STRUCTURAL STUDIES ON THERMOSTABLE PROTEINS FROM HYPERTHERMOPHILIC ARCHAEON *PYROCOCCUS KODAKARAENSIS* KOD1

HIROSHI HASHIMOTO

OSAKA UNIVERSITY

2000

PREFACE

==========

This study was carried out under the direction of Professor Yasushi Kai at the Department of Materials Chemistry, Graduate School of Engineering, Osaka University, Japan, from 1995 to 2000. The interest of this study is focused on the structure & function relationship of the thermostable proteins from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1.

The author wishes to express his sincere gratitude to Professor Yasushi Kai for his giving the author a chance to study on this field, kind guidance, helpful suggestion, and continuous encouragement throughout this work. The author is also deeply grateful to Dr. Tsuyoshi Inoue for his continuous advice, stimulating discussions, and hearty encouragement. He acknowledges Dr. Nobuko Kanehisa and Ms. Eiko Mochizuki for their encouragement and helpful suggestions. The author is also grateful to Associate Professor Masahiro Takagi, Dr. Shinsuke Fujiwara, Mr. Motomu Nishioka, Mr. Toru Yuasa (Department of Biotechnology, Osaka University) and Professor Tadayuki Imanaka (Department of Synthetic Chemistry and Biological Chemistry, Kyoto University) for stimulating discussion, and continuous guidance.

Special thanks are given to the author's co-workers, Mr. Tomoya Matsumoto (April 1997 — March 1998), Ms. Shoko Takeuchi (April 1998 — March 1999) and Ms. Hitomi Takahashi for their continuous and fruitful cooperation. He also wishes to thank Dr. Genji Kurisu, Dr. Naoki Shibata, Dr. Chunmin Li, Dr. Hajime Sugawara, Mr. Nobuya Nishio, and all others members of the research group of Professor Yasushi Kai for their helpful assistance, occasional discussions, and profound interests. The author also acknowledges to Professor Noriyoshi Sakabe at University of Tsukuba, Dr. Nobuhisa Watanabe, Dr. Mamoru Suzuki, and Dr. Noriyuki Igarashi at High Energy Accelerator Research Organization. Furthermore, the author thanks the JSPS Fellowship for Japanese Junior Science for financial support. Finally, the author would like to express his thanks to his parents, sister, brother, and friends for their understandings and hearty encouragement.

January 2000

Hiroshi HASHIMOTO

CONTENTS

==========

PREFACE

GENERAL INTRODUCTION --- 1

CHAPTER 1

Structural studies on family B DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1

1.1. INTRODUCTION --- 3

1.2. MATERIALS AND METHODS --- 6

1.3. RESULTS AND DISCUSSION --- 9

1.4. REFERENCES --- 18

CHAPTER 2

Structural study of O^6-methylguanine-DNA methyltransferase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1

2.1. INTRODUCTION --- 23

2.2. MATERIALS AND METHODS --- 26

2.3. RESULTS AND DISCUSSION --- 41

2.4. REFERENCES --- 64

CONCLUSION --- 73

LIST OF PUBLICATIONS --- 74

SUPPLEMENTARY PUBLICATIONS --- 75

GENERAL INTRODUCTION

==========

All living organisms in the world are divided into three primary kingdoms; Eucarya, Bacteria and Archaea. Archaea live in various extreme environments and can be classified as thermophiles, halophiles and methanogens. Archaea are thought to include the most ancient organisms on earth and to have some biological features of both Bacteria and Eucarya. Not surprisingly, the proteins derived from thermophiles are extremely thermostable. It is said that the enhancement of protein thermostability is an important goal of protein engineering. Although many successful examples of stabilization of protein exist, the molecular mechanism of thermostabilization is far from systematic understanding based on three dimensional structures. Structural studies on hyperthermophilic archaeal proteins may reveal the structural features essential for stability under the extreme environments in which these organisms live. The studies may also provide the structural basis required for stabilizing various mesophilic proteins for industrial applications.

The main object of the present studies is to make clear the structural features of the proteins related DNA replication and

repair, obtained from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1. This thesis consists of two chapters.

Chapter 1 deals with structural study of family B DNA polymerase from *Pyrococcus kodakaraensis* KOD1.

Chapter 2 deals with structural study of O^6-methylguanine-DNA methyltransferase from *Pyrococcus kodakaraensis* KOD1.

CHAPTER 1

==========

Structural studies on family B DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1

1.1. INTRODUCTION

DNA polymerases are a group of enzymes that use single-stranded DNA as a template for the synthesis of the complementary DNA strand. These enzymes have multifunction, synthetic mode (polymerase) and one or two degradative modes (5′-3′ and / or 3′-5′ exonucleases) and play an essential role in nucleic acid metabolism, including the process of DNA replication, repair and recombination. More than 50 DNA polymerase genes have been cloned and sequenced. Amino acid sequences deduced from their nucleotide sequences can be classified into four major types: *Escherichia coli* DNA polymerase I (family A), *E. coli* DNA polymerase II (family B), *E. coli* DNA polymerase III (family C) and others (family X) (Braithwaite & Ito, 1993). DNA polymerases from family B are called α like DNA polymerases because they have the amino acid sequence of conserved

eukaryotic DNA polymerase α (Bernad *et al.*, 1987).

In this paper, we report the crystallization and the preliminary X-ray diffraction analysis of family B DNA polymerase obtained from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1 (KOD DNA polymerase). *Pyrococcus kodakaraensis* KOD1 was isolated from a solfataric hot sprig of Kodakara Island, Kagoshima, Japan. This strain is one of the most thermostable organisms known, with an optimum growth temperature of 95°C (Morikawa *et al.*, 1994). Enzymes produced in this strain were reported to be extremely thermostable and have eukaryotic characteristics (Fujiwara *et al.*, 1996). Enzymes from hyperthermophiles are extremely thermostable and have an industrial importance. KOD DNA polymerase's optimum temperature (75 °C) is similar to those of DNA polymerase obtained from *Pyrococcus furiosus* (*Pfu* DNA polymerase), however, the KOD DNA polymerase exhibites an extension rate (100 to 130 nucleotides/sec) 5 times higher and a processivity 10 to 15 times higher than those of *Pfu* DNA polymerase (Takagi *et al.*, 1997). KOD DNA polymerase is, therefore, an enzyme with considerable biological, technological, and economic importance. Thermostable DNA polymerases are widely used in carrying out polymerase chain reaction (PCR) experiments.

The crystal structures of three thermostable DNA polymerases, *Thermus aquaticus* DNA polymerase (*taq* polymerase)(Kim *et al.*, 1995), Klentaq1 (Korolev *et al.*, 1995), a fragment analogous to Klenow fragment and *Bacillus stearothermophilus* large fragment DNA polymerase I (Kiefer *et al.*, 1997), and crystal structure of DNA polymerase α (family B) from bacteriophage RB69 (Wang *et al.*, 1997) have been solved at 3.0, 2.4, 2.1 and 2.7 Å resolution, respectively. Except for RB69 DNA polymerase, those enzymes have no similarity with KOD DNA polymerase. RB69 DNA polymerase have conserved amino acid regions with archaeal DNA polymerases (Wang *et al.*, 1997). But no information of three dimensional structures for archaeal DNA polymerase has ever been reported. The structure of KOD DNA polymerase may provide useful information for increasing the thermostability of enzymes already in use, or for designing new hyperthermostable proteins. Furthermore, structural comparison of DNA polymerases from various species may give evolutional knowledge, because DNA polymerase is essential for DNA replication and all organisms have their own DNA polymerases. Most recently, three archaeal DNA polymerases were crystallized (Goldman *et al.*, 1998; Nastopoulos *et al.*, 1998; Zhou *et al.*, 1998). Structural comparison of these archaeal DNA polymerases may reveal evolutional

relationships among these species.

1.2. MATERIALS AND METHODS

KOD DNA polymerase was overexpressed in *E. coli* strain BL21(DE3) and purified by the previously reported method (Takagi *et al.*, 1997). KOD DNA polymerase was concentrated up to about 25 Abs cm^{-1} at 280 nm in a Tris buffer at pH 8.5 containing 0.1 mM EDTA and 1 mM DTT using Centricon-50 concentrators (Amicon, Beverly, MA). All the crystallization experiments were carried out using the hanging drop vapor diffusion technique. The droplet (typically 4 μL) was prepared by mixing equal volumes (2 μL) of the protein and reservoir solutions. A preliminary screen of crystallization conditions was performed using the Crystal Screen kit (Hampton Research, Laguna Hills, CA, USA). And then, a number of precipitants including salts, polyethylene glycol solutions, and organic solvents were explored over a broad pH range from 4.0 to 10.0.

Crystals of KOD DNA polymerase suitable for diffraction experiments were obtained at 293 K with hanging drops of 2 μL of protein solution and 2 μL of reservoir solution containing 100 mM Na citrate buffer pH 5.5 and 25 ~ 30% (*v*/*v*) 2-methyl-2,4-pentanediol (MPD), equilibrated against the

reservoir solution. Crystals were thin plates in shape and with a maximum size of 0.4 x 0.1 x 0.1 mm in several months (**Figure 1**). Several crystals were washed, dissolved and analyzed by SDS-PAGE. The analysis showed that these crystals were protein with molecular weight of about 90 kDa, resulting in the protein free from the unexpected proteolysis.

All the X-ray diffraction measurements were carried out with λ = 1.00 Å synchrotron X-ray monochromated by Si(111) double mirror coupled with Weissenberg camera for macromolecule and BAS2000 imaging-plate system (Fuji Film Company Ltd.) in KEK-PF, Japan (Sakabe *et al.*, 1995; Watanabe *et al.*, 1995). The full X-ray diffraction data were processed and scaled with programs *DENZO* and *SCALEPACK* (Otwinowski *et al.*, 1997) and systematic extinctions in the intensity data were checked by program *HKLPLOT* (Eleanor Dodson, unpublished; Collaborative Computational Project, Number 4, 1994).

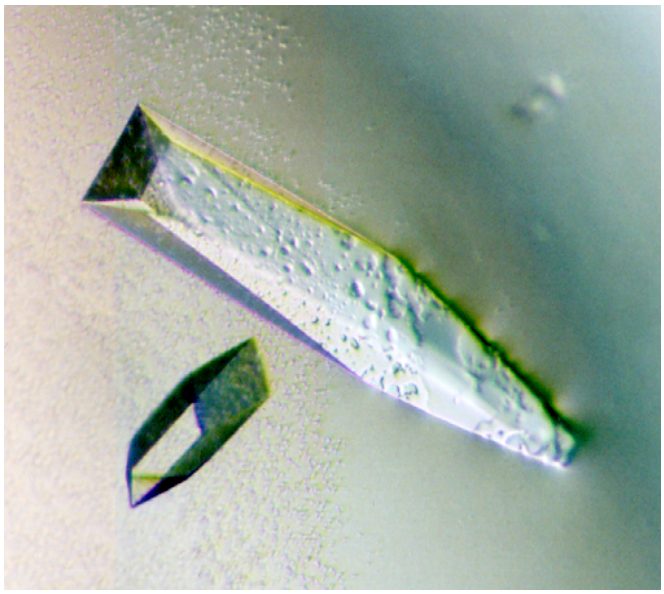

Figure 1.

Crystals of KOD DNA polymerase with maximum dimensions 0.4 x 0.1x 0.1 mm.

1.3. RESULTS AND DISCUSSION

When a KOD DNA polymerase crystal normally was mounted in a glass capillary and exposed to X-ray from synchrotron radiation, the diffraction spots were observed the Bragg spacings up to 3.0

Å. The space group was determined to be $P2_12_12_1$ with the unit-cell parameters of a = 112.8, b = 115.4 and c = 75.4 Å. The asymmetric unit contains a single molecule with a mass of 89,500, giving a crystal volume per protein mass (V_m) of 2.74 $Å^3Da^{-1}$ and the solvent content of 54.8% by volume (Matthews, 1968). As listed in **Table 1 (above)**, a full data set was collected from a single crystal, although the crystal was slightly decayed in later frames. It consists of 73,785 measurements of 18,926 unique observed reflections with an overall R_{merge} of 9.4% ($R_{merge} = \Sigma|I-<I>|/\Sigma I$) and overall $I/sigI$ of 6.6. This represents 93.0% of theoretically observable reflections at 3.0 Å resolution. The outermost shell of data between 3.11 and 3.00 Å is 83.7% complete.

Fortunately, crystallization condition of KOD DNA polymerase contains 25 ~ 30% MPD as a precipitant. MPD is widely used as one of conventional anti-freezing reagents in cryogenic protein crystallography. In this case, crystals can be cooled by liquid N_2 without dialysis against buffer solution containing anti-freezing reagents. The KOD DNA polymerase crystal was directly picked up with a nylon fiber loop from a drop of mother liquid, then the crystal was rapidly transferred to the liquid N_2 gas stream. Diffraction intensities at 100 K were observed beyond 3.0 Å

resolution (**Figure 2**). The crystal was not at all decayed during data collection under synchrotron radiation X-ray and diffracted more than at 290 K. Unit-cell parameters were determined as a = 111.9, b = 112.4 and c = 73.9 Å with the space group of $P2_12_12_1$. Cell parameters were slightly shortened by cooling. The unit-cell parameters gave Matthew's coefficient of 2.60 $Å^3Da^{-1}$ and the solvent content of 52.2% by volume. As listed in **Table 1 (below)**, a full data set was collected from one crystal. It consists of 61,509 measurements of 17,483 unique reflections with an overall R_{merge} of 8.5% and overall I/sig$I$ of 7.6. This represents 90.6% of theoretically observable reflections at 3.0 Å resolution. The outermost shell of data between 3.11 and 3.00 Å is 86.1% complete. During this experiment, diffraction from the frozen crystal have not been decayed and variation of mosaicity of the crystal are small. Therefore, data set at 100K was used at next step.

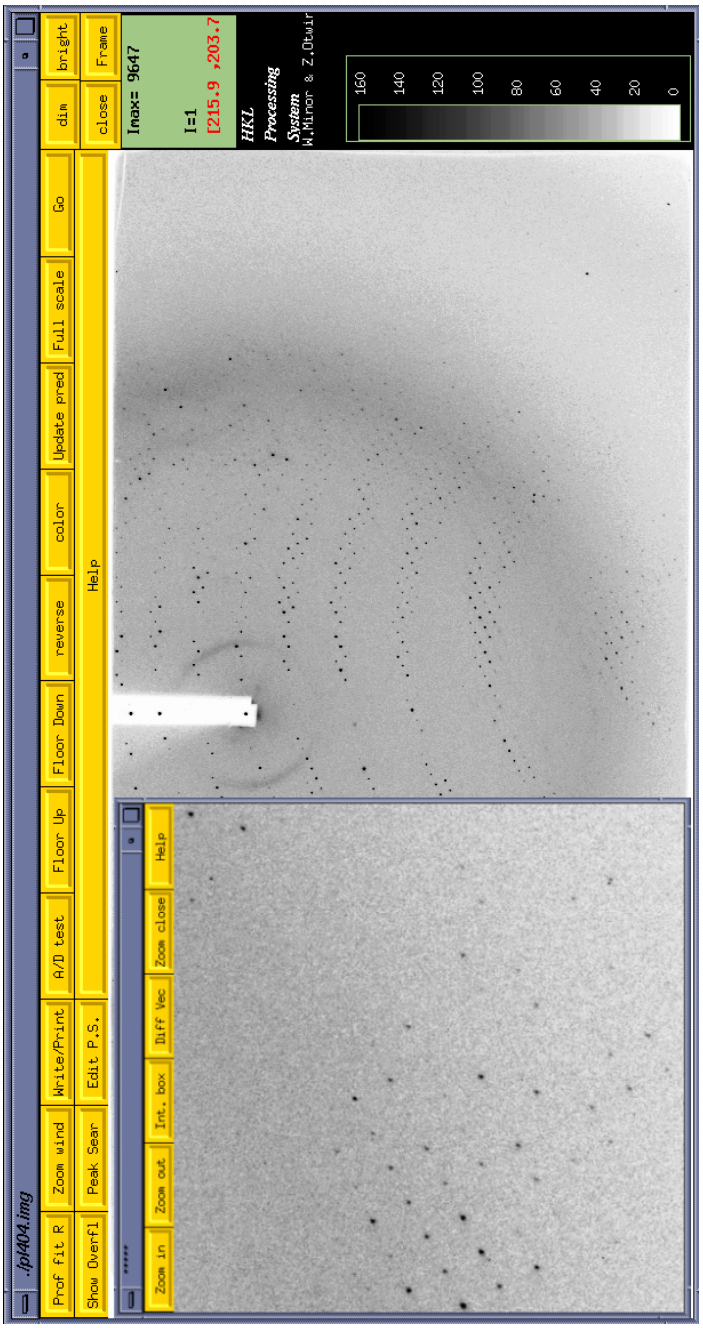

Figure 2.

Diffraction image of KOD DNA polymerase at 100 K with synchrotron radiation X-rays at Beamline 18B at KEK-PF, Japan.

Table 1.

Diffraction data statistics of a KOD DNA polymerase crystal at 290 K (above) and

100 K (below), processed and scaled with programs *DENZO* and *SCALEPACK*.

Diffraction data statistics at 290 K.

Resolution (Å)	No. of reflections (unique)	Completeness (%)	R_{merge}
40.00 - 6.46	2,123	97.0	0.039
6.46 - 5.13	2,041	98.4	0.074
5.13 - 4.48	2,006	98.5	0.082
4.48 - 4.07	1,989	97.5	0.114
4.07 - 3.78	1,922	96.0	0.165
3.78 - 3.56	1,904	94.2	0.228
3.56 - 3.38	1,824	91.1	0.290
3.38 - 3.23	1,733	87.3	0.342
3.23 - 3.11	1,717	85.5	0.389
3.11 - 3.00	1,667	83.7	0.411
overall	18,926	93.0	0.094

Diffraction data statistics at 100 K.

Resolution (Å)	No. of reflections (unique)	Completeness (%)	R_{merge}
40.00 - 6.46	1,902	91.4	0.042
6.46 - 5.13	1,819	92.9	0.071
5.13 - 4.48	1,817	93.8	0.070
4.48 - 4.07	1,777	92.6	0.089
4.07 - 3.78	1,769	91.9	0.117
3.78 - 3.56	1,718	90.0	0.153
3.56 - 3.38	1,712	90.5	0.191
3.38 - 3.23	1,687	89.0	0.230
3.23 - 3.11	1,650	87.5	0.279
3.11 - 3.00	1,632	86.1	0.316
overall	17,483	90.6	0.085

The crystal structure of KOD DNA polymerase was solved by the method of the molecular replacement using program *AMoRe* (Navaza, 1994). Recently, crystal structure of family B DNA polymerase from hyperthermophilic archaeon *Thermococcus gorgonarius* (*Tgo* DNA polymerase) was reported (Hopfner *et al.*, 1999). The coordinates of *Tgo* DNA polymerase (1tgo) reduced to polyalanine was used as the search model. Data in the resolution range of 20 ~ 3.5 Å were used in both the rotation and translation functions. Results are discussed in terms of an *AMoRe* correlation coefficient (CC). Using a Patterson cut-off radius of 36 Å, a list of 20 rotation function peaks was obtained, with the top peak having an AMoRe CC of 13.8. The top solution by translation function is CC of 43.3 with *R*-factor of 54.1%. The model was manually modified using programs O and subjected to further rounds of refinement using data in the resolution range 40 – 3.0 Å with program *CNS* (Brünger *et al.*, 1998). **Figure 3** shows the overall structure of KOD DNA polymerase. The final *R*-factor is 22.1% and R_{free} is 31.1%, with r.m.s. deviations for bond lengths and bond angles being 0.008 Å and 1.12°, respectively. The fifty residues at the tip of thumb domain are not included in the final model due to poorly defined electron density.

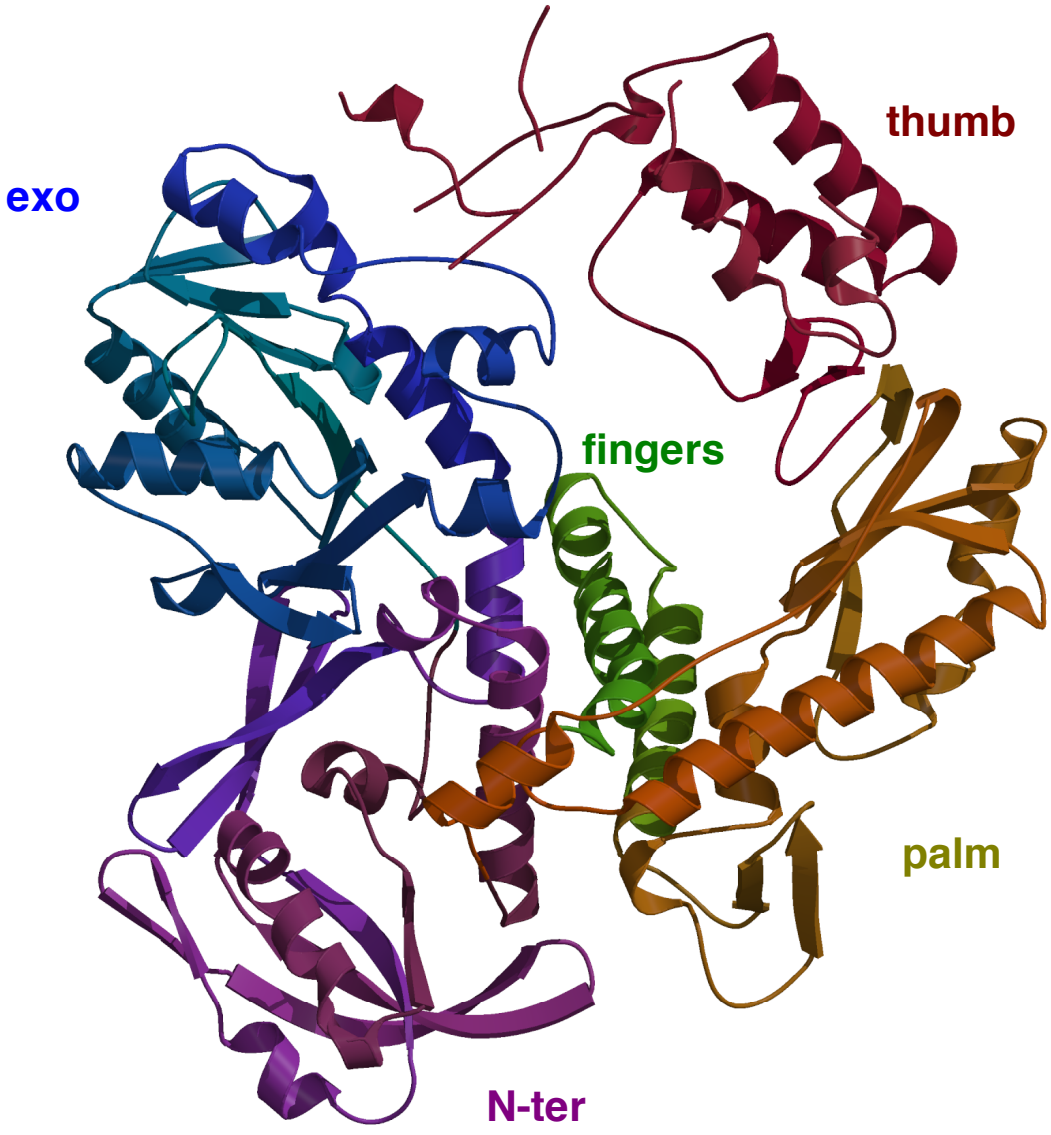

Figure 3.

Ribbon representation of the KOD DNA polymerase structure showing its five domains. The domains are NH2-terminal, exo, palm, fingers, and thumb.

Figure 4 shows the exonuclease domain of KOD DNA polymerase. Side chain of N213 interacts with side chain of N210, which is the one of highly conserved residues in exo motifs. And OG of T296 interacts with oxygen of main chain of Y146. The aromatic ring of Y146 is thought to interact with the aromatic ring of nucleotide at primer 3′ terminus. **Figure 5** shows sequence alignment for exo domains of KOD and *Pfu* DNA polymerases. In *Pfu* DNA polymerase, N213 and T296 are replaced to S213 and A296, respectively. Therefore, the differences of amino acids might be related to High PCR performance of KOD DNA polymerase.

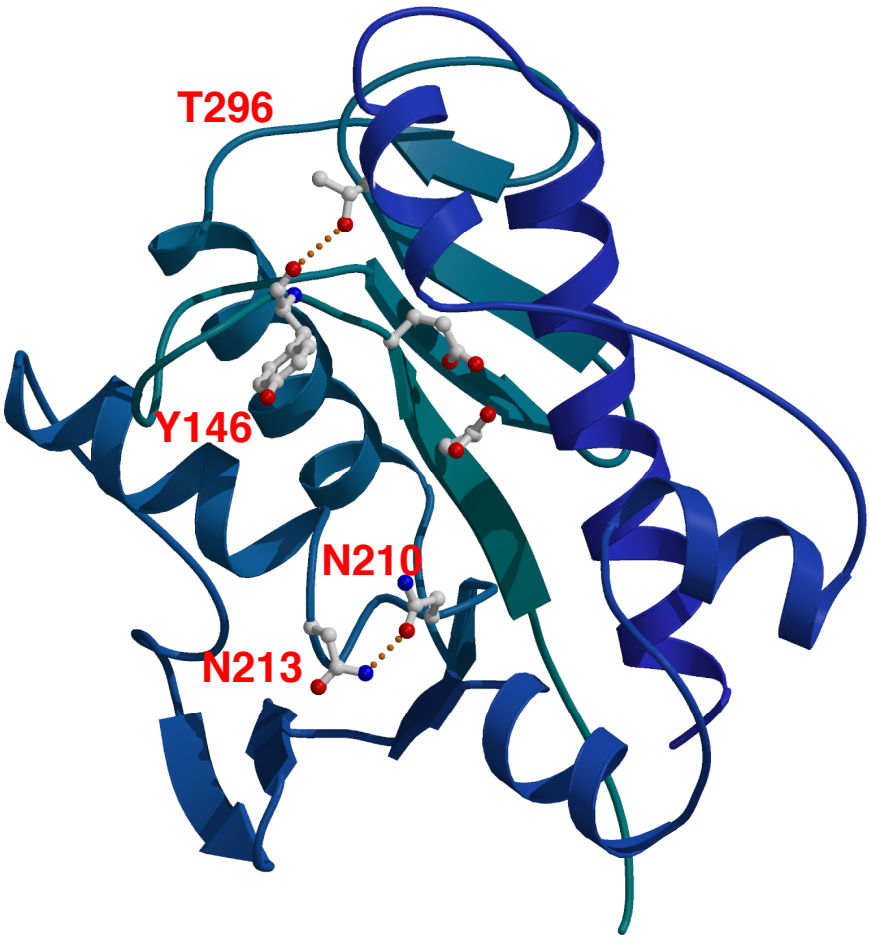

Figure 4.

Exo domain of KOD DNA polymerase**.**

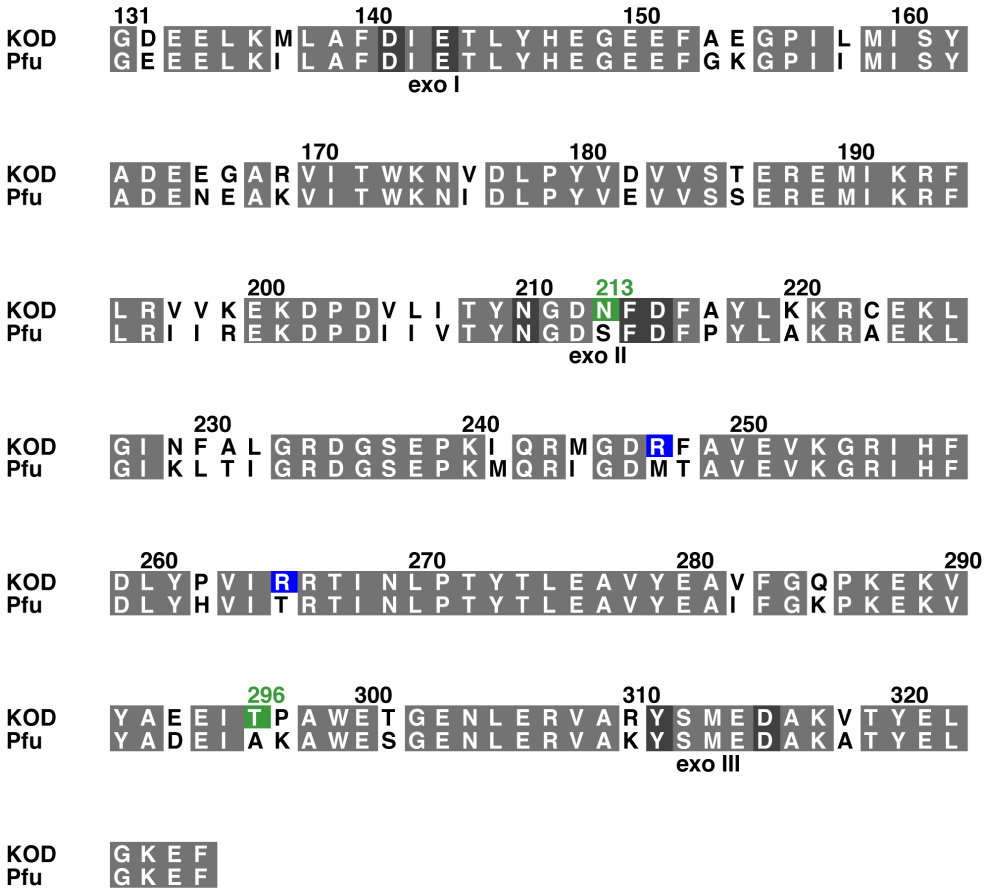

Figure 5.

Sequence alignment for exo domains of KOD and *Pfu* DNA polymerases.

1.4. REFERENCES

Bernad, A., Zaballos, A., Salas, M. and Balnco, L. (1987) Structural and functional relationships between prokaryotic and eukaryotic DNA polymerases. *EMBO J.*, **13**, 4219-4225.

Braithwaite, D. K. and Ito, J. (1993) Compilation alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787-802.

Brünger, A.T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, R., Grosse-Kunstleve, R. W., Jiang, J-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., simonson, T. and Warren, G. L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D**54***, 905-921.

Fujiwara, S., Okuyama, S. and Imanaka, T. (1996) The world of archaea: genome analysis, evolution and thermostable enzymes. *Gene* **179**, 165-170.

Goldman, S., Kim, R., Hung, L. -W., Jancarik, J. and Kim, S. -H. (1998) Purification, crystallization and preliminary X-ray crystallographic analysis of *Pyrococcus furiosus* DNA polymerase. *Acta Crystallogr. D***54**, 986-988.

Hopfner, K-P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R. and Angerer, B. (1999). Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. *Proc. Natl. Acad. Sci. USA*, **96**, 3600-3605.

Kiefer, J. R., Mao, C., Hansen, C. J., Basehore, S. L., Hogrefe, H. H., Braman, J. C. and Beese, L. S. (1997) Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure* **5**, 95-108.

Kim, Y., Eom, S. H., Wang, J., Lee, D. -S., Suh, S. W. and Steiz, T. A. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **346**, 612-616.

Korolev, S., Nayal, M., Barnes, W. M., Cera, E. D. and Waksman, G. (1995) Crystal structure of the large fragment of *Thermus*

aquaticus DNA polymerase I at 2.5 Å resolution: Structural basis for thermostability. *Proc. Natl. Acad. Sci. USA* **92**, 9264-9268.

Matthews, B. W. (1968) Solvent Content of Protein Crystals. *J. Mol. Biol.* **33**, 491-497.

Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. and Imanaka, T. (1994) Purification and characterizaion of a hyperthermophilic *Pyrococcus* sp. *Appl. Environ. Microbiol.* **60**, 4559-4566.

Nastopoulos, V., Pisani, F. M., Savino, C., Federici, L., Rossi, M. and Tsernoglou, D. (1998) Crystallization and preliminary X-ray diffraction studies of DNA polymerase from the thermophilic archaeon *Sulfolobus solfataricus*. *Acta Crystallogr. D***54**, 1002-1004.

Navaza, J. (1994) AMORE: an automated package for molecular replacement. *Acta Crystallogr. A***50**, 157-163.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enz.* **276**, 307-326.

Sakabe, N., Ikemizu, S., Sakabe, K., Higashi, T., Nakagawa, A., Watanabe, N., Adachi, S. and Sasaki, K. (1995) Weissenberg Camera for Macromolecules with Imaging Plate Data Collection System at the Photon Factory, Present Status and Future Plan. *Rev. Sci. Instrum.* **66**, 1276-1281.

Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. and Imanaka, T. (1997) Characterization of DNA Polymerase from *Pyrococcus* sp. Strain KOD1 and Its Application to PCR. *Appl. Environ. Microbiol.* **63**, 4504-4510.

Wang, J., Sattar, A. K. M. A., Wang, C. C., Karam, J. D., Konigsberg, W. H. and Steiz, T. A. (1997) Crystal structure of a pol a Family Replication DNA Polymerase from Bacteriophage RB69. *Cell*, **89**, 1087-1099.

Watanabe, N., Nakagawa, A., Adachi, S. and Sakabe, N. (1995) Macromolecular Crystallography Station BL-18B at the Photon Factory. *Rev. Sci. Instrum.* **66**, 1824-1826.

Zhou, M., Mao, C., Rodriguez, A. C., Kiefer, J. R., Kucera, R. B. and Beese, L. S. (1998) Crystallization and preliminary diffraction analysis of a hyperthermostable DNA polymerase from a *Thermococcus* archaeon. *Acta Crystallogr. D***54**, 994-995.

CHAPTER 2

==========

Structural study of O^6-methylguanine-DNA methyltransferase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1

2.1. INTRODUCTION

The exposure of DNA to alkylation agents such as *N*-methyl-*N'*-nitro-*N*-nitrosoguanidine (MNNG) yields O^6-alkylguanine residues. The formations of these derivatives are highly mutagenic and carcinogenic because O^6-alkylguanine can pair with thymine instead of cytosine, and as a result induce G-C to A-T base pair transitions during DNA replication (Coulondre & Miller, 1977). These transitions are blocked by the action of the suicidal DNA repair protein, O^6-methylguanine-DNA methyltransferase. This protein transfers an alkyl group substituted at the guanine O^6 to one of its own cysteine residue. The methylated protein is inactive in further O^6-methylguanine-DNA repair because of the covalent methylation on active site cysteine.

O^6-methylguanine-DNA methyltransferases are present in various organisms ranging from bacteria to human cells. *Escherichia coli* has two types of O^6-methylguanine-DNA methyltransferases; the Ada protein (product of the *ada* gene) and the Ogt protein (product of the *ogt* gene). The Ada protein consists of 20 kDa N-terminal and 19 kDa C-terminal regions connected by a flexible hinge. Proteolysis easily cleaves the Ada protein into N-terminal (AdaN) and C-terminal (AdaC) fragments. The C-terminal region of the Ada protein has methyltransferase activity from O^6-methylguanine-DNA. The Ogt protein, is similar to the AdaC fragment with respect to its chemical reaction and molecular size (19 kDa). Conversely, Eucarya and Archaea have only one type of O^6-methylguanine-DNA methyltransferase, with a molecular weight of 18-22 kDa. In all known O^6-methylguanine-DNA methyltransferase, there are many conserved amino acid sequences in the C-terminal domains (Moore *et al.*, 1994; Leclere *et al.*, 1998). The crystal structure of C-terminal fragment of Ada protein from *E. coli* (AdaC) was determined (Moore *et al.*, 1994). In the present paper, it is reported that the crystal structure of O^6-methylguanine-DNA methyltransferase from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1 (*Pk*-MGMT) and its structural comparison

with AdaC structure, in which the structural features of *Pk*-MGMT are shown to be essential for hyperthermostability. *Pk*-MGMT is a monomeric protein with the molecular weight of 19.5 kDa made up of 174 amino acid residues. Several oligomeric protein structures of hyperthermophiles have recently been reported and emphasized the importance of the reduction of solvent accessible surface area (ASA), and the increase of ion-pairs between subunits for thermostability. However, there are not many reports of structural studies on monomeric hyperthermophilic proteins.

The hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1 is one of the most thermostable organisms known, with an optimum growth temperature of 95°C (Morikawa *et al.*, 1994). Enzymes produced in this strain were reported to be extremely thermostable and to have eucaryotic characteristics (Fujiwara *et al.*, 1996). Recombinant *Pk*-MGMT has methyltransfer activity and is found to be thermostable at 90°C for 30 min, and is not fully denatured even at 98°C (Leclere *et al.*, 1998).

2.2. MATERIALS AND METHODS

The archaeal MGMT was overexpressed in *E. coli* strain HMS174(DE3)pLysS, purified (Leclere, *et al*. in press.), and dialyzed against 50 mM Tris-HCl buffer at pH 8.0 containing 0.1 mM EDTA. The dialyzed protein was concentrated to 10 mg/mL. All the crystallization experiments were carried out using the hanging drop vapor diffusion technique. The droplet (typically 4 μL) was prepared by mixing equal volumes (2 μL) of the protein and reservoir solutions. A number of precipitants including salts, polyethylene glycol solutions, and organic solvents were explored over a broad pH (4.0~10.0). As a result, two forms of crystals appeared.

Form I crystals (**Figure 1**) were obtained by the following procedure. The reservoir solution was prepared by mixing 400 μL of 50%(*w*/*w*) PEG 8000, 200 μL of 1.0 M Zn acetate, 100 μL of 1.0 M Na cacodylate pH 6.5 and 300 μL distilled water, giving the final concentrations of 20% PEG 8000, 200 mM Zn acetate and 100 mM Na cacodylate. Form I crystals were thin plates in shape and with a maximum size of 0.2 x 0.1 x 0.05 mm.

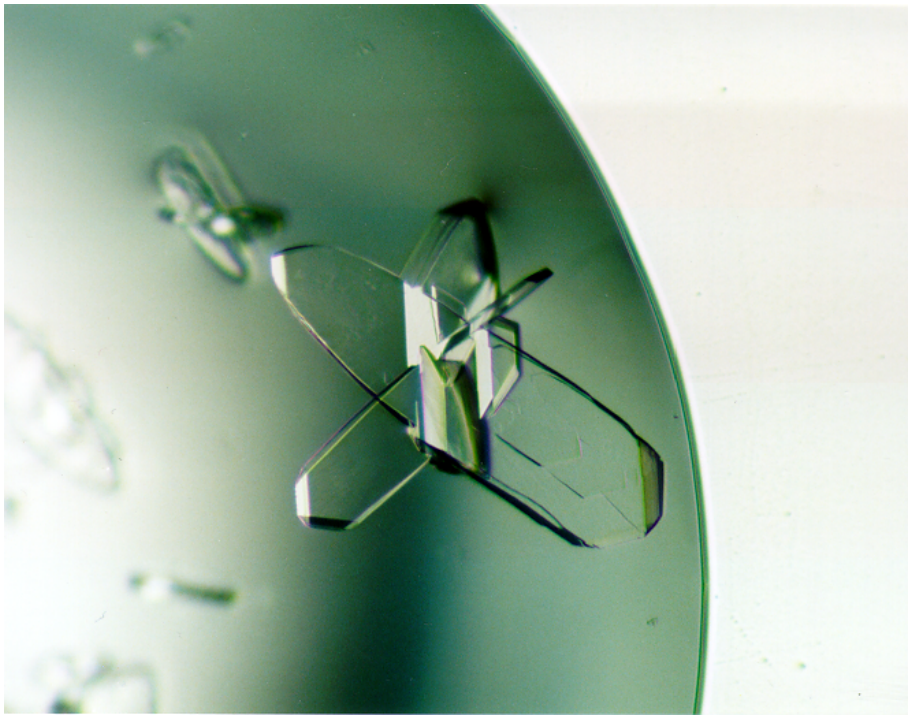

Figure 1.

Form I crystals of MGMT with the maximum dimensions 0.2 x 0.1 x 0.05 mm.

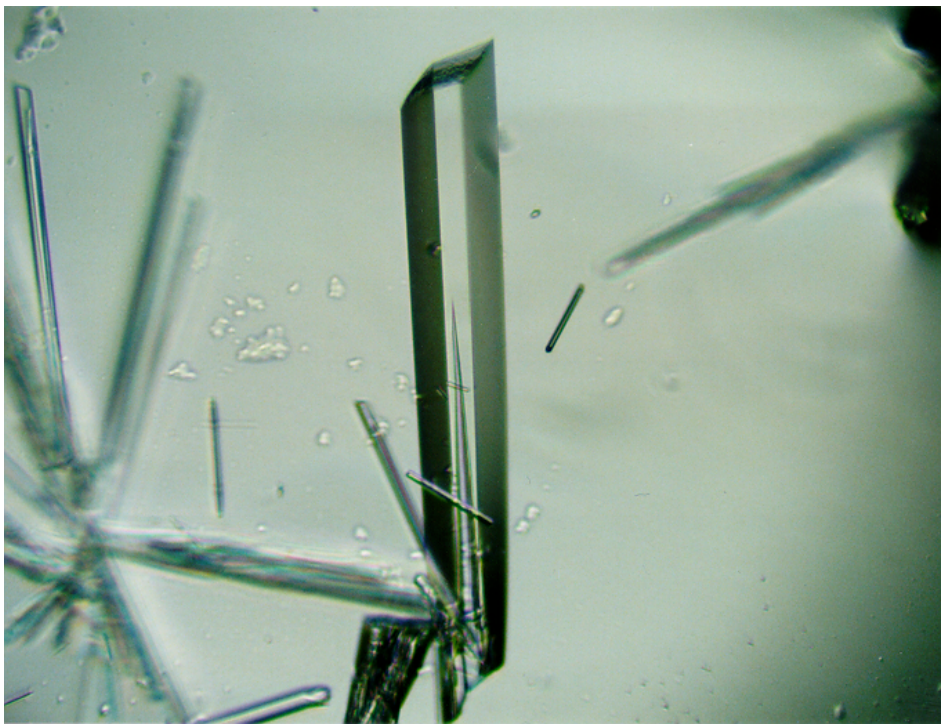

Figure 2.

Form II crystals of MGMT with the maximum dimensions 0.8 x 0.1. x 0.1 mm.

Form II crystals were obtained as follows. First, needle-shaped microcrystals were grown against a reservoir solution of 15% PEG

8000 and 200 mM ammonium sulfate. When 15%(*w*/*w*) PEG 20000 (300 μL of 50% (*w*/*w*) PEG 20000) were used instead of PEG 8000, rod-shaped crystals appeared. When the concentration of PEG 20000 was decreased down to 12% (*w*/*w*), crystals reached dimensions of 0.8 x 0.1 x 0.1 mm in a few weeks (**Figure 2**).

When form I crystals (**Figure 1**) were exposed to Cu$K\alpha$ X-ray, the diffraction spots were observed to Bragg spacings of at least 2.0 Å. The space group was determined to be *P*1 with the unit-cell parameters of a = 43.8, b = 54.2 and c = 43.8 Å, α = 75.4, β = 79.8, and γ = 85.5°. The asymmetric unit contains two molecules with a mass of 19,500, giving a crystal volume per protein mass (V_m) of 2.5 $Å^3Da^{-1}$ (Matthews, 1968). As the mosaicity of the crystals were estimated at more than 1° by *SCALEPACK* (Otwinowski & Minor, 1997), the structure determination using form I crystals was abandoned.

Diffraction intensities from form II crystals exposed to Cu$K\alpha$ X-ray were also observed to at least 2.0 Å Bragg spacings. Unit-cell parameters were determined as a = 52.8, b = 86.6 and c = 39.9 Å with the space group of $P2_12_12_1$. The asymmetric unit contains a single molecule with a mass of 19,500, giving a crystal volume per protein mass (V_m) of 2.3 $Å^3Da^{-1}$ and the solvent content of

48% by volume. The mosaicity of the crystals was estimated to be about 0.2°; form II crystals were found to be suitable for X-ray studies. As listed in **Table 1**, a data set was collected. It was consists of 73,204 measurements of 12,692 unique reflections with an overall R_{merge} of 8.6% ($R_{merge} = \Sigma|I-<I>|/\Sigma I$) and overall $I/sigI$ of 9.9. This represents 97.8% of theoretically observable reflections at 2.0 Å resolution. The outermost shell of data between 2.07 and 2.00 Å is 98.2% complete. Attempts were made to solve the MGMT structure by molecular replacement with *AMoRe* (Navaza, 1994; Collaborative Computational Project, Number 4, 1994) and *X-PLOR* (Brünger, 1990) using AdaC from *E. coli* as a probe structure; however no consistent set of rotation function solutions could be obtained. MGMT from KOD1 has low amino acid sequence homology, with AdaC except around the active site, homology with AdaC was about 13%. For this reason, molecular replacement technique might not be successful in this case. Then, structure determination by isomorphous replacement method was applied.

Table 1.

Diffraction data statistics of form II crystal exposed to Cu$K\alpha$ X-ray.

Resolution (Å)	No. of reflections (unique)	Completeness (%)	R_{merge}
20.00 - 4.30	1,335	94.1	0.042
4.30 - 3.42	1,285	97.5	0.064
3.42 - 2.99	1,286	98.5	0.080
2.99 - 2.71	1,285	98.8	0.101
2.71 - 2.52	1,252	98.4	0.119
2.52 - 2.37	1,264	98.4	0.144
2.37 - 2.25	1,254	97.6	0.162
2.25 - 2.15	1,252	98.5	0.192
2.15 - 2.07	1,240	98.7	0.230
2.07 - 2.00	1,239	98.2	0.286
overall	12,692	97.8	0.086

Isomorphous heavy-atom derivative was obtained from two weeks soaking at 20°C, using heavy-atom solutions containing 10 mM $KAu(CN)_2$, 200 mM ammonium sulfate and 15% PEG 20000. Native data was collected at 17 °C using the beamline 18B synchrotron X-ray source at Photon Factory (λ = 1.00 Å) with the Weissenberg method

(Sakabe *et al.*, 1995; Watanabe *et al.*, 1995). The final native data set was prepared by merging two native data sets. Heavy atom derivative data was collected with the simple oscillation method at room temperature on a Rigaku R-AXIS IIc imaging-plate detector system coupled to RU-300 fine-focused rotating-anode X-ray generator with Cu target (λ = 1.5418 Å). All diffraction data was indexed and processed with programs *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). X-ray diffraction statistics for isomorphous replacement are summarized in **Table 2.** The derivative data was scaled by anisotropic scaling to the native data by program *SCALEIT* (*CCP4*, 1994).

Table 2.

Data collection and scaling statistics

	Native	$KAu(CN)_2$
Wavelength	1.00 Å	1.54 Å
X-ray source	SR (KEK-PFBL18)	Cu$K\alpha$
Detector	Weissenberg camera	RAXIS-IIc
Resolution	40.0 ~ 1.8 Å	20.0 ~ 2.0 Å
No. of total reflections	92,828	69,449
No. of unique reflections	16,314	12,421
Completeness	92.4% (76.8%)	95.9% (89.5%)
R_{merge}	6.5% (16.0%)	7.6% (17.6%)
R_{deriv}		18.6% (21.3%)

Numbers in parentheses refer to the statistics for the outer shell of each data.

$R_{merge} = \Sigma_h \Sigma_i | I_{h,i} - \langle I_h \rangle | / \Sigma_h \Sigma_i I_{h,I}$; $R_{deriv} = \Sigma_h | F_{native,\ h} - F_{deriv,\ h} | / \Sigma_h F_{native,\ h}$

Table 3.

Phasing statistics

Resolution	20.0 ~ 2.0 Å
Phasing power for acentric reflections	2.39
Phasing power for centric reflections	1.89
FOM	0.52

Numbers in parentheses refer to the statistics for the outer shell of each data.

Phasing power (Ph. P) is rms ($|F_h|/E$), where the subscript h represents heavy-atom and E is the residual lack of closure.

FOM (mean figure of merit) = $\langle |\Sigma P(\alpha) e^{i\alpha} / \Sigma P(\alpha)| \rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

Phase determination was carried out by the single isomorphous replacement method with anomalous scattering (SIRAS). Difference Patterson maps for the $KAu(CN)_2$ derivative showed only one strong peak in the $u = v = w = 1/2$ Harker sections (**Figure 3**). The one heavy atom position was easily determined by manual inspection,

then refined using reflections from 20.0 to 2.0 Å resolution with program *MLPHARE* (Otwinowski, 1991), including anomalous differences. The quality of the anomalous data was judged by the anomalous difference Patterson maps and phasing statistics, which showed effective anomalous signals for phase determination. The initial SIRAS phase was improved and extended to 1.8 Å resolution by the solvent flattening and histogram matching method using program *DM* (Cowtan, 1994). Phasing statistics were summarized in **Table 3.** The quality of the 1.8 Å resolution solvent flattened SIRAS electron density map (**Figure 4 and 5a**) was excellent, and nearly the entire main and side chain were built using program *O* (Jones et al, 1994).

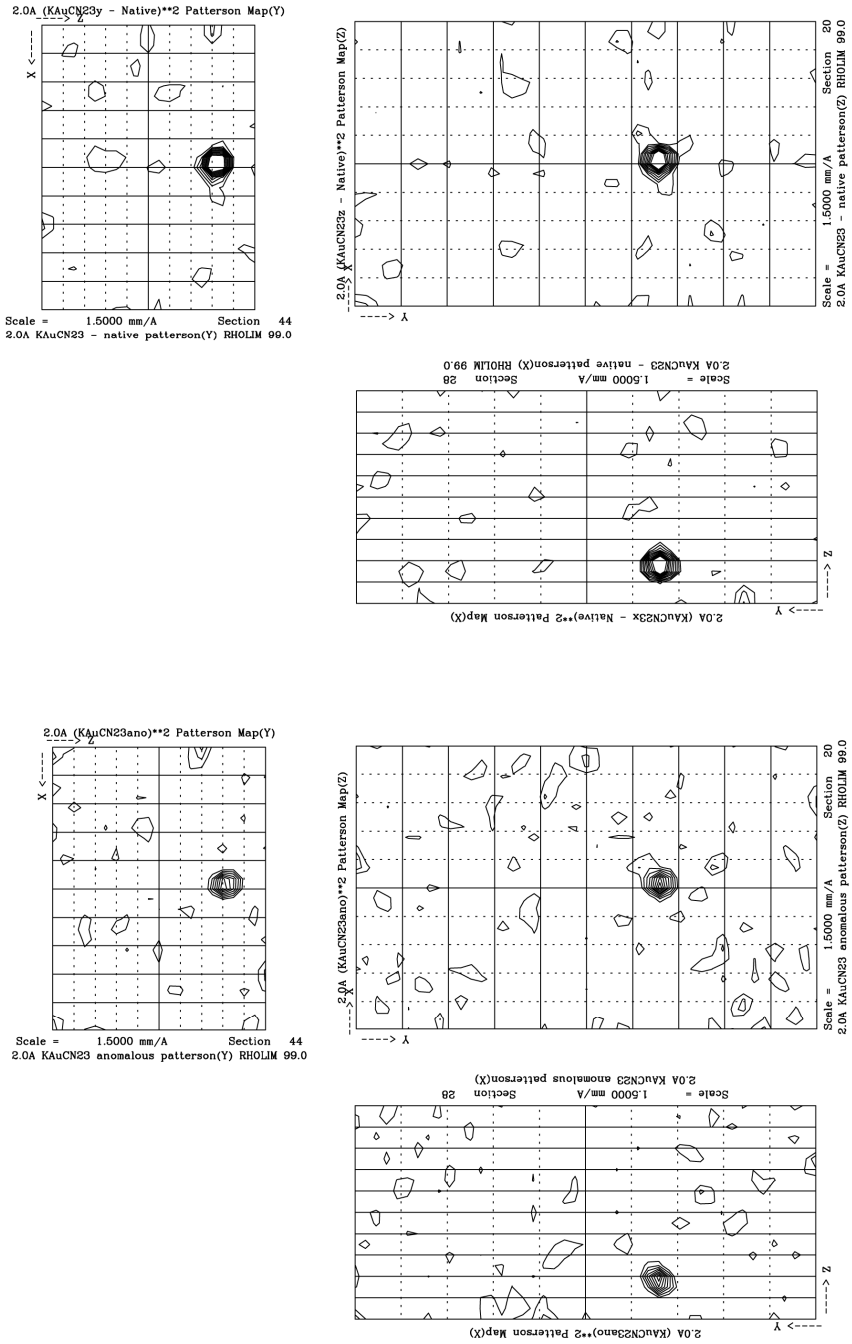

Figure 3.

Isomorphous (above) and anomalous (below) Patterson maps in the $u = v = w = 1/2$ Harker sections.

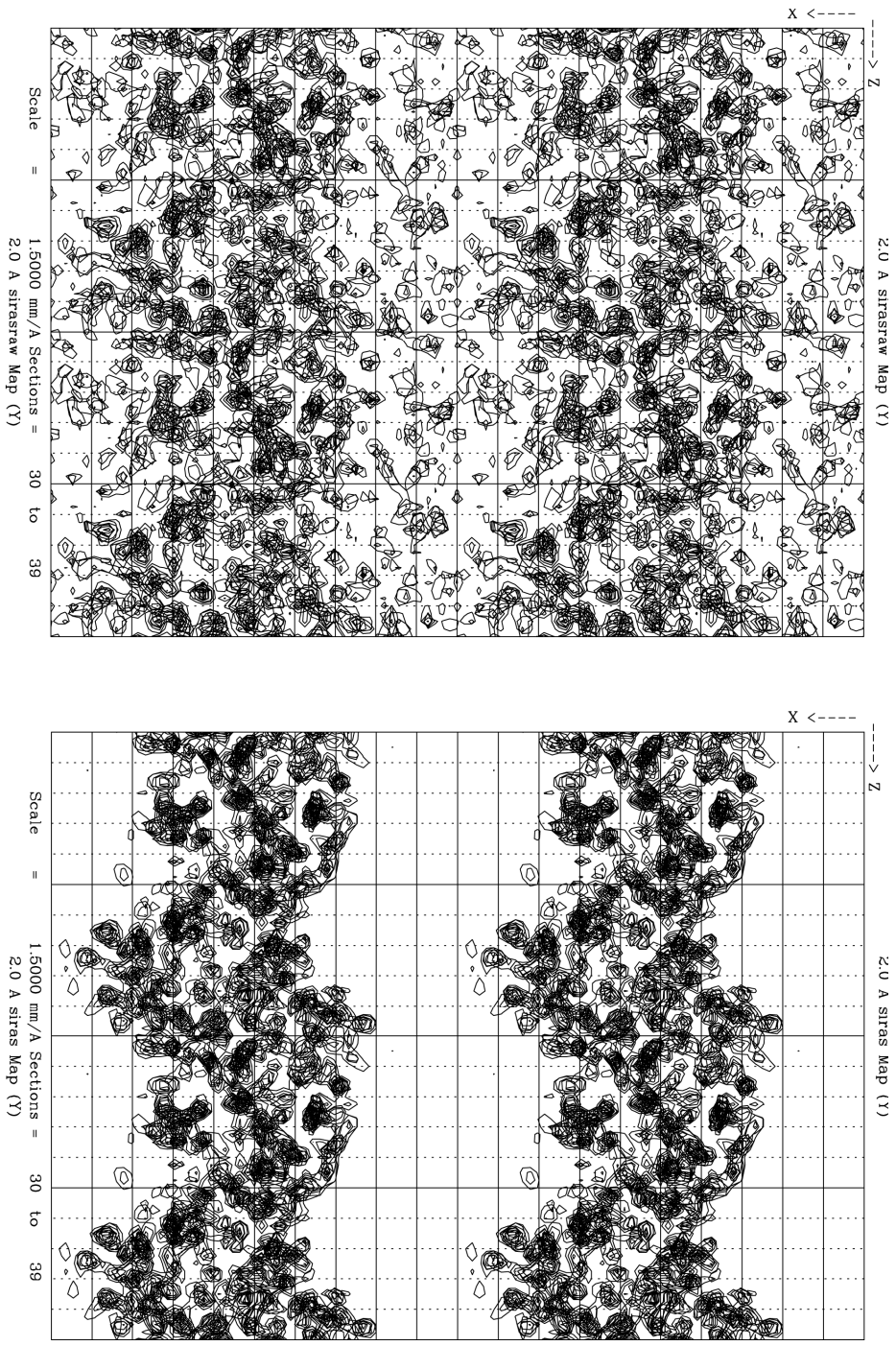

Figure 4.

Electron densigy maps calculated with SIRAS phases (above) and extended phases by program *DM* (below).

Initially, the structure was refined by program *XPLOR* (Brünger *et al.*, 1987; Brünger *et al.*, 1990; Brünger *et al.*, 1992) to the *R*-factor of 21.5% (R_{free} = 28.2%) at 5.0 to 2.0 Å resolution. Finally, the protein structure containing solvent molecules was refined by maximum-likely food using program *REFMAC* with bulk solvent correction (Murshudov *et al.*, 1997). The *R*-factor of the final model involved 169 water molecules and three SO_4^{2-} is 17.3% for 15,463 reflections from 20.0 to 1.8 Å resolution (R_{free} = 21.8%). **Figure 5b** shows the final 2*F*o-*F*c map. The crystallographic refinement statistics are summarized in **Table 4.** The geometry of the final model was checked by program *PROCHECK* (Morris *et al.*, 1992). The Ramachandran plot of the final model shows that 95.1% of the residues are in its most favored region and only one residue, Ile137, is in the disallowed regions. Ile137 is located on the sharp turn between e- and f-helices. This residue is conserved in AdaC, I142, which is also in disallowed regions. Secondary structures of the final model of *Pk*-MGMT and AdaC were defined by program *DSSP* (Kabsch & Sander, 1983). The complete refined coordinates of *Pk*-MGMT, the observed structure factor amplitudes, and the experimentally determined phases have been deposited in Brookhaven Protein Data Bank with accession code 1mgt.

Figures 6, 8 and **9** were made using programs *MOLSCRIPT* (Kraulis, 1991) and *RASTER3D* (Merritt & Murphy, 1994; Merritt & Bacon, 1997); **Figure 7** was made using program *ALSCRIPT* (Barton, 1993); and **Figure 5** was made using program *O* (Jones *et al.*, 1994)

Table 4.

Refinement statistics

Resolution	20.0 ~ 1.8 Å
No. of reflections	15,463
R-factor	17.3%
R_{free}	21.8%
r.m.s. (bonds)	0.018 Å
r.m.s. (angles)	2.5°
No. of amino acid residues	169
No. of protein atoms	1,334
No. of water molecules	129
No. of SO_4^{2-}	3

R-factor and $R_{free} = \Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c are the observed and calculated structure factor amplitudes.
R_{free} was calculated with 10% of the reflections not used during refinement.

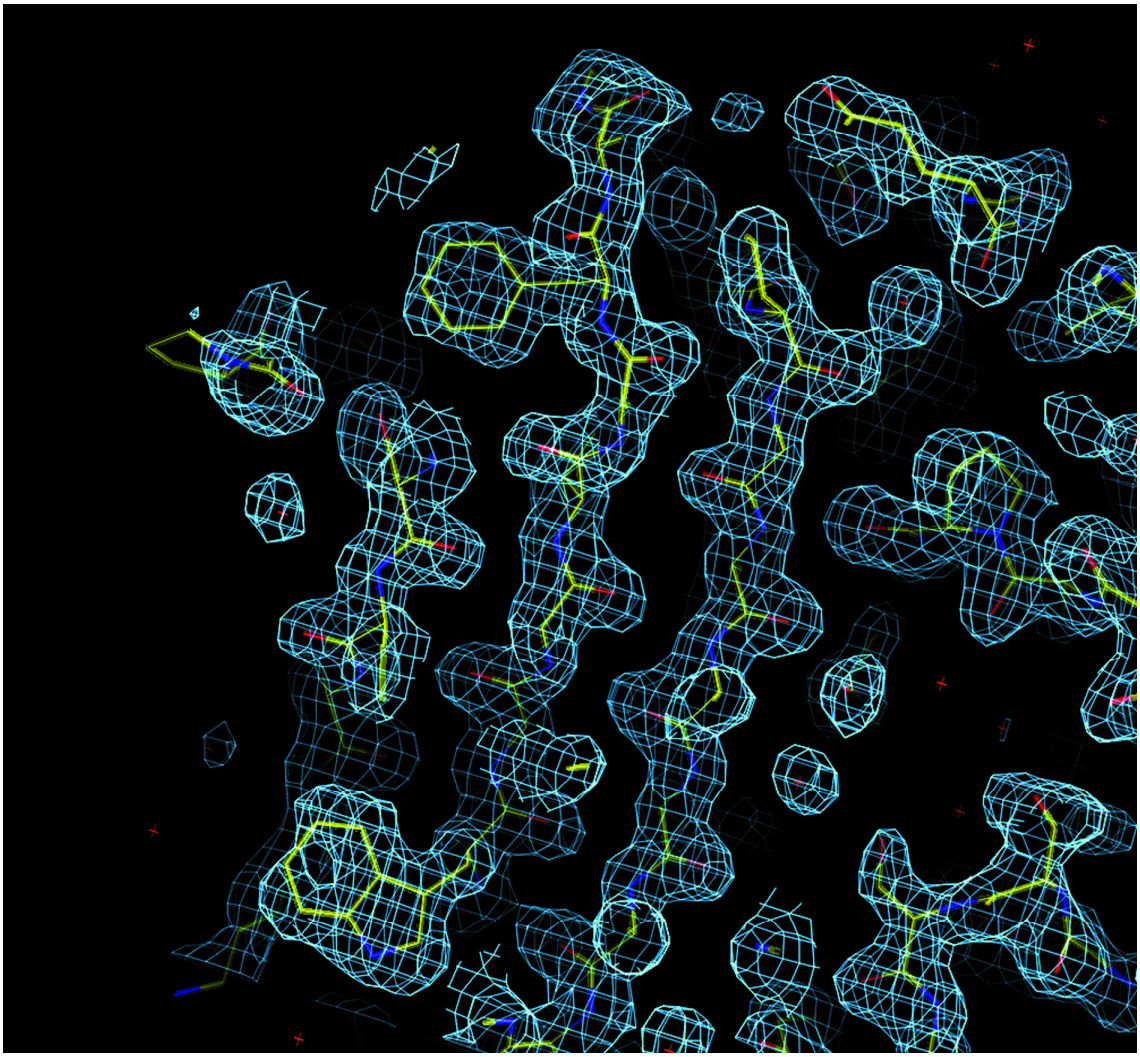

Figure 5a.

Typical sections of the initial 1.8 Å resolution solvent-flattened SIRAS electron density map contoured at 1σ, superimposed on the refined 1.8 Å resolution coordinates of *Pk*-MGMT.

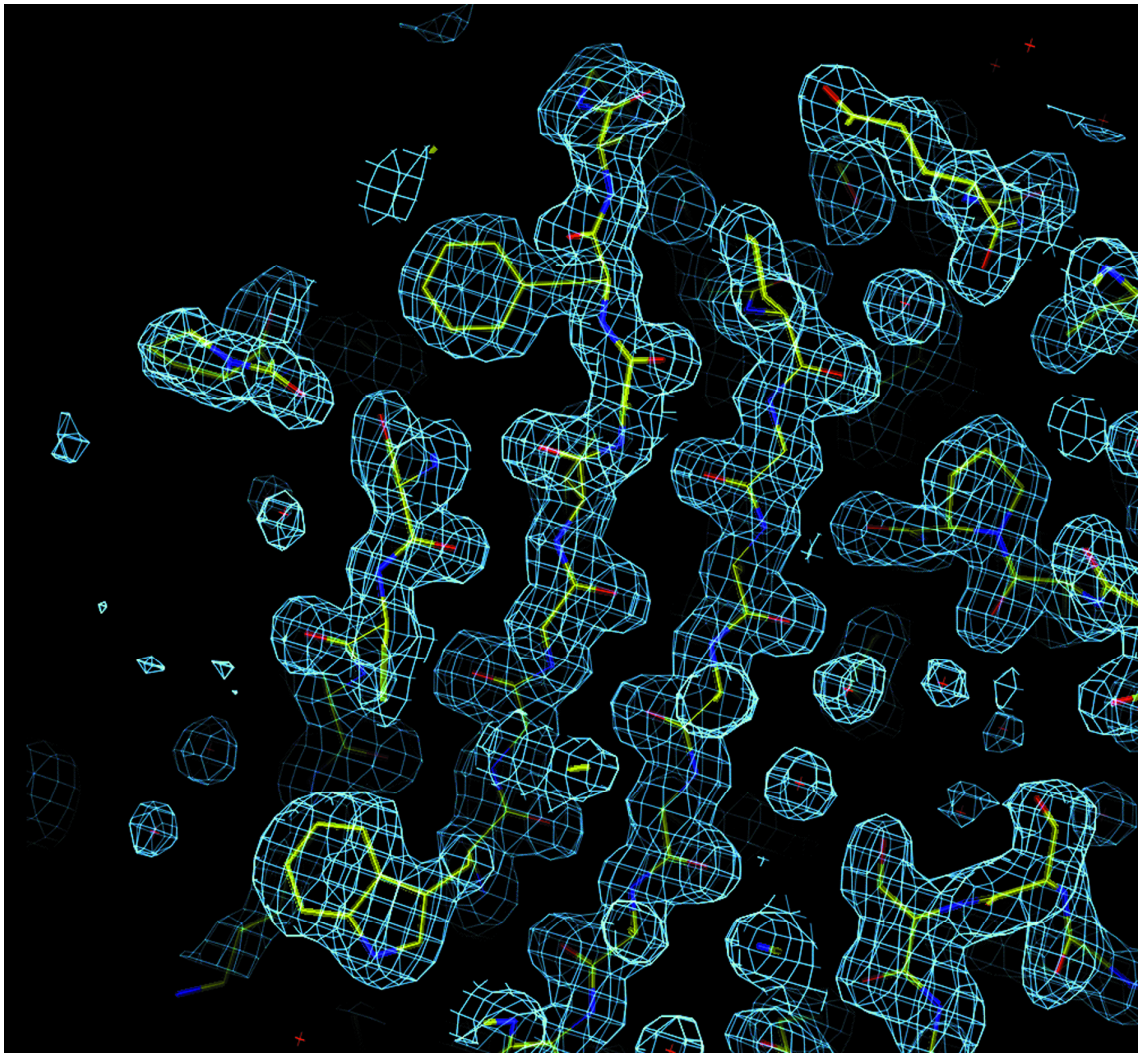

Figure 5b.

Typical sections of The 1.8 Å resolution final 2|*F*o|-|*F*c| map contoured at 1σ, superimposed on the refined 1.8 Å resolution coordinates of *Pk*-MGMT.

2.3. RESULTS AND DISCUSSION

The overall structure of *Pk*-MGMT

The X-ray structure of *Pk*-MGMT revealed that the protein has two distinct domains (**Figure 6a**); N- and C-terminal. The N-terminal domain comprises the first 85-90 residues and is made up of three strands constructing an anti-parallel β-sheet (β1, β2 and β3) and three helices (a, b1 and b2-helices), of which b2-helix is a 3_{10} helix. The C-terminal domain is made up of the remaining residues and includes five helices (c, d, e, f and g-helices) and two strands (β5 and β6) constructing a parallel β-sheet. Five residues from C-terminus (170-174) are highly disordered and are not found in the electron density map. The active site thiol of C141 exists in the f-helix (3_{10} helix) of the C-terminal domain and is buried inside of the protein similar to AdaC, shown in **Figure 6b** (C146). The reaction mechanism of the suicidal alkyltransfer is still unclear. Spectroscopic experiments suggest that a conformational change in the protein is caused when the human methyltransferase interact with DNA (Takahashi *et al.*, 1988; Chan *et al.*, 1993).

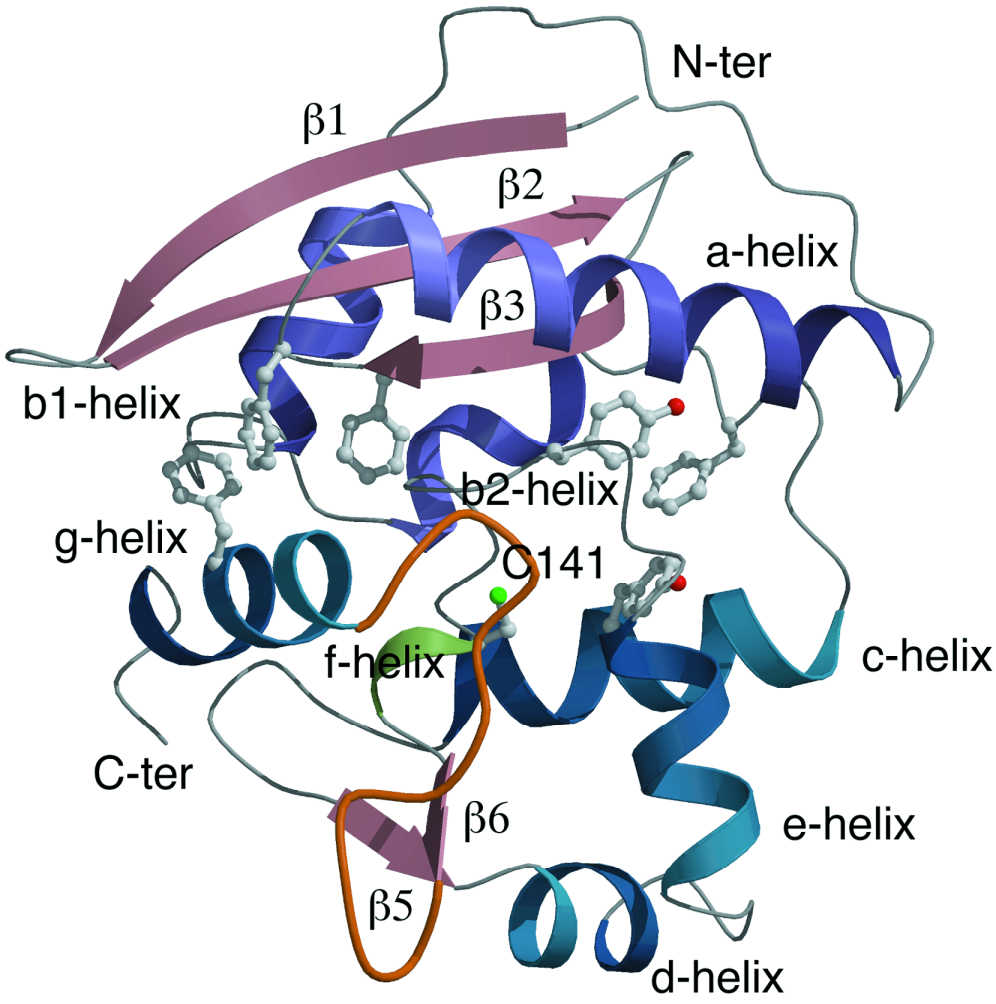

Figure 6a.

Schematic drawing of the crystal structure of *Pk*-MGMT. Aromatic clusters are displayed by ball and stick models. Wing loop from end of β6 to start of g-helix is represented by yellow coil, as shown in **Figure 9b**.

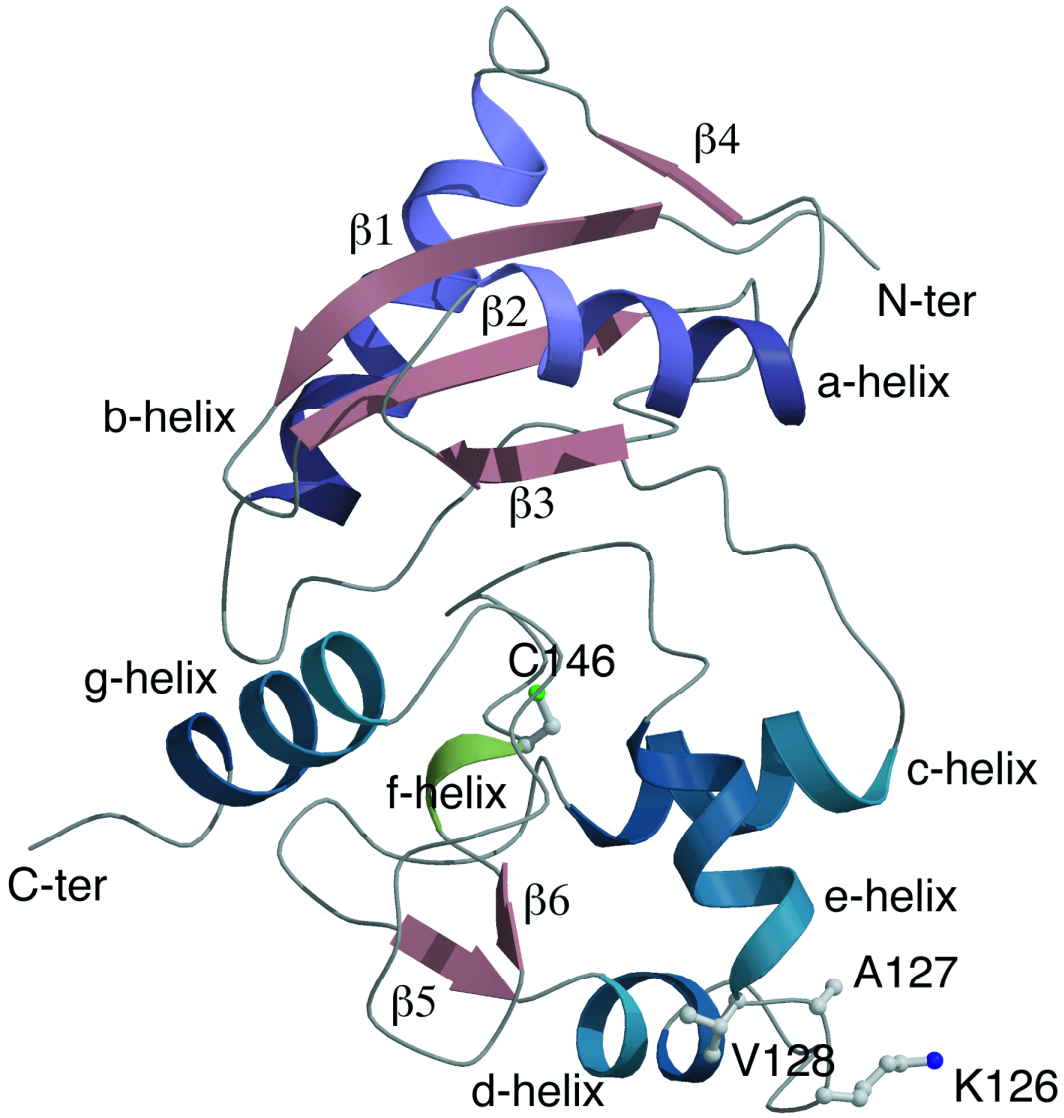

Figure 6b.

Schematic drawing of the crystal structure of AdaC. Inserted three residues, K126, A127 and V128 are shown by ball and stick models.

The experimental protein-DNA complex model is required for comprehensive understanding of the reaction mechanism of suicidal alkyltransfer and substrate recognition. However, it is difficult to prepare the protein-DNA complex for X-ray studies because the alkyltransfer reaction is S_N2 type and there is no chemical intermediate: only one transition state exists during the alkyltransfer reaction. Details of the active site structure and structural basis for DNA-binding is described in a later section.

Structural comparison of *Pk*-MGMT with AdaC

Secondary-structure assignments and structure-based sequence alignment for *Pk*-MGMT and AdaC are given in **Figure 7a.** Highlighted residues of green colored masks are identical residues between the proteins. In N-terminal domain, secondary structures of the two proteins are similar in topology except for the 3_{10}-helix in *Pk*-MGMT (b2-helix, from Q77 to E83) and one short strand in AdaC (β4). In AdaC structure, the corresponding region of b2-helix in *Pk*-MGMT is part of a long loop between N- and C-terminal domains. It has been shown that the sequence similarity of C-terminal domain is high in various methyltransferases and suggests similar structures in C-terminal domains among all the homologues (Moore

et al., 1994). C-terminal domains of *Pk*-MGMT and AdaC have identical architecture except for loop structure between d- and e-helices. **Figure 7b** shows an amino acid sequence alignment around e-helix of various homologous methyltransferases from Archaea, Bacteria and Eucarya. All archaeal methyltransferases and methyltransferase from hyperthermophilic bacterium *Aquifex aeolicus* have an identical deletion between proline and arginine. *A. aeolicus* is one of the earliest divergent and most thermophilic bacteria known, with growth temperature maxima near 95 °C. Site-direct mutagenic experiments of human MGMT suggest that d- and e-helices interact with O^6-methylguanine-DNA (Kanugula *et al.*, 1995; Goodtzova *et al.*, 1996). In *Pk*-MGMT, deletion of the three residues shortens the loop structure between d- and e-helices. The deletion contributes to the stabilization of the loop structure in the DNA-binding region, and positions the helices to keep the "helix-loop-helix" motif. The deletions may relate to the structural stability in various extreme environments and to the process of evolution in ancient organisms.

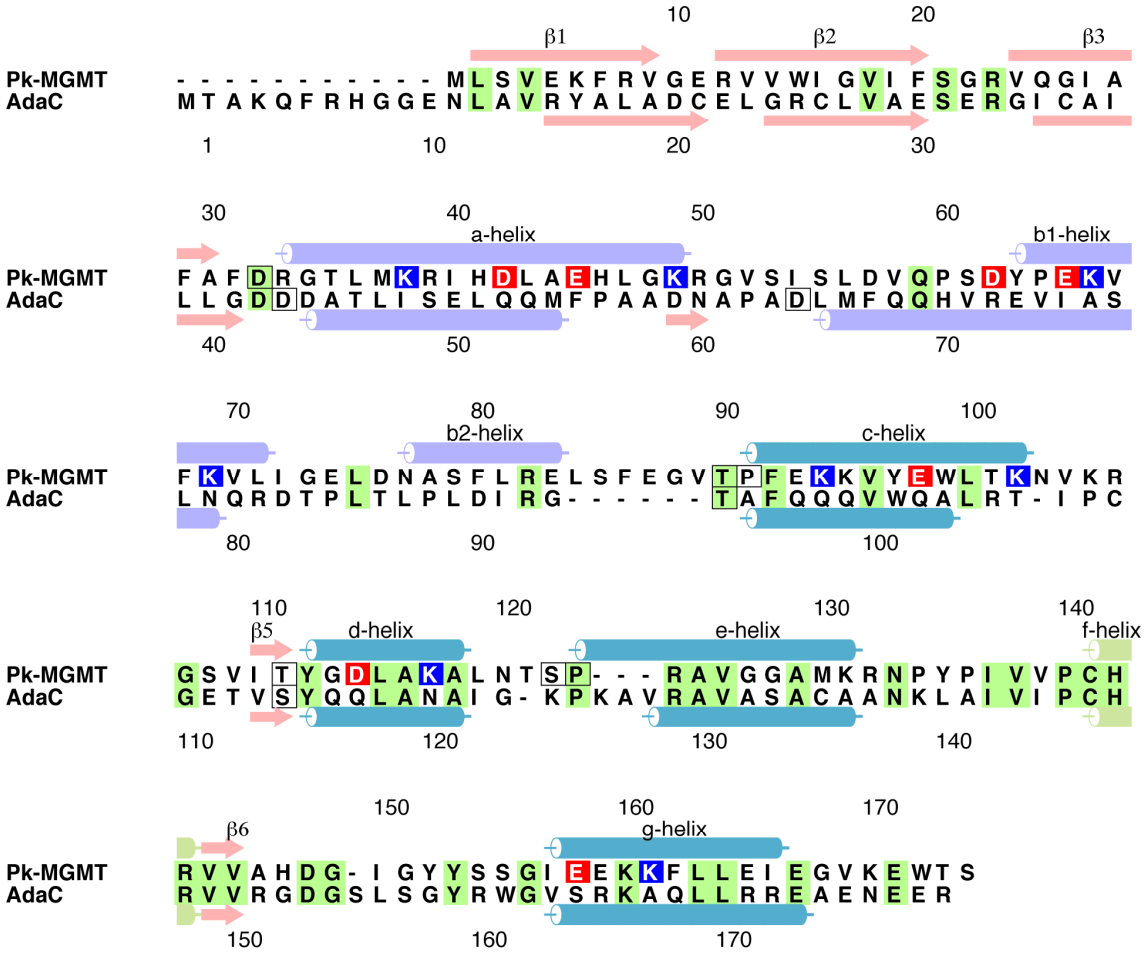

Figure 7a.

Structure-based sequence alignment of two methyltransferases. Identical residues are highlighted. N-cap residues and N1 prolines are boxed. All amino acids forming intra-helix ion-pairs are highlighted in inverse type.

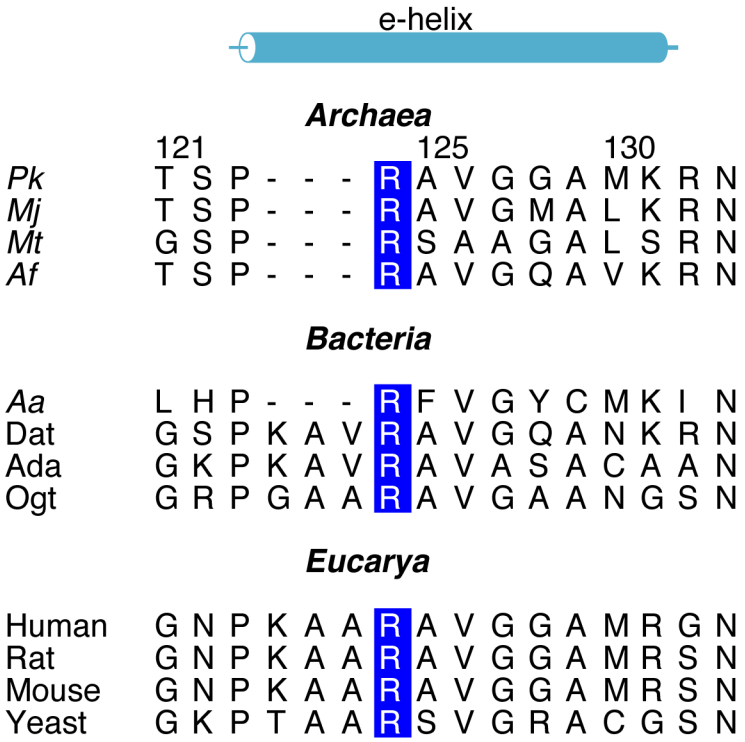

Figure 7b.

Comparison of the amino acid sequences around e-helices of O^6-methylguanine-DNA methyltransferases of various sources. The abbreviations are as follows; *Pk*: *Pyrococcus kodakaraensis*, *Mj*: *Methanococcus jannaschii*, *Mt*: *Methanobacterium thermoautotrophicum*, *Af*: *Archaeoglobus fulgidus*, and *Aa*: *Aquifex aeolicus*. Absolutely conserved arginines are highlighted in inverse. Deletion of the three residues are deleted next to proline in all hyperthermophiles.

Hyperthermostability

Many structural features have been suggested for thermostability, such as hydrophobic interaction, hydrogen bonds, helix cappings, ion-pairs and solvent accessible surface areas. These features will be discussed in the following section by comparing the *Pk*-MGMT and AdaC structure.

Helix Capping

It was shown that some amino acids tend to occur much more frequently than others in certain positions within a-helices (Richardson & Richardson, 1988; Serrano & Fersht, 1989; Matthews, 1993; Fersht & Serrano, 1993). It was reported that small polar or charged side chains (Asp, Thr, or Ser) are suitable at the N-cap position, because they stabilize the NH group of the main chain at the N3 position by forming intramolecular hydrogen bonds (where N3 is the N-cap + 3 position). Proline favors the N1 position (the first residue after the N-cap position), because it fits well in the first turn of the helix in terms of its own backbone conformation. In the present case, N-cap residues and N1 prolines are boxed in **Figure 7a.** There are four suitable residues at N-cap positions in *Pk*-MGMT (D32, T90, T111, and S122) and two N1 proline residues

(P91 and P123). Although D62 is situated at the N-cap position in the b1-helix, D62 makes the [i+4] intra-helix ion-pair with K66 as described below. Conversely, AdaC has four N-cap residues and no N1 proline. The first N1 proline in *Pk*-MGMT (P91) is replaced by A95 in AdaC. Although the second one in *Pk*-MGMT (P123) corresponds to P125 in AdaC, V128 is situated at N1 position because of the insertion of three amino residues next to P125 (**Figure 6b** and **7b**). N-cap residues and N1 prolines in *Pk*-MGMT are also shown in **Figure 9a.**

Solvent Accessible Surface Area

In general, the decrease of the solvent accessible surface area (ASA) and the increase of the fraction of buried hydrophobic atoms have been discussed as the stabilizing principles for thermostable protein (Chan *et al.*, 1995). **Table 5** shows comparison of the ASAs of *Pk*-MGMT with those of AdaC. *Pk*-MGMT (8,160 $Å^2$) shows smaller total ASA than AdaC (8,339 $Å^2$). Averaged by the number of residues, *Pk*-MGMT shows an average ASA of 48.3 $Å^2$/residue, while AdaC shows a 50.5 $Å^2$/residue. The contributions of hydrophobic, polar and charged residues to the total ASA are analyzed separately in the two proteins.

Table 5.

Comparison of solvent accessible surface areas

	Pk-MGMT	AdaC
Total ASA (Å^2)	8160	8339
Hydrophobic ASA (Å^2) (% of total)	1935 (24%)	2638 (32%)
Polar ASA (Å^2) (% of total)	1797 (22%)	2752 (33%)
Charged ASA (Å^2) (% of total)	4428 (54%)	2949 (35%)
No. of residues in crystal structure	169	165
No. of hydrophobic residues (% of total)	75 (44.4%)	78 (47.3%)
No. of polar residues (% of total)	46 (27.2%)	50 (30.3%)
No. of charged residues (% of total)	48 (28.4%)	37 (22.4%)

Then AdaC shows no significant difference in the contributions

by each property (32% for hydrophobic, 33% for polar and 35% for charged residues). As shown in the lower half of **Table 5**, there is no significant difference of hydrophobic and polar residues' compositions. Although the number of charged amino acids in *Pk*-MGMT are slightly increased compared with AdaC, *Pk*-MGMT shows a remarkably large discrepancy in contributions to ASAs by each residue type (24% for hydrophobic, 22% for polar, and 54% for charged residues). It is noteworthy that *Pk*-MGMT exposes less hydrophobic and more charged residues to the solvent region. The results suggest that hydrophobicity in the interior, and solvation effect on the surface are enhanced in *Pk*-MGMT. Comparison of the number of aromatic residues between *Pk*-MGMT and AdaC shows remarkable results. *Pk*-MGMT contains 17 aromatic residues (9 phenylalanines, 6 tyrosines and 2 tryptophans) and AdaC contains 8 aromatic residues (3 phenylalanines, 3 tyrosines and 2 tryptophans). In *Pk*-MGMT, there are many aromatic residues that construct aromatic clusters (F29-F31-F162 and F86-Y97-Y135). The clusters are located between the N- and C-terminal domains (**Figure 6a**). Therefore, it is possible considerable that hydrophobic interactions by the aromatic clusters stabilize the internal packing of *Pk*-MGMT.

Ion-pairs

Recent structural studies on hyperthermophilic proteins reveal an increase in the number of ion-pairs, and formation of ion-pair networks, compared with mesophilic versions (Day *et al.*, 1992; Hennig *et al.*, 1995; Korndörfer *et al.*, 1995; Rice *et al.*, 1996; DeDecker *et al.*, 1996; Knapp *et al.*, 1997; Russell *et al.*, 1997). The number of ion-pairs observed in *Pk*-MGMT and AdaC, are identified using cutoff distance between oppositely charged residues of 4.0 (Barlow & Thornton, 1983) and 5.0 Å. **Table 6** shows the total number of ion-pairs and the number of ion-pairs in intra- and inter-helix positions. The number of ion-pair networks formed by over four residues (Rice *et al.*, 1996) are also listed in **Table 6.** Comparison of total ion-pairs and ion-pair networks in numbers shows no significant difference between two methyltransferases in the 5 Å cutoff. In the 4 Å cutoff, the total number of ion-pairs in *Pk*-MGMT is slightly fewer. More intra- and inter-helix ion-pairs are found in *Pk*-MGMT. In the following sections, the role of those ion-pairs is discussed.

Table 6.

Comparison of the number of ion-pairs

	Pk-MGMT	AdaC
5 Å cutoff total	17	16
Intra-helix	7	0
Inter-helix	6	2
Network > 3	1	2
4 Å cutoff total	11	14
Intra-helix	4	0
Inter-helix	4	2
Network > 3	1	1

Charged residues include Arg, Lys, His, Glu and Asp.

1) Intra-helix ion-pairs

It has been reported that the helical conformation is stabilized by [i+4] or [i+3] glutamate-lysine intra-helix ion-pairs in a short

model peptides system (Marqusee & Baldwin, 1987). In *Pk*-MGMT, there are seven intra-helix ion-pairs in α-helices shown in **Figure 8a** (K38-D42, E45-K49, D62-K66, E65-K69, K94-E98, D114-K117, and E158-K161), while no intra-helix ion-pair is found in AdaC. In the present case, it is possible that the intra-helix ion-pairs contribute to reinforce the stability of helical conformation in *Pk*-MGMT. **Figure 7a** shows all the amino acid residues forming [i+4] or [i+3] intra-helix ion-pairs in *Pk*-MGMT highlighted in inverse type. Of more interesting, these amino acid residues are not absolutely conserved in AdaC. Therefore, the charged residues might have been replaced by other residues during environmental changes and evolutionary processes. In the C-terminal domain, amino acid sequence alignment suggests the existence of intra-helix ion-pairs in homologous alkyltransferases of other hyperthermophiles (*M. jannaschii*, *M. thermoautotrophicum*, *A. fulgidus*, and *A. aeolicus*).

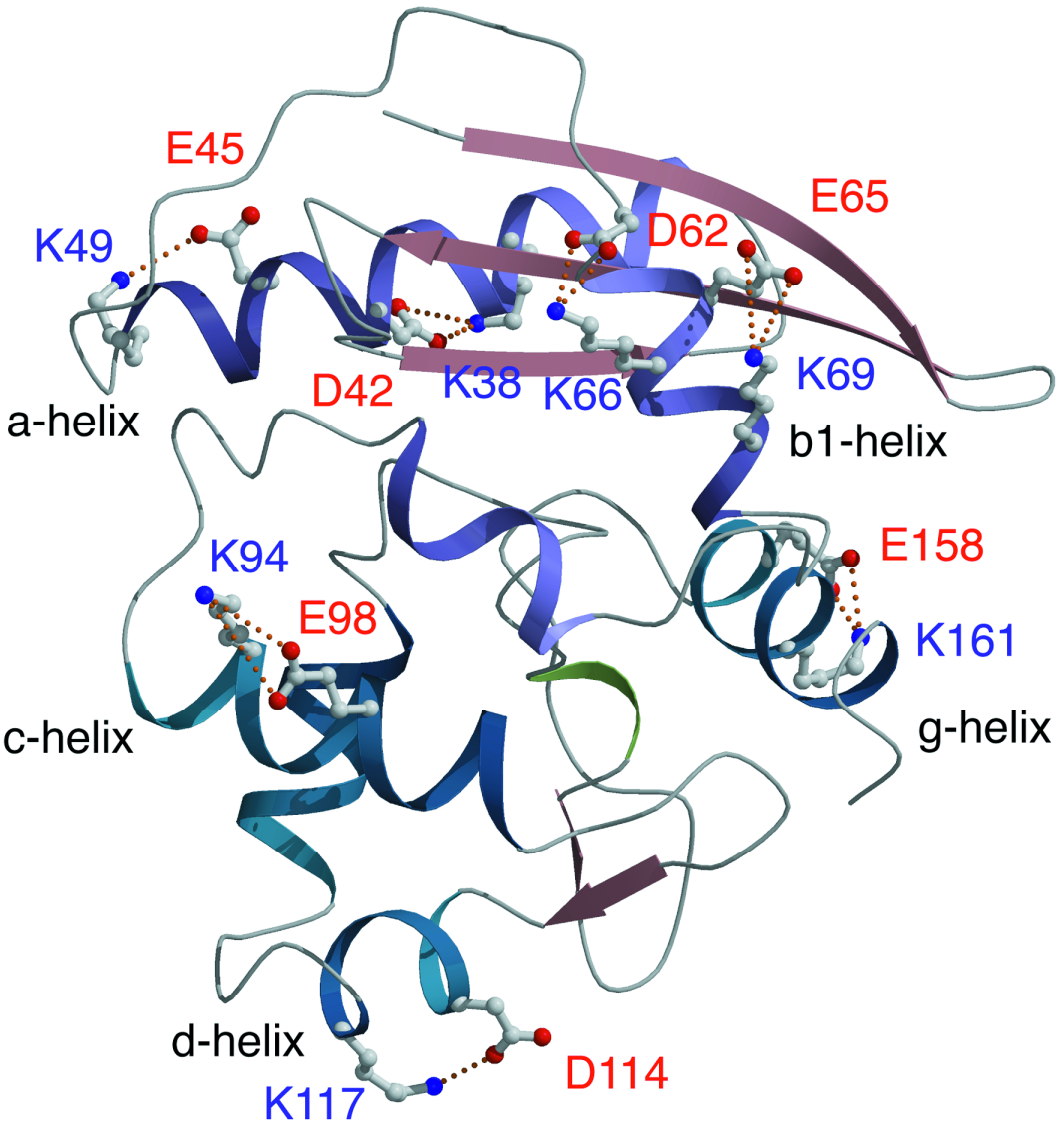

Figure 8a.

Intra-helix ion-pairs in α-helices in the *Pk*-MGMT structure.

A similar increase of Intra-helix ion-pairs compared to

mesophilic homologue is also found in TFIIb (transcription factor IIb) from *P. woesei* (1ais) and IGPS (indole-3-glycerol phosphate synthase) from *S. solfataricus* (1igs). Interestingly, there is no intra-helix ion-pair in e-helix among all a-helices of *Pk*-MGMT. Therefore, the solvent exposed ion-pair may prevent interaction with DNA, because e-helix interacts with substrate DNA. For that reason, no intra-helix ion-pair exists in only e-helix. This consideration is supported by **Figure 7b**, because it does not show the possibility for the existence of intra-helix ion-pairs within e-helix in various methyltransferases.

2) Inter-helix ion-pairs

It is generally agreed that the hydrophobic interaction is the primary means to stabilize internal packing in protein folding. Comparison of the contribution to total the ASA by hydrophobic residues revealed more of an increase of hydrophobicity in the interior of *Pk*-MGMT than in AdaC, as described above. The effect of inter-helix ion-pairs on internal packing is discussed in this section. Although there are two absolutely conserved inter-helix ion-pairs in *Pk*-MGMT; H142-E167 and R143-E167 shown in **Figure 9b** (in AdaC, H147-E173 and R148-E173), *Pk*-MGMT has four extra

inter-helix ion-pairs. The conserved inter-helix ion-pairs (H142-E167 and R143-E167) may be essential for the function of methyltransfer or for the maintenance of fundamental structure, because these conserved inter-helix ion-pairs exist near the active site. **Figure 8b** shows extra inter-helix ion-pairs in *Pk*-MGMT. R39-E159 connects a- and g-helices with a short distance (2.83 Å for NH1-OE2 and 3.02 Å for NH2-OE2). R50-E93 connects a- and c-helices with the shortest distance in *Pk*-MGMT (2.74 Å for NH1-OE1 and 2.83 Å for NH2-OE2). Subsequently, these ion-pairs connect N- and C-terminal domains. Amino acid residues forming inter-helix ion-pairs are buried relatively more than that of intra-helix ion-pairs, because the averaged ASAs of amino acid residues forming inter- and intra- ion-pairs are 39.5 and 98.9 $Å^2$/residue, respectively. Therefore inter-helix ion-pairs generally exist in the interior of the molecule in the present case. It is possible that intra-helix ion-pairs link helical conformations in the interior of the molecule and that this helix network stabilizes internal packing of the tertiary structure in cooperation with hydrophobic interaction.

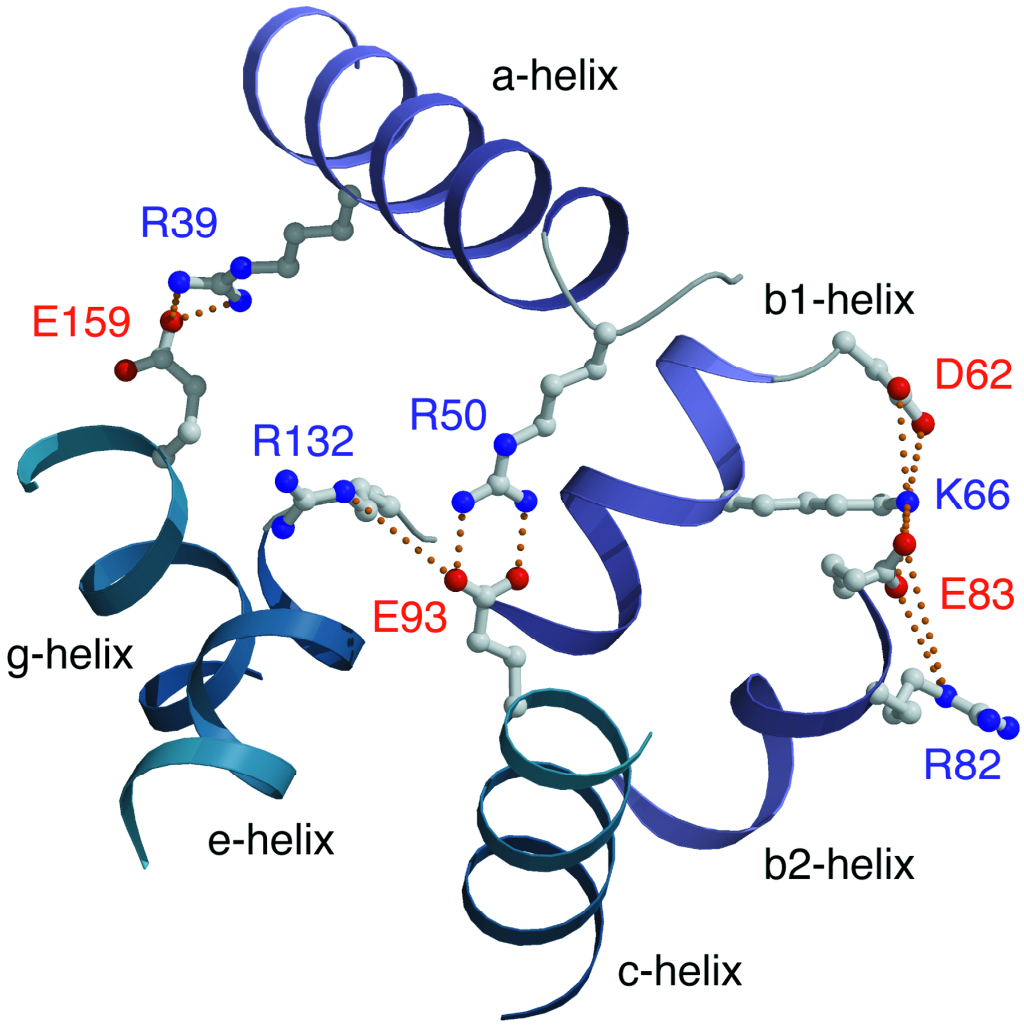

Figure 8b.

Inter-helix ion-pairs in *Pk*-MGMT are shown. R39-E159 and R50-E93 ion-pairs connect N- and C-terminal domains. D62-K66 ion-pair in b1-helix is connected to E83-R82 ion-pair in b2-helix. R82 and E83 make [i+1] intra-helix ion-pair in 3_{10}-helix. This network links b1 and b2 helices, and the 3_{10}-helix (b2-helix) is especially stabilized because of the helix-network.

Maintenance of the active site structure and alkyltransfer reaction

In the alkyltransfer reaction process, the protein is initially required to bind substrate DNA. Helix-loop-helix motif will interact with DNA (c-, d- and e-helices). Mutagenic experiments suggest that the positive charge of R124 plays an important role in interacting with negative charges of the phosphate backbone of DNA (Kanugula *et al.*, 1995). Stability of the DNA binding motif should be required in extreme environments. **Figure 9** shows the detail of the DNA binding region and the active site structure in *Pk*-MGMT. Suitable residues (T90, T111 and S122) are located in all N-cap positions of the three helices, and P91 and P123 are located in the N1 positions of the two helices. Furthermore, the two intra-helix ion-pairs (K94-E98, D114-K117) also reinforce helical conformations. The electron density maps shown in **Figure 9a** indicate two possible conformers for E98. Therefore, E98 may pair with K94 or K102, or both. E98 may change its ion-pair partner during the conformational change in the interaction with DNA. Furthermore, the E93-R132 and R50-E93 inter-helix ion-pairs provide a stable environment around the active site. The structural features described above; helix cappings, intra- and inter-helix ion-pairs, play an important role in maintaining the DNA binding

region of *Pk*-MGMT.

The next step in the reaction process is recognition of the flipped-out alkylguanine. In 1994, one of the DNA binding models was reported (Moore *et al,* 1994). In that model, C-terminal g-helix is required to swivel by movement about the loop residues from β6 to g-helix. The g-helix interacts with the major groove of DNA. **Figure 9b** shows that the g-helix is fixed by two inter-helix ion-pairs (R39-E159 and R106-E167) in *Pk*-MGMT, and the two inter-helix ion-pairs are not conserved with AdaC. The R39-E159 ion-pair connects the a- and g-helix with very short distances (2.83 Å for NH1-OE2 and 3.02 Å for NH2-OE2). The interactions related to g-helix suggest difficulty for great movement of g-helix during the reaction. It is possible that the two ion-pairs prevent great movement of the g-helix and contribute to the structural maintenance of the active site, although the helix is required to make only slight movement. Another theoretical model suggests the possibility for DNA binding without great movement of the g-helix (Vora *et al.*, 1998). **Figure 9b** shows the active site environment of *Pk*-MGMT. If the wing loop moves to expose the active site, the protein can interact with the substrate. In the crystal structure of *Pk*-MGMT, the average temperature factor of main chain atoms

in the wing loop (146~157) is 25.0 Å^2, which is higher than that of all main chain atoms (15.4 Å^2). Thus, the wing loop, including three glycines, is relatively flexible. It is possible that the wing loop moves to open the active site, and then Y112, Y152 and Y153 are exposed to the substrate and recognize the aromatic ring of flipped-out base that invites alkylguanine into the active site.

Such are the structural features which may serve to stabilize *Pk*-MGMT. Analysis of ASA reveals that *Pk*-MGMT exposes less hydrophobic and more charged residues to the solvent region, and there are many ion-pairs existing in intra- and inter-helix positions. This suggests that the intra-helix ion-pairs on the surface assist in reinforcing the stability of helical conformations (secondary structures), and inter-helix ion-pairs in the interior enhance the stability of internal packing (tertiary structure). Additional experiments such as site-directed mutagenesis together with thermodynamic stability measurements are necessary to demonstrate the effects of the intra- and inter-helix ion-pair.

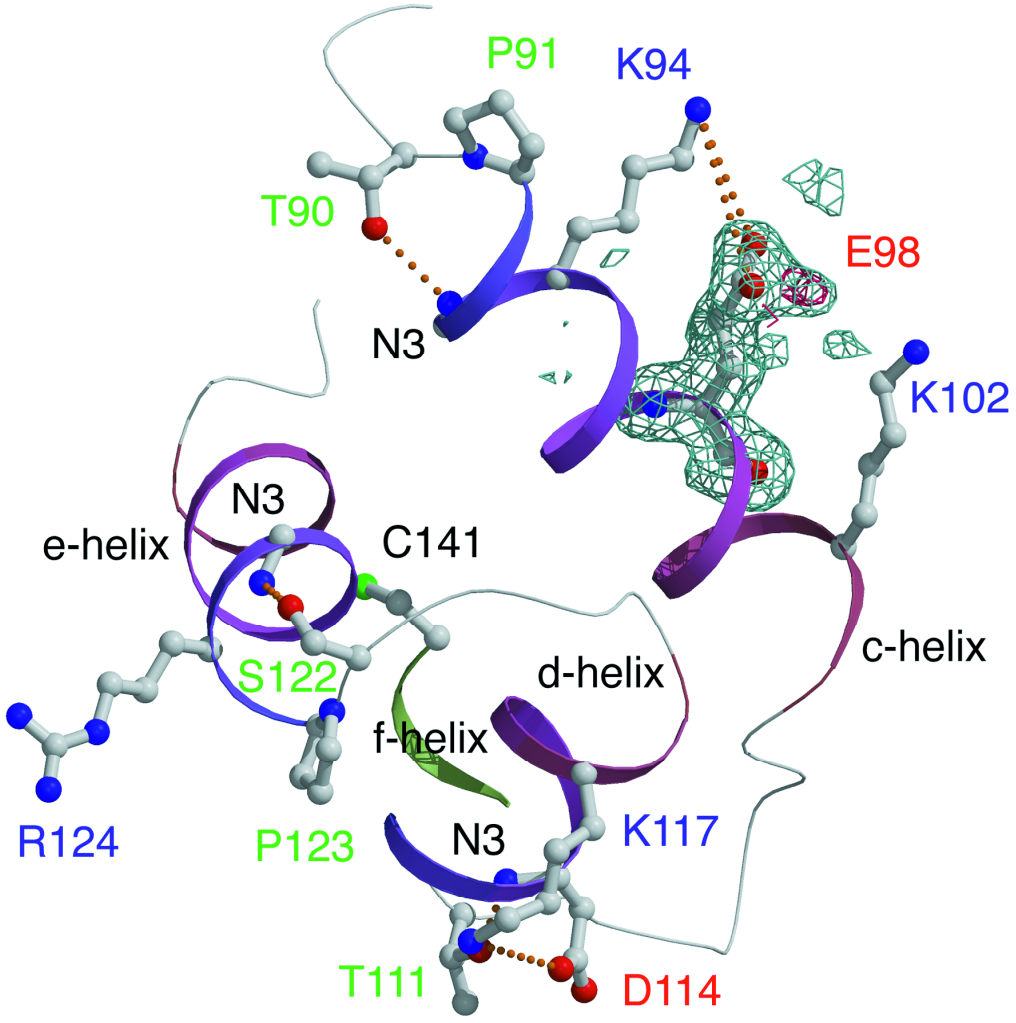

Figure 9a.

The DNA-binding region of *PK*-MGMT, as viewed from the bottom of **Figure 6a**. DNA binds from the left side. Intra-helix ion-pairs in c- and d-helices are shown by ball and stick models. Dark and light grey cages are the |*F*o|-|*F*c| map and the |*F*o|-|*F*c| omit map contoured at 2.0σ, respectively. The |*F*o|-|*F*c| omit map was calculated to omit E98 in the final models. The maps suggest dual conformation of E98.

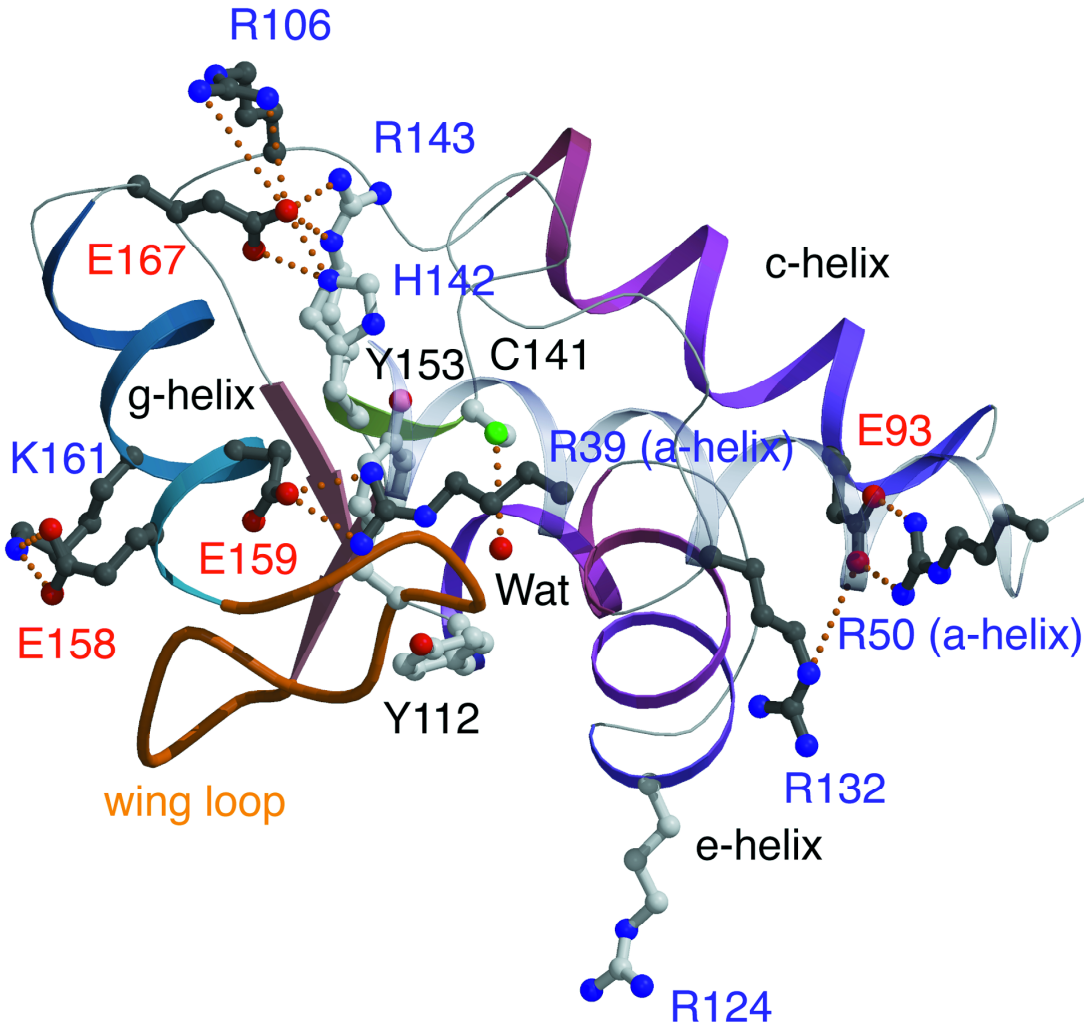

Figure 9b.

The active site environment of *Pk*-MGMT, as viewed from the top of **Figure 6a**. DNA binds from the bottom side. Extra ion-pairs are shown by ball-and-stick models (colored dark gray). The semitransparent helix is the a-helix in N-terminal domain. The helix hangs on to the active site by two inter-helix ion-pairs (R39-E159 and R50-E93). Stability of helical conformation in g-helix is reinforced by the E158-K161 intra-helix ion-pair.

2.4. REFERENCES

Barlow, D. J. & Thornton, J. M. (1983). Ion-pairs in proteins. *J. Mol. Biol.* **168**, 627-885.

Barton, G. J. (1993). ALSCRIPT: a tool to format multiple sequence alignments. *Protein Eng.* **5**, 37-40.

Brünger, A.T., Kuriyan, J. & Karplus, M. (1987). Crystallographic R factor refinement by molecular dynamics. *Science* **235**, 458-460.

Brünger, A.T. & Krukowski, A. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Crystallogr. A***46**, 585-593.

Brünger, A.T. (1992). The free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **335**, 472-474.

CCP4, The Collaborative Computational Project, Number 4. (1994). *The CCP4 Suite: Programs for Protein Crystallography, Acta Crystallogr. D***50**, 760-763.

Chan, C. L., Wu, Z., Ciardelli, T., Eastman, A. & Bresnick, E. (1993). Kinetic and DNA-binding properties of recombinant human O^6-methylguanine-DNA methyltransferase. *Arch. Biochem. Biophys.* **300**, 193-200.

Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W. W. & Rees, D. C. (1995). Structure of a Hyperthermophilic Tungstopterin Enzyme, Aldehyde Ferredoxin Oxidoreductase. *Science* **267**, 1463-1469.

Coulondre, C. & Miller, J. H. (1977). Genetic Studies of the *lac* Repressor. *J. Mol. Biol.* **117**, 577-606.

Cowtan, K. (1994). An automated procedure for phase improvement by density modification. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34-38.

Day, M. W., *et al.*, & Rees, D. C. (1992). X-ray crystal structures of the oxidized and reduced forms of the rubredoxin from the marine hyperthermophilic archaebacterium *Pyrococcus furiosus*. *Protein Sci.* **1**, 1494-1507.

DeDecker, B. S., Brien, R. O., Fleming, P. J., Geiger, J. H., Jackson, S. P. & Sigler, P. B. (1996). The Crystal Structure of a Hyperthermophilic Archaeal TATA-box Binding Protein. *J. Mol. Biol.* **264**, 1072-1084.

Fersht, A. R. & Serrano, L. (1993). Principles of protein stability derived from protein engineering experiments. *Curr. Opin. Struct. Biol.* **3**, 75-83.

Fujiwara, S., Okuyama, S. & Imanaka, T. (1996). The world of archaea: genome analysis, evolution and thermostable enzymes. *Gene* **179**, 165-170.

Goodtzova, K., Kanugula, S., Edara, S., Pegg, A. E. (1996) The role of tyrosine 114 in O^6-alkylguanine-DNA alkyltransferase activity. *Proc. Am. Assoc. Cancer Res.* **37**, A2484.

Hashimoto, H., *et al.*, & Kai, Y. (1998). Crystallization and preliminary X-ray crystallographic analysis of archaeal O^6-methylguanine-DNA methyltransferase. *Acta Crystallogr. D***54**,

1395-1396.

Hennig, M., Darimont, B., Sterner, R., Kirschner, K. & Jansonius, J. N. (1995). 2.0 Å structure of indole-3-glycerol phosphate synthase from the hyperthermophile Sulfolobus solfataricus: possible determinants of protein stability. *Structure*, **3**, 1295-1306.

Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1994). Improved methods for building protein models in electron density maps and the location of errors in the model. *Acta Crystallogr. A***47**, 110-119.

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

Kanugula, S., Goodtzova, K., Edara, S., Pegg, A. E. (1995). Alteration of arginine-128 to alanine abolishes the ability of human O^6-alkylguanine-DNA alkyltransferase to repair methylated DNA but has no effect on its reaction with O^6-benzylguanine.

Biochemistry **34**, 7113-7119.

Knapp, S., de Vos, W. M., Rice, D. & Ladenstein, R. (1997). Crystal Structure of Glutamate Dehydrogenase from the Hyperthermophilic Eubacterium *Thermotoga maritima* at 3.0 Å Resolution. *J. Mol. Biol.* **267**, 916-932.

Korndörfer, I., Steipe, B., Huber, R., Tomschy, A. & Jaenicke, R. (1995). The crystal structure of hologlyceraldehyde-3-phosphate dehydrogenase from the hyperthermophilic bacterium *Thermotoga maritima* at 2.5 Å resolution. *J. Mol. Biol.* **246**, 511-521.

Kraulis, P. J. (1991). *MOLSCRIPT*: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946-950.

Leclere, M. M., Nishioka, M., Yuasa, T., Fujiwara, S., Takagi, M & Imanaka, T. (1998). The O^6-methylguanine-DNA methyltransferase from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1: a thermostable repair enzyme. *Mol. Gen. Genet.* **258**, 69-77.

Marqusee, S. & Baldwin, R. L. (1987). Helix stabilization by $Glu^{-}...Lys^{+}$ salt bridges in short peptides of *de novo* design. *Proc. Natl. Acad. Sci. USA* **84**, 8898-8902.

Matthews, B. W. (1993). STRUCTURAL AND GENETIC ANALYSIS OF PROTEIN STABILITY. *Annu. Rev. Biochem.* **62**, 139-160.

Matthews, B. W. (1968). Solvent Content of Protein Crystals. *J. Mol. Biol.* **33**, 491-497.

Merritt, E. A. & Murphy, M. E. (1994). Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. D***50**, 869-873.

Merritt, E. A. & Bacon, D. J. (1997). Raster3D: Photorealistic Molecular Graphics. *Meth. Enzymol.* **277**, 505-524.

Moore, M. H., Gulbis, M., Dodson, E. J., Demples, B. & Moody, P. C. E. (1994). Crystal structure of a suicidal DNA repair protein: the Ada O^{6}-methylguanine-DNA methyltransferase from *E.coli*. *EMBO.*

J. **13**, 1495-1501.

Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. & Imanaka, T. (1994). Purification and characterization of a hyperthermophilic *Pyrococcus* sp. *Appl. Environ. Microbiol.* **60**, 4559-4566.

Morris, A. L., MacArthur, M. W. & Thornton, J. N. (1992). Stereochemical quality of protein structure coordinates. *Proteins* **12**, 345-364.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr. D***53**, 240-255.

Navaza, J. (1994) AMORE: an automated package for molecular replacement. *Acta Crystallogr. A***50**, 157-163.

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

Otwinowski, Z. (1991). Maximum likelihood refinement of heavy atom

parameters. In *Isomorphous Replacement and Anomalous Scattering.* (eds Wolf, W., Evans, P. R. & Leslie, A. G. W.) 80-86 (Daresbury Laboratory, Daresbury, UK).

Rice, D. W., *et al.*, & Engel, P. C. (1996). Insight into the molecular basis of thermal stability from the structure determination of *Pyrococcus furiosus* glutamate dehydrogenase. *FEMS. Microbiol. Rev.* **18**, 105-117.

Richardson, J. S. & Richardson, D. C. (1988). Amino Acid Preference for Specific Locations at the Ends of a Helices. *Science* **240**, 1648-1452.

Russell, R. J. M., Ferguson, J. M. C., Hough, D. W., Danson, M. J. & Taylor, G. L. (1997). The Crystal Structure of Citrate Synthase from the Hyperthermophilic Archaeon *Pyrococcus furiosus* at 1.9 Å Resolution. *Biochemistry* **36**, 9983-9994.

Sakabe, N., *et al.*, & Sasaki, K. (1995). Weissenberg Camera for Macromolecules with Imaging Plate Data Collection System at the Photon Factory, Present Status and Future Plan. *Rev. Sci. Instrum.*

66, 1276-1281.

Serrano, L. & Fersht, A. R. (1989). Capping and a-helix stability. *Nature* **342,** 296-299.

Takahashi, M., Sakumi, K. & Sekiguchi, M. (1988). Interaction of Ada protein with DNA examined by fluorescence anisotropy of the protein. *Biochemistry* **29,** 3431-3436.

Vora, R. A., Pegg, A. E. & Ealick, S. E. (1998). A New Model for How O^6-Methylguanine-DNA Methyltransferase Binds DNA. *Proteins* **32,** 3-6.

Watanabe, N., Nakagawa, A., Adachi, S. & Sakabe, N. (1995). Macromolecular Crystallography Station BL-18B at the Photon Factory. *Rev. Sci. Instrum.* **66,** 1824-1826.

CONCLUSION

==========

Chapter 1: The hyperthermostable family B DNA polymerase (EC 2.7.7.7) from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1 (KOD DNA polymerase) has been crystallized. The diffraction pattern of the crystal extends to 3.0 Å resolution. The crystal structure of KOD DNA polymerase was determined by molecular replacement method at 3.0 Å resolution. The crystal structure revealed the high PCR performance of KOD DNA polymerase.

Chapter 2: The crystal structure of O^6-methylguanine-DNA methyltransferase (EC 2.1.1.63) of hyperthermophilic archaeon *Pyrococcus kodakaraensis* strain KOD1 (*Pk*-MGMT) was determined by SIRAS method at 1.8 Å resolution. In the *Pk*-MGMT structure, the intra-helix ion-pairs contribute to reinforce stability of α-helices. The inter-helix ion-pairs exist in the interior of *Pk*-MGMT and stabilize internal packing of tertiary structure.

The author believes that these new findings in this study provide useful information to clear the structure-function relationship in the field of structural biology.

LIST OF PUBLICATIONS

==========

1) Hyperthermostable protein structure maintained by intra- and inter- helix Ion-pairs in archaeal O^6-methylguanine-DNA methyltransferase.
   Hashimoto, H., Inoue, T., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., & Kai, Y.
   J. Mol. Biol. **292**, 707-716, (1999).

2) Crystallographic Studies on a Family B DNA Polymerase from Hyperthermophilic Archaeon *Pyrococcus kodakaraensis* Strain KOD1.
   Hashimoto, H., Matsumoto, T., Nishioka, M., Yuasa, T., Takeuchi, S., Inoue, T., Fujiwara, S., Takagi, M., Imanaka, T., & Kai, Y.
   J. Biochem. **125**, 983-986, (1999).

3) Crystallization and preliminary X-ray crystallographic analysis of archaeal *O*^{6}-methylguanine-DNA methyltransferase.
   Hashimoto, H., Nishioka, M., Inoue, T., Fujiwara, S., Takagi, M., Imanaka, T., & Kai, Y.
   *Acta Crystallogr. D***54**, 1395-1396, (1998).

LIST OF SUPPLEMENTARY PUBLICATIONS

==========

1) Syntheses, Structures and Solution Dynamics of Anionic 5-Coordinate Pt(II) Complexes with Halide.
   Okeya, S., Hashimoto, T., Yamanaka, K., Sumino, T., Hashimoto, H., Kanehisa, N., & Kai, Y.
   Chem. Lett. 541-542, (1998).

2) Synthesis and Olefin Polymerization Catalysis of New Divalent Samarium Complexes with Bridging Bis(cyclopentadienyl) Ligands.
   Ihara, E., Nodono, M., Katsura, K., Adachi, Y., Yasuda, H., Ymagashira, M., Hashimoto, H., Kanehisa, N. & Kai, Y.
   Organometallics **17**, 3945-3956, (1998).

3) Substituent Effects on the Second Hyperpolarizability of Thioxanthen-9-one and Its Derivatives.
   Sugino, T., Hashimoto, H., Muguruma, Y., Kambe, N., Sonoda, N., Sakaguchi, T.
   Main Group Chem. **1**, 325-330, (1996).