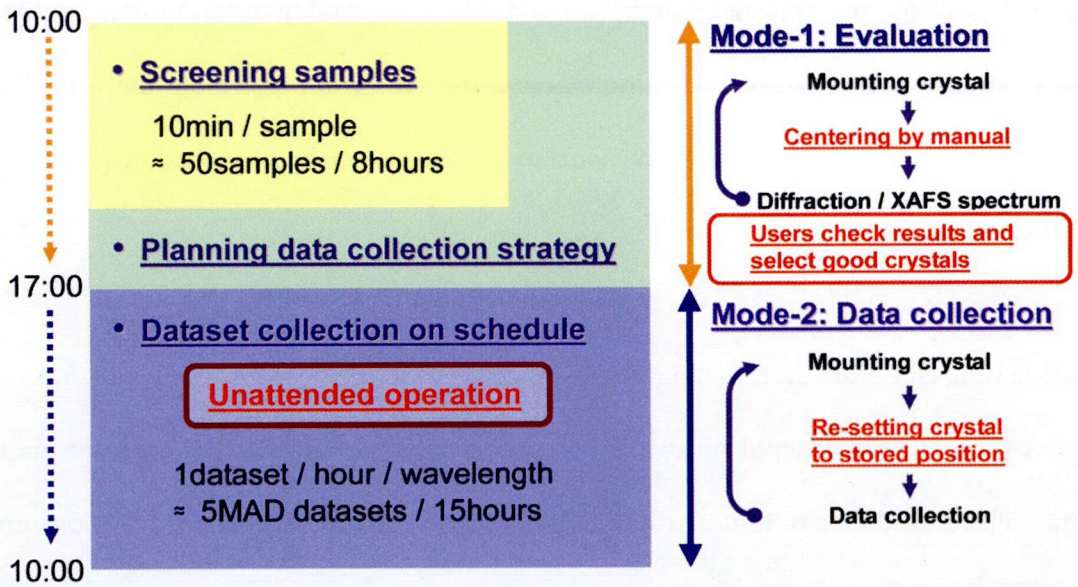


Figure A-1. Schematic diagram of the two modes of operation. In a typical procedure, mode-1 is conducted in daytime and mode-2 is operated during the night.

Mail-in data collection

Aims

Synchrotron facilities are usually at distant location from the user's laboratory, which necessitates considerable expenditure of time and money in traveling and carrying samples. Remote users would therefore benefit from being able to conduct their experiments without visiting the synchrotron facility. Although an outsourcing service is a possible solution, another possibility is remote beamline operation by users.

The aim of our mail-in system is to provide such a service with flexible use of the beamline for distant users. Such a system would both improve beamline usability for researchers and increase the efficiency of data collection. Our mail-in data collection system was designed so that: (1) the user need not visit SPring-8; (2) the user can request particular measurement conditions; (3) the user can check their measurement results and have the collected data returned to them with storage media or via the Internet; and (4) the system performs as a LIMS. To establish this system an additional component is required: a remote-access user interface to enable the user to record sample information, edit the measurement conditions, browse, and acquire measurement data.

Mail-in data collection cycle and database

A typical experimental protocol is shown in Fig. A-2. At the first stage of mail-in data collection, the remote user mounts their measurement samples in trays designed for *SPACE*. This mounting operation is performed using an offline type of mounting robot, which is dedicated to packing the sample or a special toolkit at the user's laboratory. The offline *SPACE* provides automated mounting of the samples to the sample tray for preparation and is a customized for the beamline. Another way of sample preparation is to use a compact toolkit designed for mounting crystals by hand (Fig. A-3). After the trays are prepared, the user sends the samples to SPring-8 and enters measurement schedule from their laboratory using *D-Cha*. The beamline operators load the sample trays sent by the users into *SPACE* and perform the centering of each crystal. The centering position is stored in the sample information database and can be retrieved by *SPACE* and *BSS*. In evaluation mode (mode-1), 10 minutes are re-

quired for each sample, so that the evaluation of 50 samples can be completed in about 8 hours of beamline operators' working time. At this time, the user can browse the measurement results (diffraction images and XAFS spectra) at their remote laboratory using the *D-Cha* web interface. When evaluation mode is completed, the data collection mode (mode-2) is conducted automatically overnight without an operator. The merit of this scheme is that human manipulation and inspection are completed during the daytime.

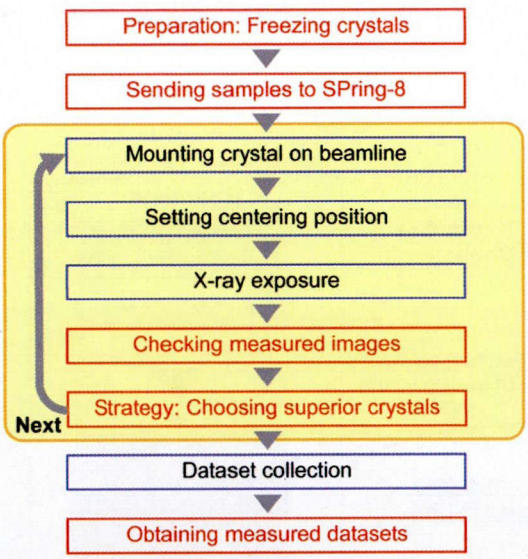


Figure A-2. Typical scheme of our mail-in data collection. Red and blue boxes show operations performed by users and beamline operators respectively. Procedures in the yellow box are repeated for each sample.

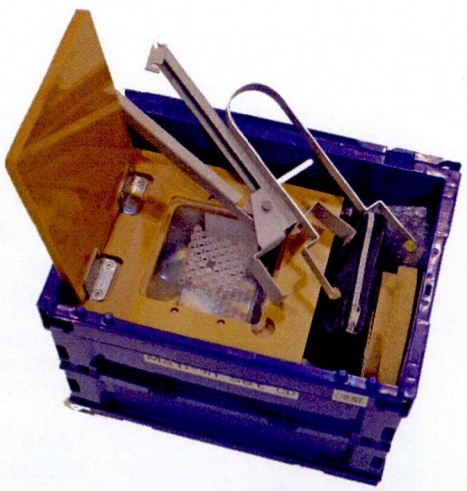


Figure A-3. Overview of the compact toolkit designed for mounting crystals by hand.

D-Cha: a web database application

Overview

D-Cha is a web application consisting of a relational database and a web server. The user can use *D-Cha* via the Internet. Fig. A-4 shows the relationships among the user, the beamline, and *D-Cha*. At the user's laboratory, the user can access *D-Cha*

with a web browser. They can enter and modify measurement conditions, and check the results at their remote laboratory. Offline-*SPACE* is used to upload information on the mounted crystals to *D-Cha* using the Hypertext Transfer Protocol (HTTP). *BSS* fetches measurement conditions from *D-Cha* and uploads experimental results. *BSS* and offline *SPACE* communicate with *D-Cha* in the eXtensible Markup Language (XML) format. *BSS* controls *SPACE* and other beamline components, such as the detector, according to the recorded schedule. In this way, the user and the beamline can communicate using *D-Cha*. Thus, the user can conduct experiments with the synchrotron facility at their laboratory via *D-Cha*.

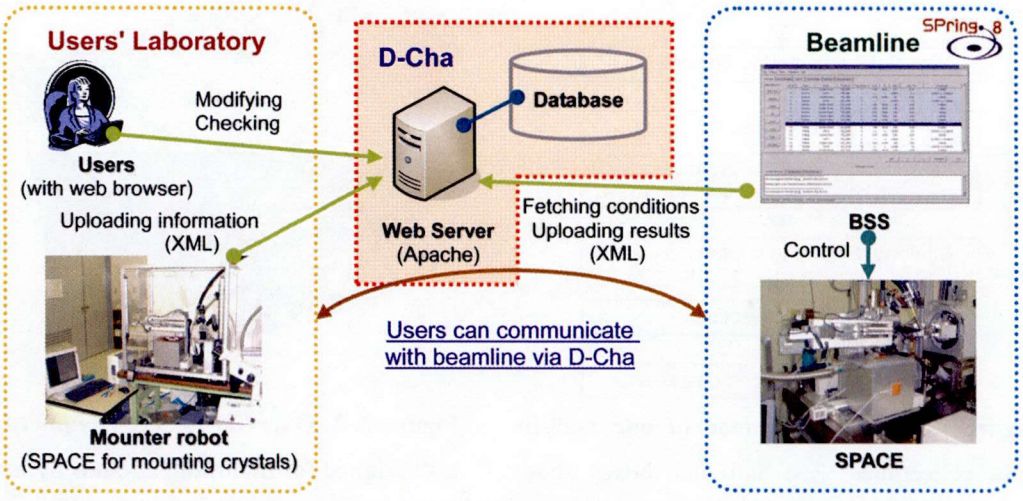


Figure A-4. Schematic diagram of relationships among *D-Cha*, the user, and the beamline. Light green lines represent connections over HTTP or secure HTTP (HTTPS).

Usage and graphical user interface (GUI) of *D-Cha*

The user first accesses *D-Cha* with a web-based GUI and logs in with a registered account name and password. In the next step, the user enters the sample information and experimental conditions using the GUI of the web browser. Two types of sample registration scheme are available on *D-Cha*. One uses the *Tray Manager* and *Crystal*

Manager windows (Figs. A-5a and -5b); alternatively, sample information can be entered via the offline *SPACE* control GUI while the samples are being mounted on the trays. When the samples are all registered, the user enters the measurement schedule for each crystal using the *Crystal Manager* window (Fig. A-5c). Measurement experiments performed by *BSS* are classified into the following four types: (i) diffraction checking; (ii) XAFS; (iii) single dataset collection; and (iv) multiwavelength anomalous dispersion (MAD) dataset collection. The first two measurement modes are in the evaluation phase (mode-1), and the last two modes are in the data collection phase (mode-2). Details of measurement conditions are entered via the *Condition* dialog window (Fig. A-5d). The appearance of the window is similar to those of *BSS*, so that *BSS* users can easily operate this GUI interface.

The user can browse the measurement results on *D-Cha*. The *Condition* dialog displays a list of the result files (Fig. A-6a). The user can browse the photographs, diffraction images, and XAFS spectra of each crystal (Figs. A-6b, -6c, and -6d) and download their raw data through these dialog windows.

Data management

D-Cha is designed to manage massive amounts of data, including sample information, measurement conditions, and the results. Trays, crystals and measurement conditions are identified by a unique tray ID, crystal ID and experiment number for each crystal ID, respectively. All information is accessible only by its owner and illegal access is prohibited to ensure privacy for each user.

D-Cha provides a simple experimental database that can store any information that the user has independently defined for the crystals and measurements (for example,

sample name, cryo condition, heavy atoms, and comments). The user can use these data freely and relate records in *D-Cha* to their own database and to the LIMS.

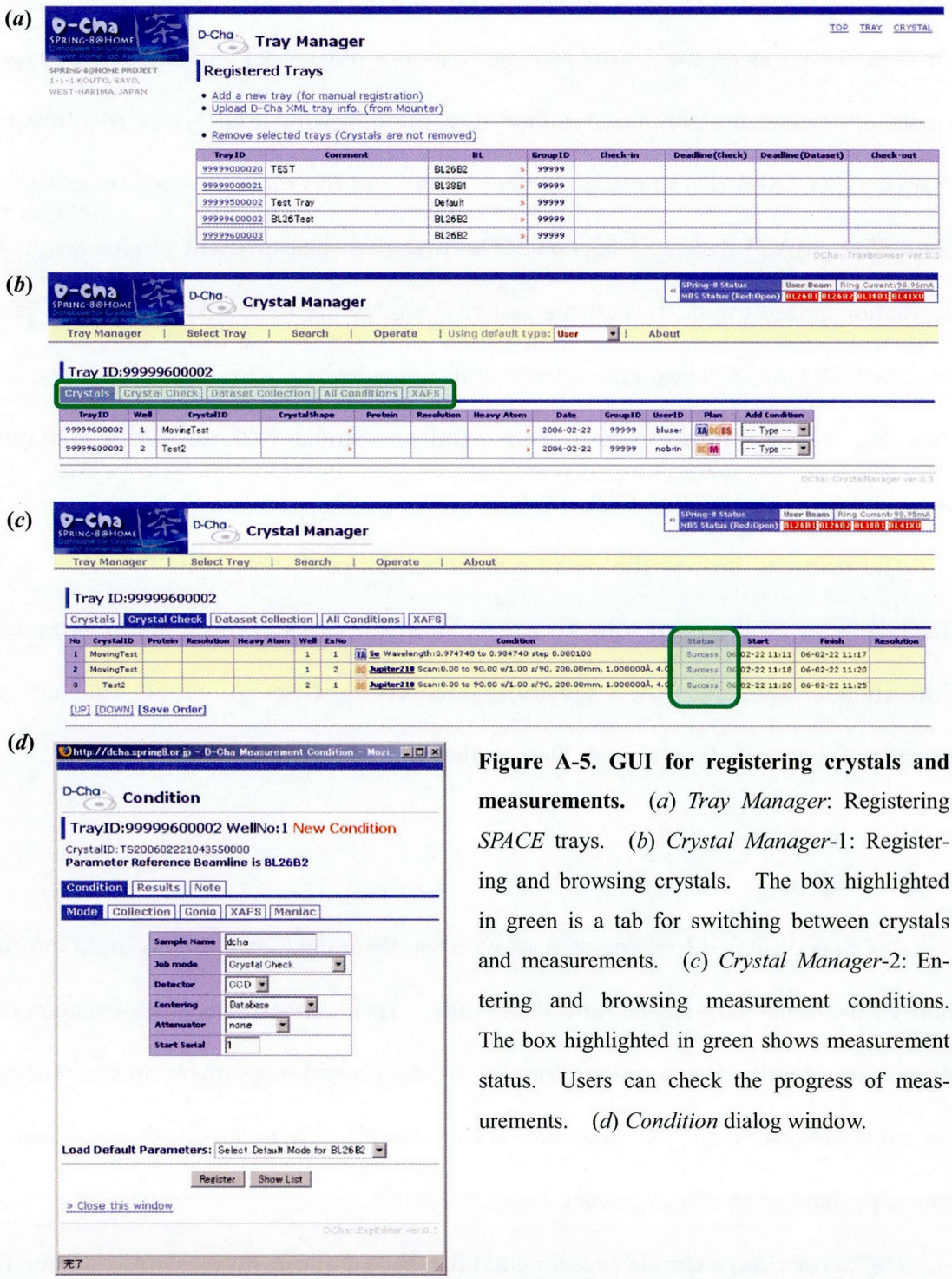


Figure A-5. GUI for registering crystals and measurements. (*a*) *Tray Manager*: Registering *SPACE* trays. (*b*) *Crystal Manager*-1: Registering and browsing crystals. The box highlighted in green is a tab for switching between crystals and measurements. (*c*) *Crystal Manager*-2: Entering and browsing measurement conditions. The box highlighted in green shows measurement status. Users can check the progress of measurements. (*d*) *Condition* dialog window.

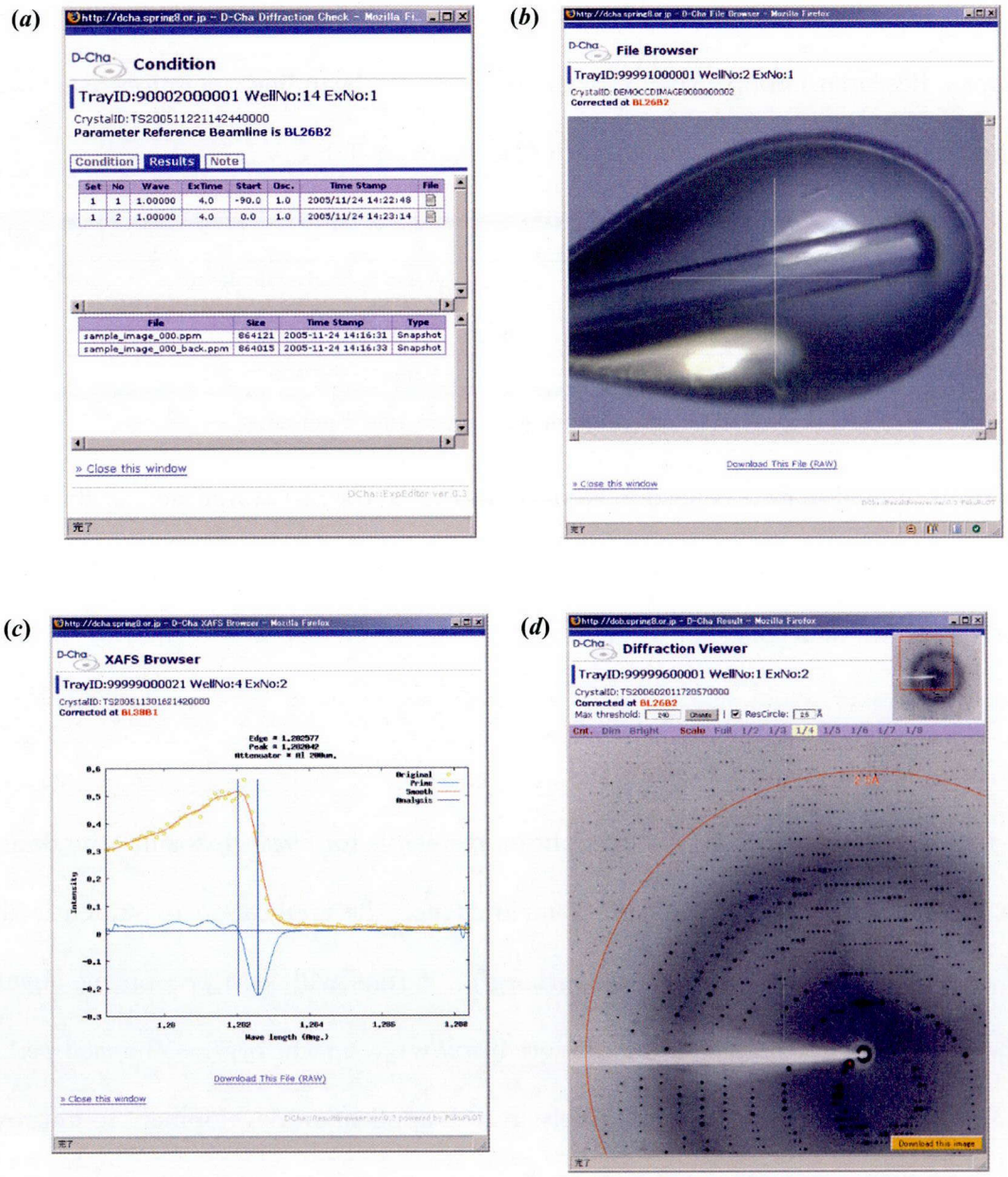


Figure A-6. Dialog windows for browsing measurement results. (*a*) *Condition* dialog with result file listing. The user can open the result browser by clicking the filename. (*b*) Crystal snapshot view. (*c*) XAFS spectrum browser. (*d*) Diffraction image viewer.

Operation result

At BL26B2, mail-in data collection runs in a routine manner. It is used by this way by facilities like the Genomic Science Center (RIKEN Yokohama), the SR System

Biology Research Group (RIKEN SPring-8 Center), and the Advanced Protein Crystallography Research Group (RIKEN SPring-8 Center). Beamline operation results for the first half of the year 2007 (2007A period) is shown in Table A-1.

Operating days	26	
Evaluated crystals	652	Average 25 crystals/daytime
Single datasets	188	
MAD experiments	17 (48 datasets)	
Total datasets	**236**	**Average 9 datasets/night**

Table A-1. Measurements made in 2007A period.

Trials of mail-in data collection were conducted at the Structural Biology Beamline for the public (BL38B1). Eleven academic users have conducted their measurements with the mail-in data collection system.

Platform, language and availability

D-Cha was developed as a web application suitable for *Firefox* (Mozilla Foundation, http://www.mozilla.org/). For ease of maintenance, the application is written in the *Perl* scripting language (http://www.perl.org/). It runs with *mod_perl* on the *Apache* Web Server (Apache Software Foundation, http://www.apache.org/). The *mod_perl* is the *Perl* interface for *Apache*, used widely in web application development to improve the performance of applications. *D-Cha* consists of several modules: *CGI::Application* (a web application framework), *DBIx::Class* (an object-relational mapper), and *Template Toolkit* for HTML processing. These modules are obtainable from the Comprehensive Perl Archive Network (CPAN, http://www.cpan.org/). *D-Cha* uses a relational database, *PostgreSQL* (http://www.postgresql.org/).

D-Cha is currently customized to work with *SPACE* and *BSS* at SPring-8 and is used at the Pharmaceutical Industry Beamline (BL32B2), BL26B1, BL26B2 and BL38B1.

Appendix B.
Further structural investigation by using neutron crystallography

Introduction

Increasing the number of solved protein structures will enable us to understand protein function as well as to create useful molecules. X-ray crystallography is the most popular technique for obtaining protein structure information. It is known that H atoms play crucial roles in the molecular recognition and catalytic reactions of enzymes. The locations of H atoms in protein molecules are usually predicted by theoretical approaches, but it is still very difficult to determine the ionization status of catalytic residues, the hydration structure and the characteristics of hydrogen-bonding interactions, especially low-barrier hydrogen-bonding interactions. There are two approaches to observing the location of H atoms: ultrahigh-resolution X-ray crystallography and neutron crystallography [80,106]. Since X-ray diffraction shows the location of the electron, the bond distance is sometimes different when observed by neutron diffraction. On the basis of these features, the author attempted to determine protein structures using both X-ray and neutron diffraction. Three different data sets were collected, including X-ray diffraction data at 100 K and room-temperature and neutron diffraction data at room temperature (using exactly the same crystal as used for X-ray data collection at room temperature). Here, the author reports successful examples of the structure determination of human immunodeficiency virus (HIV-1) protease and porcine pancreatic elastase (PPE) using a combination of neutron and X-ray diffraction data. These proteins are not thermostable. Obtaining crystals which satisfy volume for neutron diffraction measurement are challenging works. Therefore, the author at-

tempted those cases of neutron crystallography during crystallization of thermostable proteins. In next chapter, the author will describe crystallization on thermostable protein before neutron crystallography.

More than 60 000 X-ray structures have been archived in the Protein Data Bank; although neutron crystallography is a powerful tool for determining the locations of H atoms, neutron diffraction has contributed a total of at most 40 structures (http://www.pdb.org/). Therefore, further technical developments are needed for further effective utilization of neutron diffraction in protein structure determination. The author approaches to developing crystallization and changing the crystal lattice by protein engineering are also described.

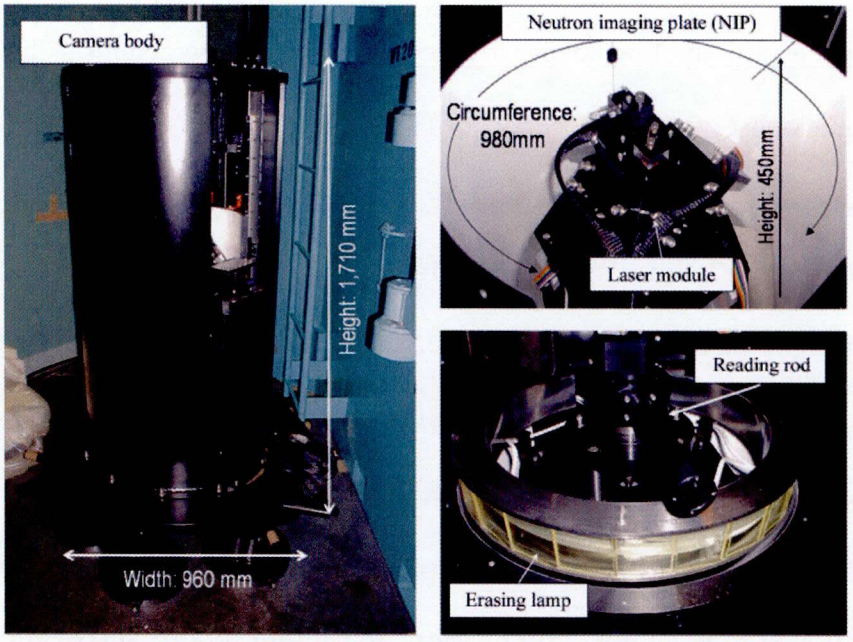


Figure B-1. The BIX-4 neutron diffractometer at JRR-3. The camera body of BIX-4 is shown (left) and its main components: the NIP and laser modules (top right) and the reading rod and erasing lamp (bottom right). The camera body (height = 1710 mm, width = 960 mm) is covered by a cylindrical shielding (not shown) composed of two layers: a 50 mm thick B4C + resin shield against neutrons and a 50 mm thick lead shield against γ rays.

Neutron structure analysis of drug-target enzymes

Neutron diffractometers

Neutron diffractometers (BIX-3 and BIX-4) equipped with neutron imaging plates (NIPs) were designed and built by Niimura and coworkers [103,107]. These diffractometers are installed at the 1G-A and 1G-B ports, respectively, of the JRR-3 research reactor at Japan Atomic Energy Agency, Tokai, Japan. The basic architectures of BIX-3 and BIX-4 are essentially the same. In Figure B-1, the appearance of the camera body of BIX-4 is shown. The NIP covers a large solid angle subtended at the sample with a

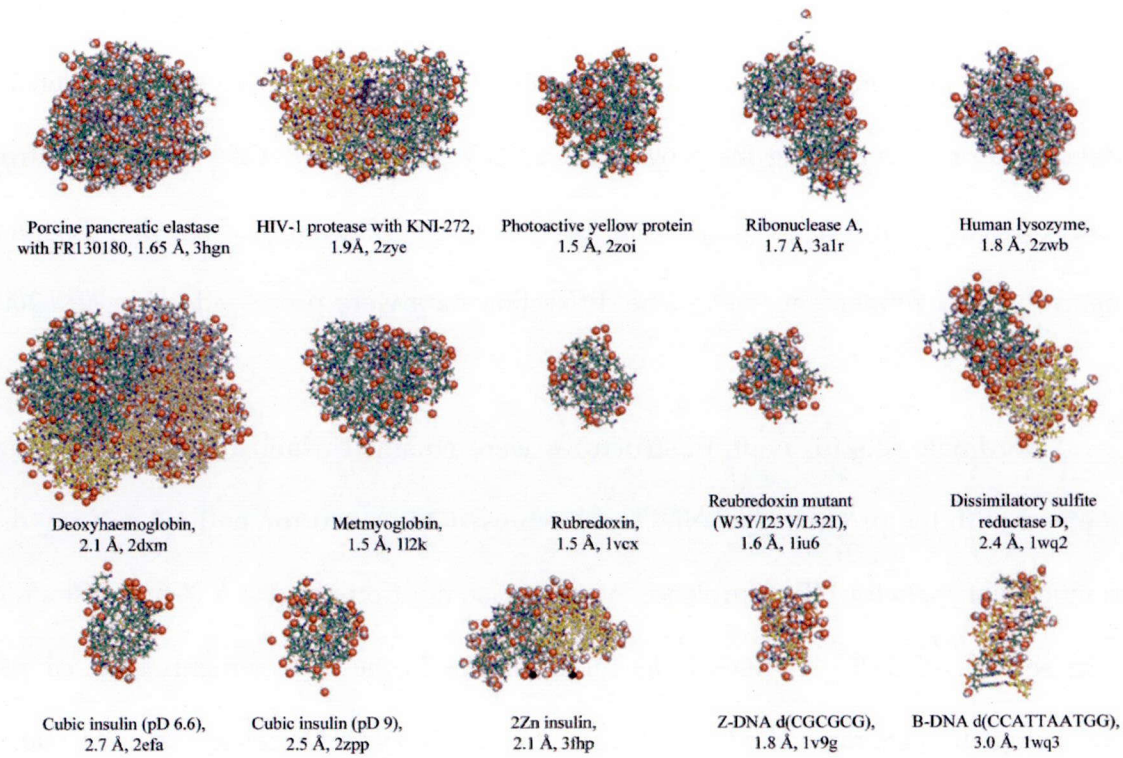


Figure B-2. Tertiary structures of biological macromolecules obtained using BIX-3 and BIX-4. A total of 15 structures of biological macromolecules have been determined using BIX-3 and BIX-4 since 2002 and are shown as space-filling representations. The molecular name, resolution and PDB code of each sample are shown under the structure models. Figures of protein structures were drawn using the program *PyMOL* (http://pymol.org/).

fine positional linearity, a high spatial accuracy and good neutron detector efficiency.

The circumference of the NIP cylinder is 980 mm ($2\theta_S$ angle range = ±140°), its height is 450 mm and the camera radius is 200 mm. The wavelengths of the incident neutrons for BIX-3 and BIX-4 are 2.9 and 2.6 Å, respectively. According to these wavelength values *i.e.* the highest possible resolution values, are 1.53 and 1.38 Å, respectively. Since 2002 BIX-3 and BIX-4 have contributed to 15 structure determinations of biological macromolecules (as shown in Fig. B-2).

Data collection and refinement

Neutron diffraction data collections using BIX-3 and BIX-4 were usually carried out at room temperature with protein crystals of 1–10 mm^3 in volume. The intensity data were processed using the programs *DENZO* and *SCALEPACK* [47]. After neutron data collection, the same crystals were used to collect X-ray diffraction data for joint neutron/X-ray refinement [104,108]. The diffraction data were processed using *HKL2000* [47].

Coordinate sets for neutron structures were obtained using the joint refinement method with the program *PHENIX* [109], which used 1.9 Å neutron and 1.4 Å X-ray diffraction data sets for HIV-1 protease [71] and 1.65 Å neutron and 1.2 Å X-ray diffraction data sets for PPE [72]. Manual model modifications in these refinements were carried out using the programs *XtalView* [110], *Coot* [50] and *QUANTA* (Accelrys Inc., San Diego, California, USA).

High-resolution X-ray diffraction data were also collected from a crystal with a volume of ~0.001 mm^3 at 100 K at the synchrotron-radiation sources SPring-8 (Hyogo, Japan) and Photon Factory (Tsukuba, Japan) and were processed using *HKL2000*. High-resolution X-ray structures for each protein were refined using *CNS* [44] followed

by *SHELX-97* [111].

Structure of HIV-1 protease in complex with the potent inhibitor KNI-272

HIV-1 protease is a dimeric aspartic protease that contains two aspartic acid residues as catalytic residues (defined as residue positions Asp25 and Asp125) and plays an essential role in viral replication. To develop HIV-1 protease inhibitors through

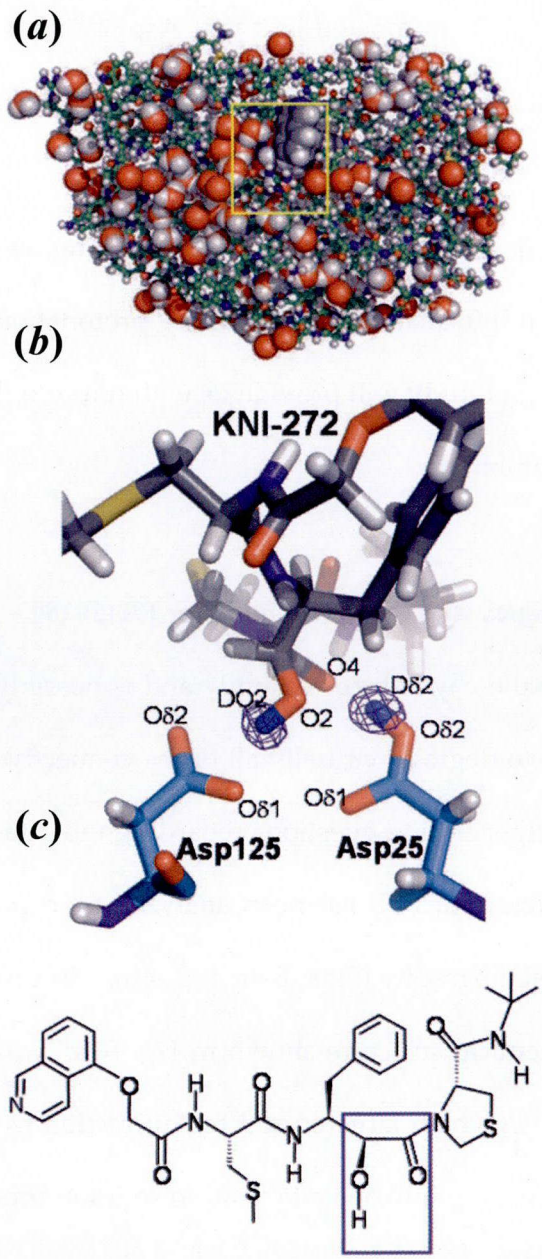


Figure B-3. Tertiary structure of HIV-1 protease in complex with its inhibitor determined by neutron diffraction. (*a*) Overall dimer structure of HIV-1 protease determined by neutron diffraction. The HIV-1 protease is shown as a ball-and-stick model; water molecules and bound inhibitor are shown in space-filling representation. H and D, C, O, N and S atoms are represented in white, cyan, red, blue and yellow, respectively. C atoms in KNI-272 are shown in dark grey. (*b*) Interaction between the active site and KNI-272 in the yellow box in (*a*). The $|F_o-F_c|$ omit nuclear density map shown at the 5σ level in blue was calculated without the contributions of the Dδ2 and DO2 atoms. (*c*) Chemical structure of the inhibitor KNI-272. The blue box indicates the hydroxymethylcarbonyl moiety.

structure-based drug design, it is useful to understand the catalytic mechanism and inhibitor recognition of HIV-1 protease. Thus, the author's group determined all atom positions, including H atoms, of HIV-1 protease in complex with KNI-272 using neutron diffraction (PDB ID: **2ZYE**; [71]; Figs. B-3a and -3b). The inhibitor KNI-272 used in this study contains a hydroxymethylcarbonyl isostere moiety (shown in a blue box in Fig. B-3c) that interacts with the catalytic residues of HIV-1 protease. The neutron diffraction analysis directly showed that Asp25 is protonated and that Asp125 is unprotonated (Fig. B-3b). Although the catalytic mechanism of HIV-1 protease has been a matter of some debate, the results demonstrate that Asp25 provides a proton to the carbonyl group of the substrate and Asp125 contributes to activating the attacking water molecule as a nucleophile. The structural information, including the protonation states of the catalytic residues, determined in this study will provide us with important information for the design of more effective inhibitors.

Structures of porcine pancreatic elastase in complex with the potent inhibitor FR130180

Elastase is a serine protease classified in the chymotrypsin family and is possibly the most destructive enzyme, with an ability to degrade virtually all of the connective components in the body. To help resolve long-standing questions regarding the catalytic activity of the serine proteases, the structure of PPE has been analyzed by combined high-resolution neutron and X-ray crystallography (Figs. B-4a and -4b). In this analysis, the peptidic inhibitor FR130180 (chemical structure shown in Fig. B-4c) was used to mimic the tetrahedral intermediate. A single large crystal for diffraction experiments was prepared by repeating macroseeding into a crystallization solution prepared with deuterated reagents [112].

The tertiary structure of PPE in complex with FR130180 was determined to 1.65 Å resolution by neutron crystallography and also to 1.20 Å resolution by X-ray crystallography using diffraction data obtained at room temperature from the same crystal (PDB ID: **3HGN**; [72]). The complex structure includes a total of 1792 H and D atoms and 190 hydration water molecules. The neutron analysis showed that the O atom of

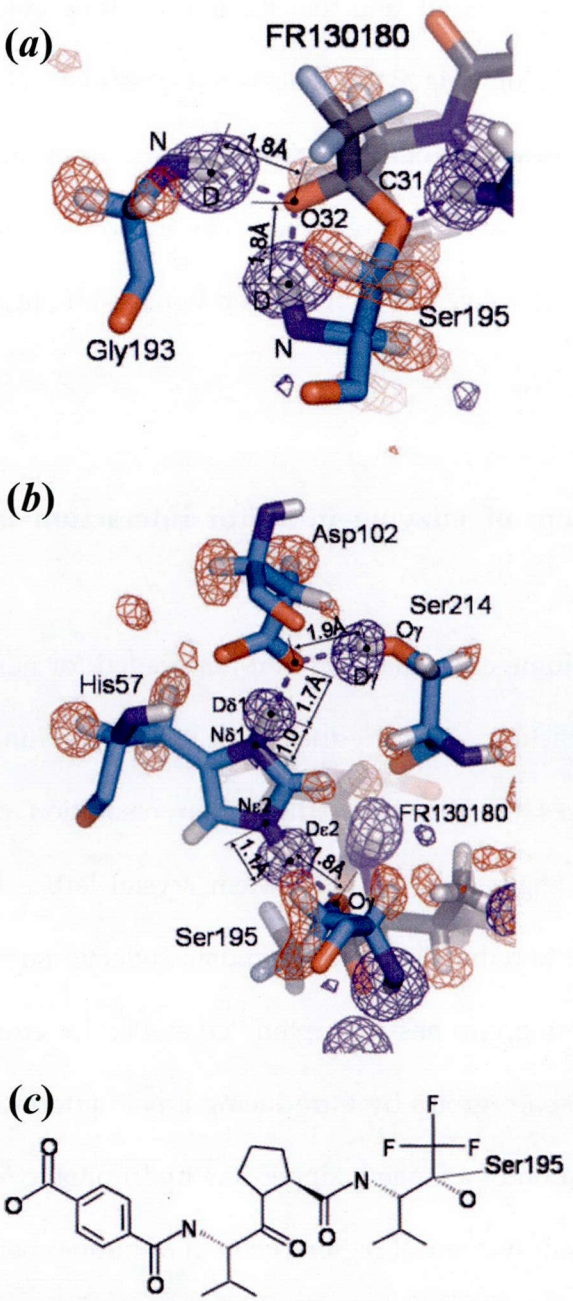


Figure B-4. Inhibitor recognition of PPE determined by neutron diffraction. (*a*) Tertiary structure of the oxyanion hole of PPE. H and D, C, O and N atoms are represented in white, cyan, red and blue, respectively. C atoms in FR130180 are shown in dark grey. The $|F_o-F_c|$ omit nuclear density maps were calculated without the contributions of H and D atoms. The blue and red contours show +5.0σ and −4.5σ densities, respectively. (*b*) Tertiary structure of the catalytic triad of PPE. The $|F_o-F_c|$ nuclear density maps were calculated as above. (*c*) Chemical structure of FR130180.

the oxopropyl group of the inhibitor was present as an oxygen anion rather than a hydroxyl group (Fig. B-4a) because the O atom (O32) was directed towards the D (H) atoms of the backbone amides of Gly193 and Ser195 and no nuclear density was observed for deuterium attached to O32.

The neutron and X-ray data also show that the hydrogen bond between His57 and Asp102 (chymotrypsin numbering) is 2.6 Å in length and that the hydrogen-bonding hydrogen is 0.8–1.0 Å from the histidine N atom (Fig. B-4b). This is not consistent with a low-barrier hydrogen bond, which would be predicted to position the hydrogen midway between the donor and acceptor atoms. The observed interaction between His57 and Asp102 is essentially a short but conventional hydrogen bond, sometimes described as a short ionic hydrogen bond.

Future aspects of solving the problem of enzyme–inhibitor interaction by neutron crystallography

It is generally known that a large-volume crystal (1.0–10 mm^3) is needed for neutron diffraction because of the weak diffraction. Since reducing the unit-cell volume of a protein crystal empirically improves the maximum diffraction resolution by strengthening the diffraction intensities, engineering of the protein crystal lattice to change the space group should contribute to reducing the neutron data-collection time by improving the symmetry. The author's group has attempted "crystal lattice engineering" to change the packing and the space group by introducing a new interface. The first attempts involved the incorporation of a leucine zipper-like hydrophobic interface (comprising four leucine residues) into a helical region (helix 2) of human pancreatic ribonuclease 1 (RNase 1). The mutant RNase 1 was successfully crystallized in

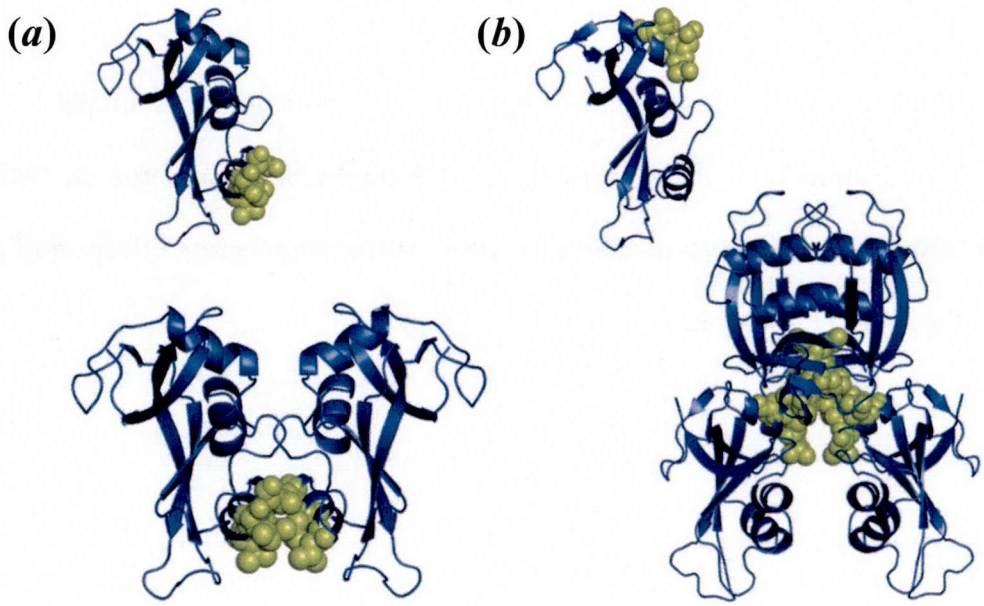


Figure B-5. Leucine-induced artificial association of mutant human RNase-1. (*a*) Association of the 4L mutant human RNase 1 in which four leucines were incorporated into helix 2. Two mutant RNases were associated by leucines (shown as van der Waals representations). (*b*) Association of the N4L mutant human RNase 1 in which four leucines were incorporated into helix 3. Four mutant RNases were associated; the incorporated leucines are shown as van der Waals representations.

space group $I4_1$, with a unit-cell volume of 412 678.7 Å^3, and the crystal structure was subsequently determined by X-ray crystallography [113]. The overall structure of 4L-RNase 1, in which amino acids at four positions were mutated to leucine, is quite similar to that of bovine RNase A and the introduced leucine residues formed the designed crystal interface (Fig. B-5a). To further characterize the role of the introduced leucine residues in the crystallization of RNase 1, the number of leucines was reduced to three or two (3L-RNase 1 and 2L-RNase 1, respectively). Both mutants crystallized and a similar hydrophobic interface to that in 4L-RNase 1 was observed. A related approach of engineering crystal contacts at helix 3 of RNase 1 (N4L-RNase 1) was also evaluated. N4L-RNase 1 was also successfully crystallized in space group $P6_122$, with a unit-cell volume of 941 059.7 Å^3, and formed the expected hydrophobic packing in-

terface. In Figure B-5b, the crystal packings of 4L-RNase 1 and N4L-RNase 1 are shown. These results suggest that the appropriate introduction of a unique leucine zipper-like hydrophobic interface can promote intermolecular symmetry, *i.e.* twofold for more efficient protein crystallization in crystal lattice engineering efforts, and may optimize the crystal lattice for neutron crystallography.

References

1. New Enzyme for More Efficient Corn Ethanol Production. (2005). http://www.greencarcongress.com/2005/06/new_enzyme_for_.html

2. Yokoyama, S., Matsuo, Y., Hirota, H., Kigawa, T., Shirouzu, M., Kuroda, Y., Kurumizaka, H., Kawaguchi, S., Ito, Y., Shibata, T., Kainosho, M., Nishimura, Y., Inoue, Y. & Kuramitsu, S. Structural genomics projects in Japan. *Progress in Biophysics & Molecular Biology* **73**, 363–376 (2000).

3. Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. & Rees, D. C. Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463–1469 (1995).

4. Rice, D. W., Yip, K. S., Stillman, T. J., Britton, K. L., Fuentes, A., Connerton, I., Pasquo, A., Scandura, R. & Engel, P. C. Insights into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS Microbiol. Rev.* **18**, 105–117 (1996).

5. Fang, T.-Y., Tseng, W.-C., Chung, Y.-T. & Pan, C.-H. Mutations on aromatic residues of the active site to alter selectivity of the *Sulfolobus solfataricus* maltooligosyltrehalose synthase. *J. Agric. Food Chem.* **54**, 3585–3590 (2006).

6. Fang, T.-Y., Tseng, W.-C., Pan, C.-H., Chun, Y.-T. & Wang, M.-Y. Protein engineering of *Sulfolobus solfataricus* maltooligosyltrehalose synthase to alter its selectivity. *J. Agric. Food Chem.* **55**, 5588–5594 (2007).

7. Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J. & Withers, S. G. Changing enzymatic reaction mechanisms by glucosidase to an inverting enzyme. *J. Am. Chem. Soc.* **116**, 11594–11595 (1994).

8. Yokoyama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. Structural genomics projects in Japan. *Nat. Struct. Biol.* **7 Suppl**, 943–945 (2000).

9. Hiraki, M., Watanabe, S., Yamada, Y., Matsugaki, N., Igarashi, N., Gaponov, Y. & Wakatsuki, S. Automated protein crystallography in KEK-PF. *KEK Mechanical Encineering Center Workshop on Accelerator Structure Development* (2007).

10. Ueno, G., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. Sample management system for a vast amount of frozen crystals at SPring-8. *J. Appl. Crystallogr.* **37**, 867–873 (2004).

11. Ueno, G., Kanda, H., Hirose, R., Ida, K., Kumasaka, T. & Yamamoto, M. RIKEN structural genomics beamlines at the SPring-8; high throughput protein crystallography with automated beamline operation. *J. Struct. Funct. Genomics* **7**, 15–22 (2006).

12. Yamamoto, M., Hirata, K., Hikima, T., Kawano, Y. & Ueno, G. Protein micro-crystallography with a new micro-beam beamline. *Yakugaku Zasshi* **130**, 641–648 (2010).

13. Ueno, G., Kanda, H., Kumasaka, T. & Yamamoto, M. Beamline Scheduling Software: administration software for automatic operation of the RIKEN structural genomics beamlines at SPring-8. *J. Synchrotron Radiat.* **12**, 380–384 (2005).

14. Okazaki, N., Hasegawa, K., Ueno, G., Murakami, H., Kumasaka, T. & Yamamoto, M. Mail-in data collection at SPring-8 protein crystallography beamlines. *J. Synchrotron Radiat.* **15**, 288–291 (2008).

15. Okazaki, N., Adachi, M., Tamada, T., Kurihara, K., Ooga, T., Kamiya, N., Kuramitsu, S. & Kuroki, R. Crystallization and preliminary neutron diffraction studies of ADP-ribose pyrophosphatase-I from *Thermus thermophilus* HB8. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **68**, 49–52 (2012).

16. de-Araujo, P. S. The role of trehalose in cell stress. *Braz. J. Med. Biol. Res.* **29**, 873–875 (1996).

17. Newman, Y. M., Ring, S. G. & Colaco, C. The role of trehalose and other carbohydrates in biopreservation. *Biotechnology & Genetic Engineering Reviews* **11**, 263–294 (1993).

18. Potts, M. Desiccation tolerance of prokaryotes. *Microbiol. Rev.* **58**, 755–805 (1994).

19. Schiraldi, C., Di Lernia, I. & De Rosa, M. Trehalose production: exploiting novel approaches. *Trends Biotechnol.* **20**, 420–425 (2002).

20. Nakada, T., Maruta, K., Tsusaki, K., Kubota, M., Chaen, H., Sugimoto, T., Kurimoto, M. & Tsujisaka, Y. Purification and properties of a novel enzyme, maltooligosyl trehalose synthase, from *Arthrobacter* sp. Q36. *Biosci., Biotechnol., Biochem.* **59**, 2210–2214 (1995).

21. Nakada, T., Maruta, K., Mitsuzumi, H., Kubota, M., Chaen, H., Sugimoto, T., Kurimoto, M. & Tsujisaka, Y. Purification and characterization of a novel enzyme, maltooligosyl trehalose trehalohydrolase, from *Arthrobacter* sp. Q36. *Biosci., Biotechnol., Biochem.* **59**, 2215–2218 (1995).

22. Maruta, K., Nakada, T., Kubota, M., Chaen, H., Sugimoto, T., Kurimoto, M. & Tsujisaka, Y. Formation of trehalose from maltooligosaccharides by a novel enzymatic system. *Biosci., Biotechnol., Biochem.* **59**, 1829–1834 (1995).

23. Nakada, T., Ikegami, S., Chaen, H., Kubota, M., Fukuda, S., Sugimoto, T., Kurimoto, M. & Tsujisaka, Y. Purification and characterization of thermostable maltooligosyl trehalose trehalohydrolase from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. *Biosci., Biotechnol., Biochem.* **60**, 267–270 (1996).

24. Nakada, T., Ikegami, S., Chaen, H., Kubota, M., Fukuda, S., Sugimoto, T., Kurimoto, M. & Tsujisaka, Y. Purification and characterization of thermostable maltooligosyl trehalose trehalose synthase from the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. *Biosci., Biotechnol., Biochem.* **60**, 263–266 (1996).

25. Kato, M., Miura, Y., Kettoku, M., Komeda, T., Iwamatsu, A. & Kobayashi, K. Reaction mechanism of a new glycosyltrehalose-hydrolyzing enzyme isolated from the hyperthermophilic archaeum, *Sulfolobus solfataricus* KM1. *Biosci., Biotechnol., Biochem.* **60**, 925–928 (1996).

26. Kato, M., Miura, Y., Kettoku, M., Shindo, K., Iwamatsu, A. & Kobayashi, K. Reaction mechanism of a new glycosyltrehalose-producing enzyme isolated from the hyperthermophilic archaeum, *Sulfolobus solfataricus* KM1. *Biosci., Biotechnol., Biochem.* **60**, 921–924 (1996).

27. Kato, M., Miura, Y., Kettoku, M., Shindo, K., Iwamatsu, A. & Kobayashi, K. Purification and characterization of new trehalose-producing enzymes isolated from the hyperthermophilic archae, *Sulfolobus solfataricus* KM1. *Biosci., Biotechnol., Biochem.* **60**, 546–550 (1996).

28. Svensson, B. Protein engineering in the α-amylase family: catalytic mechanism, substrate specificity, and stability. *Plant Mol. Biol.* **25**, 141–157 (1994).

29. Henrissat, B. & Daviest, G. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* **7**, 637–644 (1997).

30. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. & Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233–D238 (2009).

31. Kato, M., Takehara, K., Kettoku, M., Kobayashi, K. & Shimizu, T. Subsite structure and catalytic mechanism of a new glycosyltrehalose-producing

enzyme isolated from the hyperthermophilic archaeum, *Sulgolobus solfataricus* KM1. *Biosci., Biotechnol., Biochem.* **64**, 319–326 (2000).

32. Feese, M. D., Kato, Y., Tamada, T., Kato, M., Komeda, T., Miura, Y., Hirose, M., Kondo, K., Kobayashi, K. & Kuroki, R. Crystal structure of glycosyltrehalose trehalohydrolase from the hyperthermophilic archaeum *Sulfolobus solfataricus*. *J. Mol. Biol.* **301**, 451–464 (2000).

33. Kobayashi, M., Kubota, M. & Matsuura, Y. Refined structure and functional implications of trehalose synthase from *Sulfolobus acidocaldarius*. *J. Appl. Glycosci.* **50**, 1–8 (2003).

34. Cielo, C. B. C., Okazaki, S., Suzuki, A., Mizushima, T., Masui, R., Kuramitsu, S. & Yamane, T. Structure of ST0929, a putative glycosyl transferase from *Sulfolobus tokodaii*. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **66**, 397–400 (2010).

35. Brzozowski, A. M. & Davies, G. J. Structure of the *Aspergillus oryzae* α-Amylase complexed with the inhibitor acarbose at 2.0 Å resolution. *Biochemistry* **36**, 10837–10845 (1997).

36. Boel, E., Brady, L., Brzozowski, A. M., Derewenda, Z., Dodson, G. G., Jensen, V. J., Petersen, S. B., Swift, H., Thim, L. & Woldike, H. F. Calcium binding in α-amylases: an X-ray diffraction study at 2.1-Å resolution of two enzymes from *Aspergillus*. *Biochemistry* **29**, 6244–6249 (1990).

37. Kubota, M., Maruta, K., Fukuda, S., Kurimoto, M., Tsujisaka, Y., Kobayashil, M. & Matsuura, Y. Structure and function analysis of malto-oligosyltrehalose synthase. *J. Appl. Glycosci.* **48**, 153–161 (2001).

38. Janeček, Š. Sequence similarities in $(\alpha/\beta)_8$-barrel enzymes revealed by conserved regions of α-amylase. *FEBS Lett.* **316**, 23–26 (1993).

39. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).

40. Koshrand, D. E. J. Stereochemistry and the mechanism of enzymatic reactions. *Biological Reviews* **28**, 416–436 (1953).

41. Withers, S. G. & Aebersold, R. Approaches to labeling and identification of active site residues in glycosidases. *Protein Sci.* **4**, 361–372 (1995).

42. Hasegawa, K., Kubota, M. & Matsuura, Y. Roles of catalytic residues in alpha-amylases as evidenced by the structures of the product-complexed mutants of a maltotetraose-forming amylase. *Protein Eng.* **12**, 819–824 (1999).

43. Varrot, A., Frandsen, T. P., Von Ossowski, I., Boyer, V., Cottaz, S., Driguez, H., Schülein, M. & Davies, G. J. Structural basis for ligand binding and processivity in cellobiohydrolase Cel6A from *Humicola insolens*. *Structure* **11**, 855–864 (2003).

44. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. Crystallography & NMR

System: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **D54**, 905–921 (1998).

45. Ohtaki, A., Mizuno, M., Tonozuka, T., Sakano, Y. & Kamitori, S. Complex structures of *Thermoactinomyces vulgaris* R-47 α-amylase 2 with acarbose and cyclodextrins demonstrate the multiple substrate recognition mechanism. *J. Biol. Chem.* **279**, 31033–31040 (2004).

46. Uitdehaag, J. C. M., Mosi, R., Kalk, K. H., Van der Veen, B. A., Dijkhuizen, L., Withers, S. G. & Dijkstra, B. W. X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nat. Struct. Biol.* **6**, 432–436 (1999).

47. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).

48. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).

49. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **53**, 240–255 (1997).

50. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132 (2004).

51. Lee, B. & Richards, F. M. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400 (1971).

52. Kawabata, T. MATRAS: a program for protein 3D structure comparison. *Nucleic Acids Res.* **31**, 3367–3369 (2003).

53. Lovell, S. C., Davis, I. W., Arendall, W. B., De Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. Structure validation by Cα geometry: ϕ, ψ and Cβ deviation. *Proteins* **50**, 437–450 (2003).

54. Collaborative Computational Project, N. 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763 (1994).

55. Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M. & Henrissat, B. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Engineering, Design & Selection* **19**, 555–562 (2006).

56. Elbein, A. D. The metabolism of α,α-trehalose. *Adv. Carbohydr. Chem. Biochem.* **30**, 227–256 (1974).

57. Kato, M., Miura, Y., Kettoku, M., Shindo, K., Iwamatsu, A., Kobayashi, K. & Komeda, T. Gene cloning and expression of new trehalose-producing enzymes from the hyperthermophilic archaeum *Sulfolobus solfataricus* KM1. *Biosci., Biotechnol., Biochem.* **60**, 1882–1885 (1996).

58. Miura, Y., Kettoku, M., Kato, M., Kobayashi, K. & Kondo, K. High level production of thermostable α-amylase from Sulfolobus solfataricus in high-cell density culture of the food yeast *Candida utilis*. *J. Mol. Microbiol. Biotechnol.* **1**, 129–134 (1999).

59. Takata, H., Kuriki, T., Okada, S., Takesada, Y., Iizuka, M., Minamiura, N. & Imanaka, T. Action of neopullulanase. *J. Biol. Chem.* **267**, 18447–18452 (1992).

60. Kuroki, R., Weaver, L. H. & Matthews, B. W. A covalent enzyme-substrate intermediate with saccharide distortion in a mutant T4 lysozyme. *Science* **262**, 2030–2033 (1993).

61. Kuroki, R., Weaver, L. H. & Matthews, B. W. Structure-based design of a lysozyme with altered catalytic activity. *Nat. Struct. Biol.* **2**, 1007–1011 (1995).

62. Kuroki, R., Weaver, L. H. & Matthews, B. W. Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 8949–8954 (1999).

63. Qian, M., Nahoum, V., Bonicel, J., Bischoff, H., Henrissat, B. & Payan, F. Enzyme-catalyzed condensation reaction in a mammalian α-amylase. High-resolution structural analysis of an enzyme-inhibitor complex. *Biochemistry* **40**, 7700–7709 (2001).

64. Timmins, J., Leiros, H.-K. S., Leonard, G., Leiros, I. & McSweeney, S. Crystal structure of maltooligosyltrehalose trehalohydrolase from *Deinococcus radiodurans* in complex with disaccharides. *J. Mol. Biol.* **347**, 949–963 (2005).

65. Hassinen, T. & Peräkylä, M. New energy terms for reduced protein models implemented in an off-lattice force field. *J. Comput. Chem.* **22**, 1229–1242 (2001).

66. MacGregor, E. A., Janeček, Š. & Svensson, B. Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta* **1546**, 1–20 (2001).

67. Fang, T.-Y., Tseng, W.-C., Shih, T.-Y. & Wang, M.-Y. Identification of the essential catalytic residues and selectivity-related residues of maltooligosyltrehalose trehalohydrolase from the thermophilic archaeon *Sulfolobus solfataricus* ATCC 35092. *J. Agric. Food Chem.* **56**, 5628–5633 (2008).

68. Kuroki, R., Ito, Y., Kato, Y. & Imoto, T. A covalent enzyme-substrate adduct in a mutant hen egg white lysozyme (D52E). *J. Biol. Chem.* **272**, 19976–19981 (1997).

69. Kondo, K., Miura, Y., Sone, H., Kobayashi, K. & Iijima, H. High-level expression of a sweet protein, monellin, in the food yeast *Candida utilis*. *Nat. Biotechnol.* **15**, 453–457 (1997).

70. Kobayashi, K., Kato, M., Miura, Y., Kettoku, M., Komeda, T. & Iwamatsu, A. Gene cloning and expression of new trehalose-producing enzymes from the hyperthermophilic archaeum *Sulfolobus solfataricus* KM1. *Biosci., Biotechnol., Biochem.* **60**, 1882–1885 (1996).

71. Adachi, M., Ohhara, T., Kurihara, K., Tamada, T., Honjo, E., Okazaki, N. & Arai, S. Structure of HIV-1 protease in complex with potent inhibitor KNI-272

determined by high-resolution X-ray. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 4641–4646 (2009).

72. Tamada, T., Kinoshita, T., Kurihara, K., Adachi, M., Ohhara, T., Imai, K., Kuroki, R. & Tada, T. Combined high-resolution neutron and X-ray analysis of inhibited elastase confirms the active-site oxyanion hole but rules against a low-barrier hydrogen bond. *J. Am. Chem. Soc.* **131**, 11033–11040 (2009).

73. Okazaki, N., Kumei, M., Manzoku, M., Kuramitsu, S., Shirouzu, M., Shinkai, A. & Yokoyama, S. Structure of a UPF0150-family protein from *Thermus thermophilus* HB8. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **63**, 173–177 (2007).

74. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-jones, S., Howe, K. L., Marshall, M. & Sonnhammer, E. L. L. The Pfam Protein Families Database. *Nucleic Acids Res.* **30**, 276–280 (2002).

75. Bessman, M. J., Frick, D. N. & O'Handley, S. F. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.* **271**, 25059–25062 (1996).

76. Lin, J., Abeygunawardana, C., Frick, D. N., Bessman, M. J. & Mildvan, A. S. Solution structure of the quaternary MutT-M2+-AMPCPP-M2+ complex and mechanism of its pyrophosphohydrolase action. *Biochemistry* **36**, 1199–1211 (1997).

77. Ramakrishnan, V. & White, S. W. The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature* **358**, 768–771 (1992).

78. Bycroft, M., Grunert, S., Murzin, A. G., Proctor, M. & Johnston, D. S. NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5 Hy2. *EMBO J.* **14**, 3563–3571 (1995).

79. Kharrat, A., Macias, M. J., Gibson, T. J., Nilges, M. & Pastore, A. Structure of the dsRNA binding domain of *E. coli* RNase III. *EMBO J.* **14**, 3572–3584 (1995).

80. Myles, D. A. A. Neutron protein crystallography: current status and a brighter future. *Curr. Opin. Struct. Biol.* **16**, 630–637 (2006).

81. Yokota, T., Tonozuka, T., Shimura, Y., Ichikawa, K., Kamitori, S. & Sakano, Y. Structures of *Thermoactinomyces vulgaris* R-47 α-amylase II complexed with substrate analogues. *Biosci., Biotechnol., Biochem.* **65**, 619–626 (2001).

82. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637 (1983).

83. Barlow, D. J. & Thornton, J. M. Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867–885 (1983).

84. Hashimoto, H., Inoue, T., Nishioka, M., Fujiwara, S., Takagi, M. & Kai, Y. Hyperthermostable protein structure maintained by intra and inter-helix

ion-pairs in archaeal O6-methylguanine-DNA methyltransferase. *J. Mol. Biol.* **292**, 707–716 (1999).

85. Chan, C.-H., Wilbanks, C. C., Makhatadze, G. I. & Wong, K.-B. Electrostatic contribution of surface charge residues to the stability of a thermophilic protein: benchmarking experimental and predicted pKa values. *PloS one* **7**, e30296 (2012).

86. Ishikawa, K., Nakamura, H., Morikawa, K. & Kanaya, S. Stabilization of *Escherichia coli* ribonuclease HI by cavity-filling mutations within a hydrophobic core. *Biochemistry* **32**, 6171–6178 (1993).

87. Marqusee, S. & Baldwin, R. L. Helix stabilization by Glu-...Lys+ salt bridges in short peptides of de novo design. *Proc. Natl. Acad. Sci. U. S. A.* **84**, 8898–8902 (1987).

88. Okazaki, N., Arimori, T., Nakazawa, M., Miyatake, K., Ueda, M. & Tamada, T. Crystallization and preliminary X-ray diffraction studies of the catalytic domain of a novel chitinase, a member of GH family 23, from the moderately thermophilic bacterium *Ralstonia* sp. A-471. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **67**, 494–497 (2011).

89. Kuroki, R., Okazaki, N., Adachi, M., Ohhara, T., Kurihara, K. & Tamada, T. Towards investigation of the inhibitor-recognition mechanisms of drug-target proteins by neutron crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 1126–1130 (2010).

90. Okazaki, N., Tamada, T., Feese, M. D., Kato, M., Miura, Y., Komeda, T., Kobayashi, K., Kondo, K., Blaber, M. & Kuroki, R. Substrate recognition mechanism of a glycosyltrehalose trehalohydrolase from *Sulfolobus solfataricus* KM1. *Protein Sci.* **21**, 539–552 (2012).

91. Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. An automated system for micro-batch protein crystallization and screening. *J. Appl. Crystallogr.* **23**, 297–302 (1990).

92. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 849–861 (1999).

93. Terwilliger, T. C. Map-likelihood phasing. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 1763–1775 (2001).

94. Morris, R. J., Perrakis, A. & Lamzin, V. S. ARP/wARP and automatic interpretation of protein electron density maps. *Methods Enzymol.* **374**, 229–244 (2003).

95. Laskowski, R. A., Moss, D. S. & Thornton, J. M. Main-chain bond lengths and bond angles in protein structures. *J. Mol. Biol.* **231**, 1049–1067 (1993).

96. Ueda, M., Ohata, K., Konishi, T., Sutrisno, A., Okada, H., Nakazawa, M. & Miyatake, K. A novel goose-type lysozyme gene with chitinolytic activity from the moderately thermophilic bacterium *Ralstonia* sp. A-471: cloning, sequencing, and expression. *Appl. Microbiol. Biotechnol.* **81**, 1077–1085 (2009).

97. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

98. Weaver, L. H., Grütter, M. G. & Matthews, B. W. The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the "goose-type" lysozymes lack a catalytic aspartate residue. *J. Mol. Biol.* **245**, 54–68 (1995).

99. Avila-Sakar, A. J., Creutz, C. E. & Kretsinger, R. H. Crystal structure of bovine annexin VI in a calcium-bound state. *Biochim. Biophys. Acta* **1387**, 103–116 (1998).

100. Garau, G., Bebrone, C., Anne, C., Galleni, M., Frère, J.-M. & Dideberg, O. A metallo-β-lactamase enzyme in action: crystal structures of the monozinc carbapenemase CphA and its complex with biapenem. *J. Mol. Biol.* **345**, 785–795 (2005).

101. Altschul, S. F., Madden, T. L., Schäffer, a a, Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

102. Yoshiba, S., Nakagawa, N., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S. & Kuramitsu, S. Overproduction, crystallization and preliminary diffraction data of ADP-ribose pyrophosphatase from *Thermus thermophilus* HB8. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **D59**, 1840–1842 (2003).

103. Tanaka, I., Kurihara, K., Chatake, T. & Niimura, N. A high-performance neutron diffractometer for biological crystallography (BIX-3). *J. Appl. Crystallogr.* **35**, 34–40 (2002).

104. Adams, P. D., Mustyakimov, M., Afonine, P. V & Langan, P. Generalized X-ray and neutron crystallographic analysis: more accurate and complete structures for biological macromolecules. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **65**, 567–573 (2009).

105. McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* **9**, 401–406 (2002).

106. Niimura, N. & Bau, R. Neutron protein crystallography: beyond the folding structure of biological macromolecules. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **64**, 12–22 (2008).

107. Kurihara, K., Tanaka, I., Chatake, T., Adams, M. W. W., Jenney, F. E., Moiseeva, N., Bau, R. & Niimura, N. Neutron crystallographic study on rubredoxin from Pyrococcus furiosus by BIX-3, a single-crystal diffractometer for biomacromolecules. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 11215–11220 (2004).

108. Wlodawer, A. & Hendrickson, W. A. A procedure for joint refinement of macromolecular structures with X-ray and neutron diffraction data from single crystals. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **A38**, 239–247 (1982).

109. Adams, P. D., Afonine, P. V, Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213–221 (2010).

110. McRee, D. E. A visual protein crystallographic software system for X11/Xview. *J. Mol. Graphics* **10**, 44–46 (1992).

111. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr., Sect. A: Found. Crystallogr.* **64**, 112–122 (2008).

112. Kinoshita, T., Tamada, T., Imai, K., Kurihara, K., Ohhara, T., Tada, T. & Kuroki, R. Crystallization of porcine pancreatic elastase and a preliminary neutron diffraction experiment. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **63**, 315–317 (2007).

113. Yamada, H., Tamada, T., Kosaka, M., Miyata, K., Fujiki, S., Tano, M., Moriya, M., Yamanishi, M., Honjo, E., Tada, H., Ino, T., Yamaguchi, H., Futami, J., Seno, M., Nomoto, T., Hirata, T., Yoshimura, M. & Kuroki, R. "Crystal lattice engineering ," an approach to engineer protein crystal contacts by creating

intermolecular symmetry: crystallization and structure determination of a mutant human RNase 1 with a hydrophobic interface of leucines. *Protein Sci.* **16,** 1389–1397 (2007).

List of publications

1. **Structure of a UPF0150-family protein from *Thermus thermophilus* HB8**

   Nobuo Okazaki, Maki Kumei, Miho Manzoku, Seiki Kuramitsu, Mikako Shirouzu, Akeo Shinkai, Shigeyuki Yokoyama

   Acta Cryst., **F63**, 173–177 (2007).

2. **Mail-in data collection at SPring-8 protein crystallography beamlines**

   Nobuo Okazaki, Kazuya Hasegawa, Go Ueno, Hironori Murakami, Takashi Kumasaka, Masaki Yamamoto

   J. Synchrotron Radiat., **15**, 288–291 (2008).

3. **Towards investigation of the inhibitor-recognition mechanisms of drug-target proteins by neutron crystallography**

   Ryota Kuroki, Nobuo Okazaki, Motoyasu Adachi, Takashi Ohhara, Kazuo Kurihara, Taro Tamada

   Acta Cryst., **D66**, 1126–1130 (2010).

4. **Crystallization and preliminary X-ray diffraction studies of a novel chitinase, a member of GH family 23, from the moderately thermophilic bacterium *Ralstonia* sp. A-471**

   Nobuo Okazaki, Takao Arimori, Masami Nakazawa, Kazutaka Miyatake, Mitsuhiro Ueda, Taro Tamada

   Acta Cryst., **F67**, 494–497 (2011).

5. **Crystallization and preliminary neutron diffraction studies of ADP-ribose pyrophosphatase-I from *Thermus thermophilus* HB8**

Nobuo Okazaki, Motoyasu Adachi, Taro Tamada, Kazuo Kurihara, Takushi Ooga, Nobuo Kamiya, Seiki Kuramitsu, Ryota Kuroki

Acta Cryst., **F68**, 49–52 (2012).

6. **Substrate recognition mechanism of a glycosyltrehalose trehalohydrolase (GTHase) from *Sulfolobus solfataricus* KM1**

Nobuo Okazaki, Taro Tamada, Michael D. Feese, Masaru Kato, Yutaka Miura, Toshihiro Komeda, Kazuo Kobayashi, Keiji Kondo, Michael Blaber, Ryota Kuroki

Protein Science, **21**, 539–552 (2012).

List of supplementary publications

1. **Mechanism of metal activation of human hematopoietic prostaglandin D synthase**

   Tsuyoshi Inoue, Daisuke Irikura, Nobuo Okazaki, Shigehiro Kinugasa, Hiroyoshi Matsumura, Nobuko Uodome, Masaki Yamamoto, Takashi Kumasaka, Masashi Miyano, Yasushi Kai, Yoshihiro Urade

   Nat. Struct. Biol., **10**, 291–296 (2003). Erratum in: *Nat. Struct. Biol.*, **10**, 409 (2003).

2. **First determination of the inhibitor complex structure of human hematopoietic prostagrandin D synthase**

   Tsuyoshi Inoue, Yousuke Okano, Yuji Kado, Kousuke Aritake, Daisuke Irikura, Nobuko Uodome, Nobuo Okazaki, Shigehiro Kinugasa, Hideyuki Shishitani, Hiroyoshi Matumura, Yasushi Kai, Yoshihiro Urade

   J. Biochem., **135**, 279–283 (2004).

3. **Purification, crystallization and preliminary crystallographic analysis of the glycine-cleavage system component T-protein from *Pyrococcus horikoshii* OT3**

   Neratur K. Lokanath, Chizu Kuroishi, Nobuo Okazaki, Naoki Kunishima

   Acta Cryst., **D60**, 1450–1452 (2004).

4. **Crystal structure of a component of glycine cleavage system: T-protein from *Pyrococcus horikoshii* OT3 at 1.5 Å resolution**

   Neratur K. Lokanath, Chizu Kuroishi, Nobuo Okazaki, Naoki Kunishima

   Proteins, **58**, 769–773 (2005).

5. **Crystal structure of novel NADP-dependent 3-hydroxyisobutyrate dehydrogenase from *Thermus thermophilus* HB8**

   Neratur K. Lokanath, Noriyasu Ohshima, Koji Takio, Ikuya Shiromizu, Chizu Kuroishi, Nobuo Okazaki, Seiki Kuramitsu, Shigeyuki Yokoyama, Masashi Miyano, Naoki Kunishima

   J. Mol. Biol., **352**, 905–917 (2005).

6. **Crystal structure of alanyl-tRNA synthetase editing-domain homolog (PH0574) from a hyperthermophile, *Pyrococcus horikoshii* OT3 at 1.45 Å resolution**

   Jun Ishijima, Yumiko Uchida, Chizu Kuroishi, Chizuru Tuzuki, Naoko Takahashi, Nobuo Okazaki, Katsuhide Yutani, Masashi Miyano

   Proteins, **62**, 1133–1137 (2006).

7. **Structure of the stand-alone RAM-domain protein from *Thermus thermophilus* HB8**

   Noboru Nakano, Nobuo Okazaki, Shinya Satoh, Koji Takio, Seiki Kuramitsu, Akeo Shinkai, Shigeyuki Yokoyama

   Acta Cryst., **F62**, 855–860 (2006).

8. **Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography**

   Motoyasu Adachi, Takashi Ohhara, Kazuo Kurihara, Taro Tamada, Eijiro Honjo, Nobuo Okazaki, Shigeki Arai, Yoshinari Shoyama, Kaname Kimura, Hiroyoshi Matsumura, Shigeru Sugiyama, Hiroaki Adachi, Kazufumi Takano, Yusuke Mori, Koushi Hidaka, Tooru Kimura, Yoshio Hayashi, Yoshiaki Kiso, Ryota Kuroki

   Proc. Natl. Acad. Sci. U. S. A., **106**, 4641–4646 (2009).

9. **A structural mechanism for dimeric to tetrameric oligomer conversion in *Halomonas* sp. nucleoside diphosphate kinase**

   Shigeki Arai, Yasushi Yonezawa, Nobuo Okazaki, Fumiko Matsumoto, Taro Tamada, Hiroko Tokunaga, Matsujiro Ishibashi, Michael Blaber, Masao Tokunaga, Ryota Kuroki

   Protein Science, **21**, 498–510 (2012).

10. **Human hematopoietic prostaglandin D synthase inhibitor complex structures**

    Yuji Kado, Kousuke Aritake, Nobuko Uodome, Yousuke Okano, Nobuo Okazaki, Hiroyoshi Matsumura, Yoshihiro Urade, Tsuyoshi Inoue

    J. Biochem., **151**, 447–455 (2012).

List of supplementary publications in Japanese

1. **タンパク質結晶化の自動化の試み－タンパク質結晶をより理論的に進める未知はあるか？－**

   菅原光明, 仲村勇樹, 黒石千寿, 浮田陽子, 岡﨑伸生, 粂井麻希, 田中智之, 小林正則, 高秀幸, 宮野雅司

   日本結晶成長学会誌 **30**, 395–397 (2003).

2. **最新機器への挑戦－全自動タンパク質結晶化観察ロボット TERA**

   菅原光明, 岡﨑伸生, 宮野雅司

   化学 **59**, 52–53 (2004).

3. **施設の現状と進展 構造生物**

   長谷川和也, 酒井久伸, 岡﨑伸生, 清水伸隆, 河本正秀

   SPring-8 年報 86–89 (2005).

4. **施設の現状と進展 BL38B1 構造生物学 III**

   長谷川和也, 酒井久伸, 岡﨑伸生, 清水伸隆, 河本正秀, 熊坂崇, 山本雅貴

   SPring-8 年報 92–94 (2006).

5. **施設の現状と進展 BL38B1 構造生物学 III**

   馬場清喜, 岡﨑伸生, 長谷川和也, 清水伸隆, 酒井久伸, 河本正秀, 山本雅貴, 熊坂崇

   SPring-8 年報 91–93 (2007).

6. **ISILON IQ ユーザー事例**

酒井久伸, 長谷川和也, 馬場清喜, 岡崎伸生, 清水伸隆, 河本正秀, 熊坂崇, 山本雅貴

http://www.isilon.co.jp/wp-content/uploads/2010/08/spring8.pdf

7. **中性子回折による蛋白質単結晶の構造解析**

栗原和男, 岡崎伸生, 黒木良太

Radioisotopes , **59**, 263–77 (2010).

Acknowledgements

The author wishes to express his special gratitude to Dr. Ryota Kuroki and Dr. Taro Tamada for their interesting suggestion and cooperation. Studies in Chapter I, II, and a part of III were achieved as the work at Japan Atomic Energy Agency (JAEA; Ibaraki) from 2008 to 2012. The author sincerely appreciates Prof. Michael Blaber, Dr. Kazuo Kurihara, Dr. Motoyasu Adachi, Dr. Takashi Ohhara, Dr. Shigeki Arai, Dr. Takao Arimori, Dr. Michael D. Feese, and the colleagues of Molecular Structural Biology Group, Quantum Beam Science Directorate, JAEA. The author also thanks Ms. Kyoko Takehara for her excellent technical assistance in Chapter II. The author is grateful to the great supports of the beamline staffs of SPring-8 and Photon Factory. The author also indebted to the members of the SOSHO project at Osaka University, especially Associate Prof. Yoshihiro Matsumura, Dr. Hiroaki Adachi, and Prof. Yusuke Mori. The author is grateful to Associate Prof. Takayoshi Kinoshita and Prof. Toshiji Tada of Osaka Prefecture Univ. for study in Chapter III. Development of *D-Cha* was achieved in Japan Synchrotron Radiation Research Institute (JASRI; Hyogo) from 2005 to 2008. The author is grateful to Dr. Masaki Yamamoto (RIKEN), Dr. Takashi Kumasaka, Dr. Kazuya Hasegawa, Dr. Seiki Baba, Dr. Masahide Kawamoto, Dr. Nobutaka Shimizu, Dr. Go Ueno (RIKEN), Mr. Hironori Murakami (RIKEN), Ms. Izumi Wada, and the colleagues in JASRI Structural Biology Group and RIKEN Research Infrastructure Group. The author also would like to thank all *D-Cha* users for very useful suggestions, which led to the system improvements. Study of the TTHA0281 protein in Chapter III was supported by the RIKEN Structural Genomics/Proteomics Initiative and carried out as a part of the Protein 3000 project from 2003 to 2006. The author thanks Dr. Koji Takio for the gel-filtration analysis of the TTHA0281 protein. The author is grateful to Prof.

Shigeyuki Yokoyama, Prof. Seiki Kuramitsu, Dr. Akeo Shinkai, Prof. Neratur K. Lokanath, Dr. Mitsuaki Sugahara, Ms. Maki Kumei, Mr. Yuki Nakamura, Mr. Tomoyuki Tanaka, Ms. Keiko Sakamoto, Ms. Aki Sasayama, and the colleagues in RIKEN High-throughput Factory.

Special thanks are given to Prof. Tsuyoshi Inoue (Osaka Univ.), Prof. Takashi Hayashi (Osaka Univ.), and Prof. Hiroshi Uyama (Osaka Univ.) for their enthusiastic guidance, helpful suggestion, and great encouragement. The author sincerely appreciates Prof. Yasushi Kai (Fukui Univ. of Tech.), Dr. Nobuko Kanehisa (Osaka Univ.), Associate Prof. Hiroshi Hashimoto (Yokohama City Univ.), Assistant Prof. Eiichi Mizohata (Osaka Univ.), Dr. Yuji Kado (Osaka Univ.), Ms. Miho Takeichi (Osaka Univ.), and all colleagues work with the author for interesting suggestion, helpful advice, and supports. The author thanks to the great associates for their welcome distraction and encouragement: Dr. Hiroshi Nakagawa (JAEA), Ms. Mizue Meguro (Waseda Univ.), Dr. Fumiko Matsumoto (JAEA), Dr. Mitsugu Yamada (JAEA), Ms. Rumi Shimizu (JAEA), Dr. Kunio Hirata (RIKEN), Dr. Atsushi Nisawa (RIKEN), and many colleagues.

Last but not least, the author also wants to thank to Hikari, Michiru, and family, who inspired, encouraged and fully-supported.