Studies on membrane-bound peptidases and a sugar transporter in the hyperthermophilic archaeon *Thermococcus kodakaraensis*

Rie Matsumi

2008

PREFACE

This is a thesis submitted by the author to Kyoto University for the degree of Doctor of Engineering. The study presented here has been performed under the supervision of Professor Tadayuki Imanaka in the Laboratory of Biochemical Engineering, Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, 2002-2008.

The author would like to express her sincerest gratitude to Professor Tadayuki Imanaka for his invaluable guidance, worthy suggestions, and ceaseless encouragement throughout the course of this study. The author is deeply grateful to Associate Professor Haruyuki Atomi for fruitful discussion, precious suggestions, and heartwarming support throughout this work. The author is also thankful to Assistant Professor Toshiaki Fukui for his valuable suggestions and helpful discussion. The author is also grateful to Assistant Professor Tamotsu Kanai for his helpful suggestions and assistance. The author would like to acknowledge Assistant Professor Satoshi Ezaki for his guidance during the early times in the Imanaka Laboratory (-2002), through which the author was able to learn many fundamental techniques of microbiology.

The author cannot forget to express her deep gratitude to all the colleagues, past and present, in the Imanaka Laboratory. Without their continuous support and encouragement, she would not have been able to complete this study. The author would like to emphasize that the studies reported here were made possible by the accumulation of knowledge and technology established by the previous members of the Laboratory.

The author acknowledges Professor Tadayuki Imanaka for financial support through projects funded by the Japan Science and Technology Corporation for Core Research for Evolutional Science and Technology (CREST) (2000-2002), the National Project on Protein Structural and Functional Analyses (2002-2006), and Grant-in-Aid for Scientific Research on Priority Areas “System Cell Engineering by Multi-scale Manipulation” (2006-present) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

The author also acknowledges research funding (Grant-in Aid, 2006-2008) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

Rie Matsumi

Laboratory of Biochemical Engineering

Department of Synthetic Chemistry and Biological Chemistry

Graduate School of Engineering

Kyoto University

2008

Kyoto

TABLE OF CONTENTS

		Pages
GENERAL INTRODUCTION		**1**
SYNOPSIS		**22**
Part I	**Studies on membrane-bound peptidases from *Thermococcus kodakaraensis***	
Chapter 1	**Biochemical properties of a putative signal peptide peptidase from *Thermococcus kodakaraensis***	**28**
Chapter 2	**Identification of the amino acid residues essential for proteolytic activity in an archaeal signal peptide peptidase**	**59**
Chapter 3	**Biochemical properties of a novel membrane-bound peptidase from *Thermococcus kodakaraensis***	**85**
Chapter 4	**Identification of the amino acid residues essential for proteolytic activity of the membrane-bound peptidase $SppB_{Tk}$ from *Thermcoccus kodakaraensis***	**110**
Part II	**Studies on a sugar transporter in *Thermococcus kodakaraensis* and development of a gene disruption system based on antibiotic resistance**	
Chapter 5	**Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance**	**130**
GENERAL CONCLUSIONS		**157**
LIST OF RELATED PUBLICATIONS		**160**
OTHER PUBLICATIONS		**161**

GENERAL INTRODUCTION

1. Hyperthermophilic archaea and the evolution of life

All living organisms on our planet can be divided into three domains; *Bacteria*, *Eucarya* and *Archaea*. The *Archaea* were first recognized in the 1970's by Carl Woese through phylogenetic analyses based on sequences of 16S or 18S ribosomal RNA genes (1, 2) (Fig. 1). The *Archaea* are further classified into two major kingdoms; *Crenarchaeota* and *Euryarchaeota*. The *Archaea* are mainly comprised from the methanogens, halophiles, sulfate reducers, and (hyper)thermophiles.

Hyperthermophiles are defined as organisms that grow optimally at temperatures above 80˚C (3). They were discovered in the early 1980's (4, 5) and represent one of the six major groups of extremophiles; hyperthermophiles, psychrophiles (cold temperature), alkaliphiles (high pH), acidophiles (low pH), halophiles (high salinity) and piezophiles (high pressure) (6). Most hyperthermophiles are members of the *Archaea* with the exception of several bacterial genera, *Aquifex*, *Thermotoga* and *Geothermobacterium*. All hyperthermophiles are positioned near the root of the phylogenetic tree and a majority are obligate anaerobes. Together with the assumption that the earth was a thermal and anaerobic environment in the early periods when life originated, it has been proposed that the origin of life, or at least the Last Universal Common Ancestor (LUCA), was a hyperthermophilic form of life (7).

There has been one strong argument against the proposal that life originated in the form of hyperthermophiles (the hot origin of life). This was based on the structure and distribution of an enzyme called reverse gyrase. Reverse gyrase introduces positive supercoils in covalently closed DNA (8) and is the only enzyme/gene that is present in

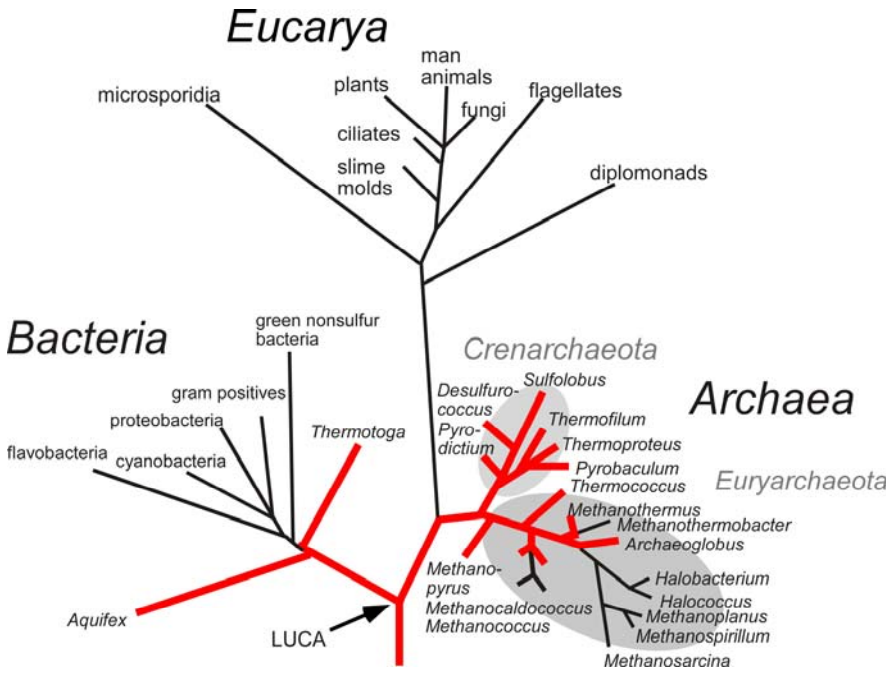


Fig. 1. A rooted phylogenetic tree of life based on 16S/18S ribosomal RNA sequences. The tree reveals the three domains of life, *Archaea*, *Bacteria* and *Eucarya*. The *Euryarchaeota* and *Crenarchaeota* branches are shaded. LUCA is indicated by an arrow. Hyperthermophiles are indicated in red.

all hyperthermophilic organisms but absent in all mesophilic organisms, *i. e.* it is the one and only hyperthermophile-specific protein (9). On the other hand, the structure of reverse gyrase suggests that it is not at all a primitive enzyme in terms of protein evolution. Reverse gyrase is formed by the association of two entirely different enzymes belonging to the DNA/RNA helicase and the topoisomerase families (10, 11), and thus could only have evolved after the diversification of the respective protein families. If reverse gyrase were to be a prerequisite for life at high temperatures, this would provide a convincing argument that contradicts with the hot origin of life. The evolution of the two protein families could only have occurred in less thermophilic organisms (11) (Fig. 2).

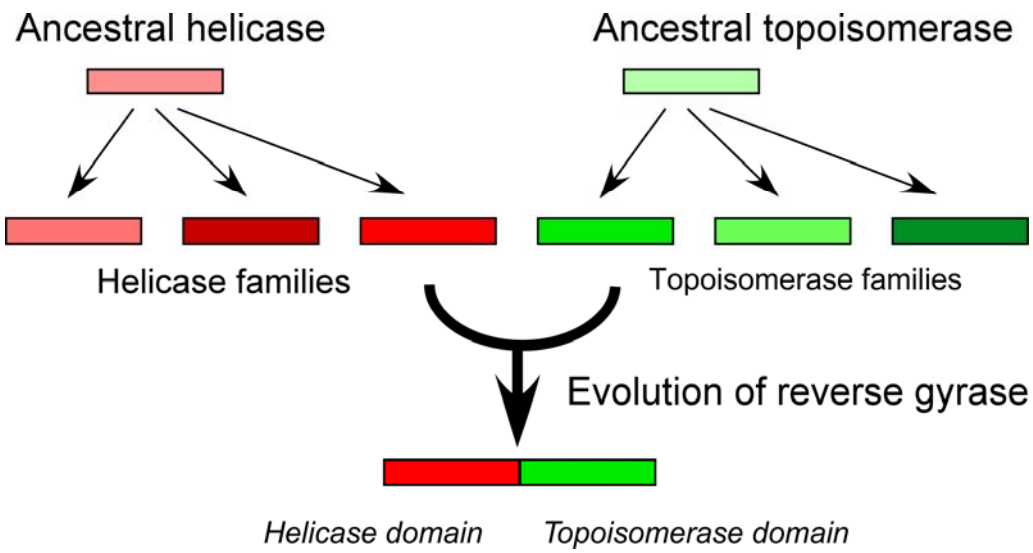


Fig. 2. The evolution of reverse gyrase. The function of the enzyme is dependent on two domains (helicase and topoisomerase) that are members of two entirely different protein superfamilies.

In order to evaluate the necessity of reverse gyrase for hyperthermophilic life, prior to the studies reported in this thesis, the author attempted to disrupt the reverse gyrase gene of the hyperthermophilic archaeon, *Thermococcus kodakaraensis* (the organism is described below) (12). Disruption of the gene was not lethal, but specific growth rates of the mutant strain declined with increases in temperature above 80˚C compared with the original strain. The gene disruption strain was able to grow at 90˚C but not at higher temperatures (Fig. 3). The results indicate that reverse gyrase provides a significant advantage for life at high temperatures (>80˚C), and helps in understanding why all organisms isolated from hyperthermophilic environments until now harbor a reverse gyrase. The results also revive the possibilities of a hot origin of life in which primitive hyperthermophiles without a reverse gyrase might have been the first organisms to evolve, most likely at temperatures below 90˚C.

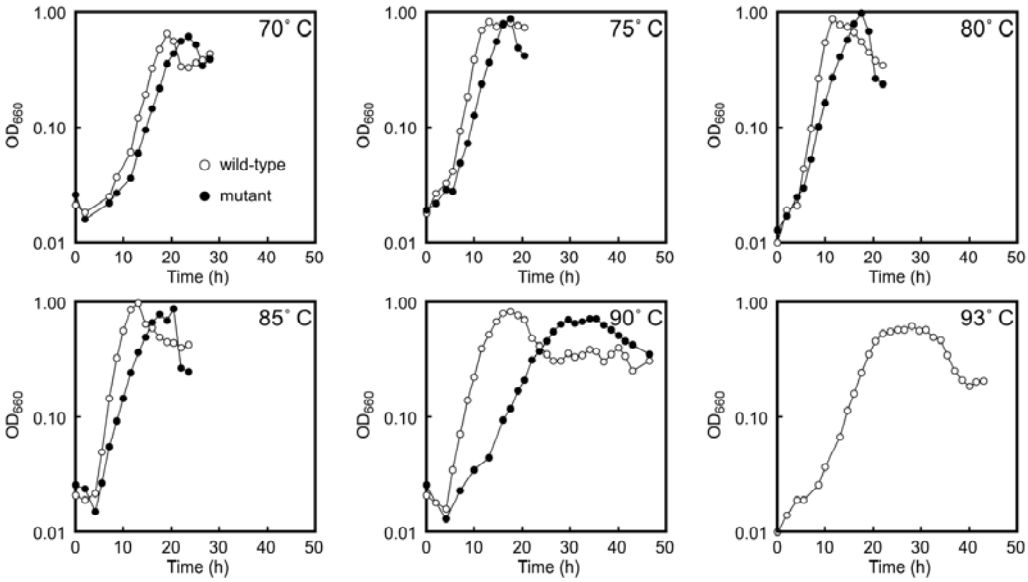


Fig. 3. Growth characteristics of the wild-type *T. kodakaraensis* strain and a strain lacking the reverse gyrase gene (mutant) at various temperatures. Temperatures are indicated in each panel and cell growth was monitored by measuring the optical density of the medium at 660 nm.

2. Biochemical characteristics of the *Archaea*

Members of the *Archaea* and *Bacteria* are prokaryotes, and share many common characteristics including similar cell size, the absence of organelles, circular structures of their chromosomes, and the presence of gene operons. However, the *Archaea* display numerous features that are not found in the *Bacteria*; some unique to the *Archaea* and others that had previously been presumed to be traits specific to the *Eucarya*. One feature that is specific to the *Archaea* is their unusual membrane composition. Archaeal lipids consist of C_{20}-C_{40} isoprenoid backbones linked to glycerol via ether bonds (13). The transcription mechanism and machinery utilized in the *Archaea* are closely related to those of the *Eucarya* (14-16). σ-factor dependent RNA polymerase is not found in the *Archaea*, and instead transcription initiation is brought about by the presence of the TATA box, and factors such as TATA-binding protein,

Transcription Factor IIB and a eukaryotic-type RNA polymerase. DNA replication and repair mechanisms in the *Archaea* also closely resemble those that function in eukaryotes (17, 18).

As the *Archaea* represent a third form of life that can be distinguished from the bacteria and eukaryotes, complete genome sequences of many archaeal strains have been determined. At present, the entire genome sequences of 47 archaeal strains, including 21 hyperthermophiles, are available. This has revealed that hyperthermophiles can maintain life with the function of only 1,500-3,000 genes. The small number of genes suggests that their metabolism and biological machinery are simple, which should provide an advantage in examining the basic mechanisms of various biological phenomena. Studies on hyperthermophiles are also attractive in terms of enzyme application (19, 20). All proteins in a hyperthermophile must properly function at high temperature ranges, making them much more (thermo)stable when compared to the conventional enzymes from mesophiles utilized at present.

3. Protein secretion and signal peptide peptidase

3-1. Protein secretion in bacteria and eukaryotes

In bacteria and eukaryotes, the majority of proteins to be secreted from the cell are produced with an extension in their amino-termini that directs the precursor proteins towards the secretion machinery (21, 22). These regions are called signal peptides and are cleaved during the translocation process through the membranes of the cytoplasm (bacteria) or endoplasmic reticulum (ER) (eukaryotes) (23). The main mode of protein transport in bacteria is post-translational, and the precursor proteins are maintained in a translocation-competent conformation by cytosolic chaperones such as SecB. The

precursor protein is then delivered to the ATPase SecA, which provides the driving force of protein translocation through the translocation channel composed of SecY, SecE and SecG (22, 24, 25). In eukaryotes, the majority of proteins are transported into the ER in a co-translational manner. The cytoplasmic Signal Recognition Particle (SRP) recognizes the signal peptide during the early stage of translation. This triggers an arrest in translation, during which the SRP-nascent chain-ribosome complex is directed to the protein translocation channel on the ER membrane. Release of SRP allows translation to resume, and the ribosome, which remains docked to the channel, provides the driving force of protein translocation through protein synthesis. The three proteins that constitute the eukaryotic channel are Sec61α (SecY homolog), Sec61β (distantly related to SecG) and Sec61γ (SecE homolog) (21, 26).

3-2. Protein secretion machinery in the *Archaea*

In the *Archaea*, research on protein secretion and the machinery involved are still at an early stage. Sequence comparison of genome data have shed light on the features of the archaeal signal peptide (27), and have also indicated the presence or absence of individual components corresponding to eukaryotic or bacterial factors participating in protein secretion (28-30). The archaeal genomes harbor homologs corresponding to the protein-conducting channel found in bacteria and eukaryotes (31, 32), and the crystal structure of the channel from *Methanocaldococcus jannaschii* has been determined (33). The archaeal genomes do not harbor homologs of SecA nor SecB (34), but possess genes corresponding to the components of an SRP (35, 36), suggesting a eukaryotic mode of protein translocation.

3-3. Signal peptides, signal peptidases and signal peptide peptidases

The general features of signal peptides present in the amino-terminal regions of proteins exported from bacterial cells have been clarified (37). A positively charged domain (n-region) with basic residues is near the extreme amino-terminus, followed by a central hydrophobic domain (h-region) and a carboxy-terminal hydrophilic domain (c-region) that includes the cleavage site for processing of the precursor protein. The signal peptides that are responsible for protein translocation to the eukaryotic ER display almost identical characteristics. It has been reported that eukaryotic signal peptides are slightly more hydrophobic than their bacterial counterparts. In both cases, the sequence that designates the cleavage site is Ala-X-Ala (Fig. 4).

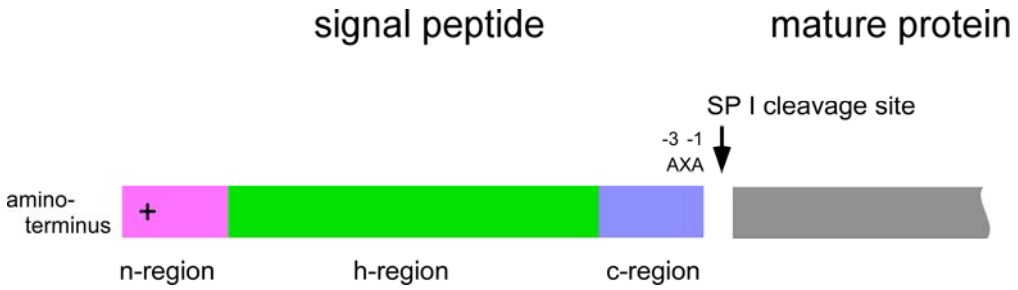


Fig. 4. A diagram illustrating the general features of signal peptides present in the amino-terminus of secretion proteins. Signal peptides are comprised from three regions; a positively charged amino-terminus (n-region, 1-5 amino acids) shown in pink, a core of at least six hydrophobic amino acids (h-region, 7-15 amino acids) shown in green, and an uncharged polar carboxy-terminal region (c-region, 3-7 amino acids) shown in blue. The Ala-X-Ala signature that designates the cleavage site by signal peptidases is indicated as AXA.

Signal peptidases (SPs) cleave the precursor protein during the translocation process, resulting in the release of the signal peptide (Fig. 5). Signal peptidase I (SP I) is the major enzyme in both eukaryotes and bacteria and is essential for cell viability. The

eukaryotic enzyme is a component of a multi-protein complex whereas the bacterial enzyme acts independently. All SP I proteins from both bacteria and eukaryotes harbor a number of conserved regions (boxes A-E) that are considered to play important roles in catalysis (37).

Signal peptide peptidases (SPPs) are enzymes considered to cleave the signal peptide chains of secreted proteins after they are released from the precursor proteins by SP (37, 38) (Fig. 5). In bacteria, SPP was first identified as a membrane-anchored protein involved in the breakdown of the signal peptide of the outer membrane lipoprotein in *Escherichia coli* (39-41). Gene disruption studies strongly implied that SPPs along with other cytoplasmic proteases including oligopeptidase A are involved in signal peptide degradation (42, 43). It is now presumed that in *E. coli*, SPP initiates the degradation by introducing endoproteolytic cuts into the signal peptide, whereas the other cytoplasmic proteases are responsible for complete degradation of the smaller fragments into free amino acids (44). Eukaryotic SPPs are intramembrane enzymes, with activity dependent on two aspartate residues. It had long been presumed that the signal peptides, after their removal from the precursor protein, have no active function in the cell and are simply degraded to free amino acids. However, it is now known that in human cells, peptide fragments generated after cleavage by SPP exhibit vital regulatory functions in immune surveillance (45, 46). The function of SPP has also been reported to be necessary for the proper development of *Drosophila* larvae (47).

In the *Archaea*, genes encoding proteins with putative signal peptides similar to the bacterial and eukaryotic sequences are found in abundance (8-32% of total genes) on the archaeal genomes (48). Archaeal signal peptides are 20-30 residues in length and contain the classical n-, h- and c-regions. In terms of SPs, genes resembling the

eukaryotic SP I gene are found on the archaeal genomes (49). The SP I from *Methanococcus voltae* has been characterized, and has been found to exhibit the expected SP activity. A catalytic triad comprised from Ser52, His122, and Asp148 has been determined to be critical for its peptidase activity (50, 51). Although not described in this thesis, the author was not able to disrupt the corresponding gene in *T. kodakaraensis*, suggesting that as in the case of bacteria and eukaryotes, the archaeal SP is also essential for cell viability. In contrast to the progress on SPs, the presence or examination of archaeal SPPs has not been reported in the *Archaea*.

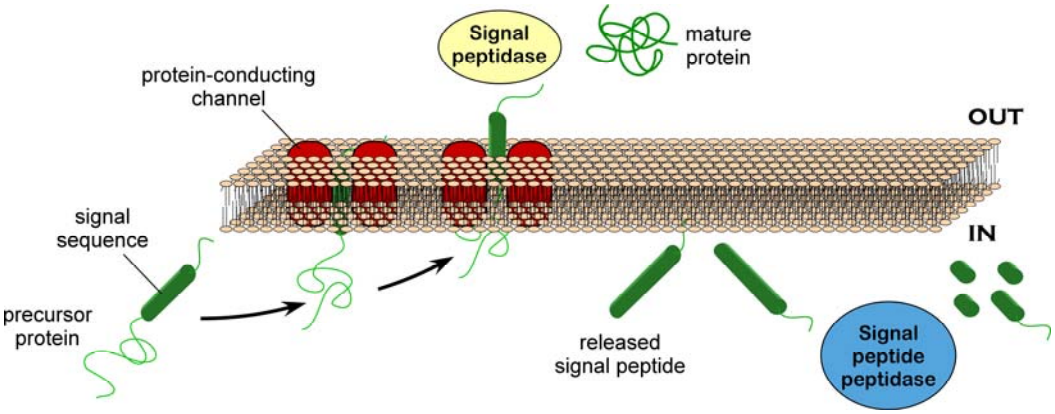


Fig. 5. Schematic model of the function of signal peptidases (SPs) and signal peptide peptidases (SPPs) in the protein secretion pathway. For simplicity, various factors related to protein secretion such as SRP and SecA are not included. The topologies of SPs and SPPs relative to the cytoplasmic membrane do not necessarily reflect their topologies in the cell.

4. Serine proteases

Proteases are ubiquitous in nature and play indispensable roles in both intracellular and extracellular processes, as well as in the regulation of physiological pathways, including the degradation of misfolded proteins, processing short-lived signaling proteins, and signal peptide cleavage (52). Proteases also represent one of the

most heavily utilized enzymes in various industrial processes and detergents (53-56). Most of the proteases known to date fall into four major groups by nature of their active center; the serine proteases, cysteine proteases, aspartic proteases, and the metallo-proteases. A limited number of proteases that utilize other residues such as threonine are also known.

The standard mechanism for serine proteases involves a catalytic triad (Fig. 6). The catalytic triad consists of a histidine general base, which abstracts the proton from a serine to act as a nucleophile and attack the carbonyl group of an amide bond within the protein substrate. The third player in the triad, an acidic residue, acts to orient the histidine residue and neutralize the charged histidine intermediate (52). A number of serine proteases have recently been identified that contain an essential lysine, but no essential histidines. It has been proposed that this group of proteases uses a Ser-Lys dyad mechanism, whereby an ε-amino group of a lysine side chain acts as the general base to increase the nucleo-philicity of the active site serine (52). According to the MEROPS database (The peptidase database at http://merops.sanger.ac.uk/index.htm) (57), there are now 42 different families of serine proteases including S1 (chymotrypsin), S8 (subtilisin) and S14 (Clp peptidase).

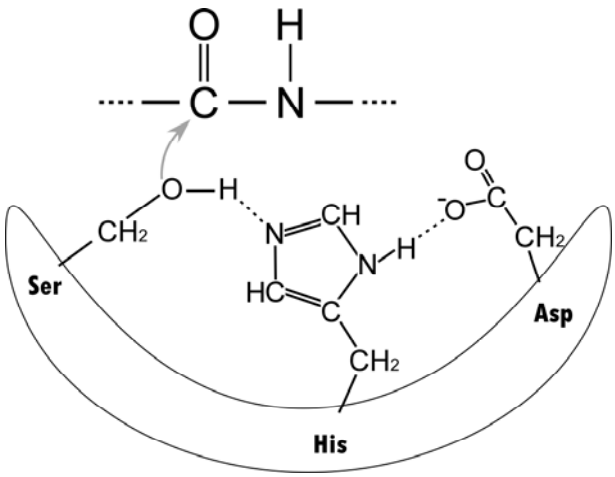


Fig. 6. A diagram illustrating the catalytic triad of serine proteases.

5. *Thermococcus kodakaraensis* KOD1

Thermococcus kodakaraensis KOD1 is a hyperthermophilic archaeon isolated from a solfatara on Kodakara Island, Kagoshima, Japan (58, 59) (Fig. 7A). The strain is an obligate anaerobe and grows optimally at 85˚C. Only heterotrophic growth has been observed, and the strain can efficiently utilize and/or degrade amino acids, peptides, pyruvate, and a number of polymers such as chitin, pullulan and starch. The author contributed in determining the complete genome sequence of *T. kodakaraensis* (60). This is the only *Thermococcus* genome sequence available at present, and together with the *Pyrococcus* genome sequences, has provided valuable knowledge towards understanding the various cellular activities of sulfur-reducing heterotrophic archaea. The genome consists of 2,088,737 bp and harbors 2,306 predicted open reading frames, among which half were annotatable (Fig. 7B). As expected from the growth characteristics of *T. kodakaraensis*, the genome sequence revealed the presence of a large number of putative extracellular enzymes, including chitinase (61), α-amylase (62), amylopullulanase, and subtilisin-like protease (63).

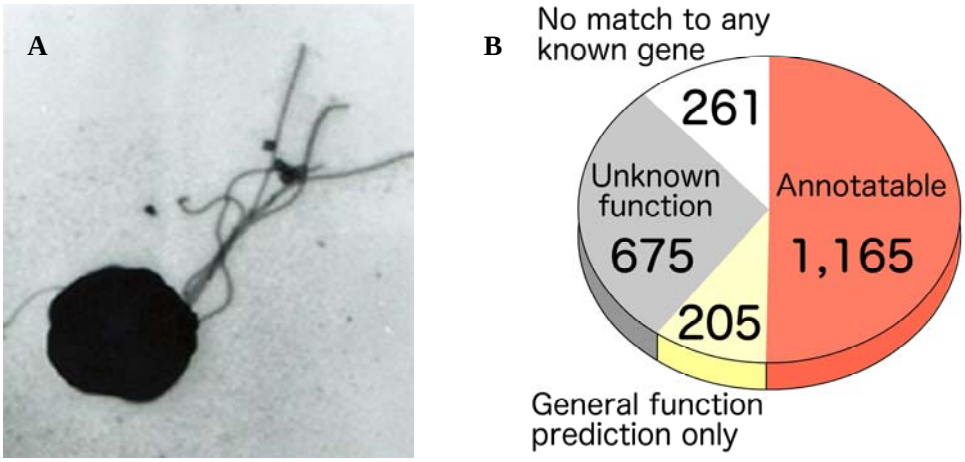


Fig. 7. A; Electron micrograph of *T. kodakaraensis*. B; Annotation results of the *T. kodakaraensis* genome sequence.

6. Gene disruption systems in hyperthermophiles

In spite of the many attractive aspects of hyperthermophilic archaea, progress in the research on these organisms has been constantly hampered by the limitation of tools available for genetic manipulation. In particular, the development of gene disruption technology, which is one of the most straightforward methods in examining gene function *in vivo*, has been limited to only two organisms among the hyperthermophilic archaea. One was developed in *T. kodakaraensis* from the *Euryarchaeota* (64, 65) and the other was constructed in *Sulfolobus solfataricus* from the *Crenarchaeota* (66). Both systems rely on homologous recombination. The former system utilizes various host strains with amino acid/nucleotide auxotrophy and corresponding marker genes that complement the auxotrophy (Fig. 8). The latter utilizes a *lacS*-deficient host strain and a modified but active *lacS* marker gene with selection based on lactose-dependent growth. The two systems have proved to be powerful tools in examining gene function in the respective organisms (12, 67-69). Application of these methodologies or developing new gene disruption technologies towards other hyperthermophilic archaea will be important to broaden our understanding on these unique organisms.

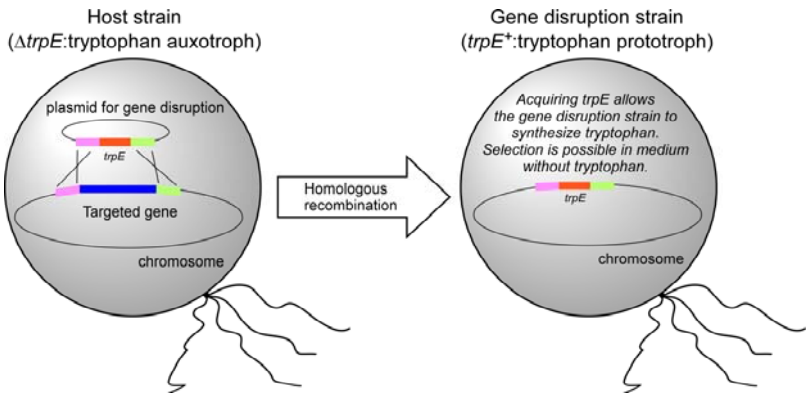


Fig. 8. A model illustrating gene disruption in *T. kodakaraensis* via double crossover recombination.

7. Objectives of the study

Part I of this study focuses on the identification and biochemical characterization of signal peptide peptidase in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. As described above, among the various components of the archaeal secretion machinery, the signal peptide peptidases had not been identified nor examined in the *Archaea*. In addition, the two putative peptidases examined in this study were representatives of serine proteases whose catalytic mechanisms had not been elucidated. The characterization of these peptidases was expected to provide valuable information on the degradation mechanism of signal peptides in the *Archaea*, and also clarify the catalytic mechanism of two types of serine proteases. Depending on their biochemical properties, the peptidases were also recognized as potential thermostable biocatalysts for application in protein cleavage/degradation. In Part II, the author set out to develop a gene disruption system in hyperthermophilic archaea based on antibiotic resistance. This type of system had not been developed in hyperthermophiles, and was expected to provide a general method for gene disruption in these organisms that did not require the construction of a specific host cell deficient in a particular cellular function. The genes chosen for disruption were those presumed to encode the components of a sugar transporter on the membrane of *T. kodakaraensis*.

REFERENCES

1. **Woese, C. R. & Fox, G. E.**, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. USA*, 74, 5088-5090, 1977.
2. **Woese, C. R., Magrum, L. J. & Fox, G. E.**, Archaebacteria. *J. Mol. Evol.*, 11, 245-251, 1978.

3. **Stetter, K. O.**, Hyperthermophilic procaryotes. *FEMS Microbiol. Rev.*, 18, 149-158, 1996.

4. **Stetter, K. O.**, Ultrathin mycelia-forming organisms from submarine volcanic areas having an optimum growth temperature of 105˚C. *Nature*, 300, 258-260, 1982.

5. **Stetter, K. O., Köning, H. & Stackebrandt, E.**, *Pyrodictium* gen. nov., a new genus of submarine disc-shaped sulphur reducing archaebacteria growing optimally at 105˚C. *System. Appl. Microbiol.*, 4, 535-551, 1983.

6. **Stetter, K. O.**, Extremophiles and their adaptation to hot environments. *FEBS Lett.*, 452, 22-25, 1999.

7. **Di Giulio, M.**, The universal ancestor was a thermophile or a hyperthermophile: tests and further evidence. *J. Theor. Biol.*, 221, 425-436, 2003.

8. **Kikuchi, A. & Asai, K.**, Reverse gyrase-a topoisomerase which introduces positive superhelical turns into DNA. *Nature*, 309, 677-681, 1984.

9. **Forterre, P.**, A hot story from comparative genomics: reverse gyrase is the only hyperthermophile-specific protein. *Trends Genet.*, 18, 236-237, 2002.

10. **Confalonieri, F., Elie, C., Nadal, M., de La Tour, C. B., Forterre, P. & Duguet, M.**, Reverse gyrase: a helicase-like domain and a type I topoisomerase in the same polypeptide. *Proc. Natl. Acad. Sci. USA*, 90, 4753-4757, 1993.

11. **Forterre, P.**, A hot topic: the origin of hyperthermophiles. *Cell*, 85, 789-792, 1996.

12. **Atomi, H., Matsumi, R. & Imanaka, T.**, Reverse gyrase is not a prerequisite for hyperthermophilic life. *J. Bacteriol.*, 186, 4829-4833, 2004.

13. **Hanford, M. J. & Peeples, T. L.**, Archaeal tetraether lipids: unique structures

and applications. *Appl. Biochem. Biotechnol.*, 97, 45-62, 2002.

14. **Bell, S. D.**, Archaeal transcriptional regulation-variation on a bacterial theme? *Trends Microbiol.*, 13, 262-265, 2005.

15. **Geiduschek, E. P. & Ouhammouch, M.**, Archaeal transcription and its regulators. *Mol. Microbiol.*, 56, 1397-1407, 2005.

16. **Reeve, J. N.**, Archaeal chromatin and transcription. *Mol. Microbiol.*, 48, 587-798, 2003.

17. **Barry, E. R. & Bell, S. D.**, DNA replication in the archaea. *Microbiol. Mol. Biol. Rev.*, 70, 876-887, 2006.

18. **Kelman, Z. & White, M. F.**, Archaeal DNA replication and repair. *Curr. Opin. Microbiol.*, 8, 669-676, 2005.

19. **Adams, M. W. W. & Kelly, R. M.**, Finding and using hyperthermophilic enzymes. *Trends Biotechnol.*, 16, 329-332, 1998.

20. **Imanaka, T. & Atomi, H.**, Catalyzing "hot" reactions: enzymes from hyperthermophilic Archaea. *Chem. Rec.*, 2, 149-163, 2002.

21. **Corsi, A. K. & Schekman, R.**, Mechanism of polypeptide translocation into the endoplasmic reticulum. *J. Biol. Chem.*, 271, 30299-30302, 1996.

22. **Economou, A.**, Following the leader: bacterial protein export through the Sec pathway. *Trends Microbiol.*, 7, 315-320, 1999.

23. **Izard, J. W. & Kendall, D. A.**, Signal peptides: exquisitely designed transport promoters. *Mol. Microbiol.*, 13, 765-773, 1994.

24. **de Keyzer, J., van der Does, C. & Driessen, A. J. M.**, The bacterial translocase: a dynamic protein channel complex. *Cell. Mol. Life Sci.*, 60, 2034-2052, 2003.

25. **Oliver, D. B.**, SecA protein: autoregulated ATPase catalysing preprotein insertion and translocation across the *Escherichia coli* inner membrane. *Mol. Microbiol.*, 7, 159-165, 1993.

26. **Nagai, K., Oubridge, C., Kuglstatter, A., Menichelli, E., Isel, C. & Jovine, L.**, Structure, function and evolution of the signal recognition particle. *EMBO J.*, 22, 3479-3485, 2003.

27. **Pohlschröder, M., Giménez, M. I. & Jarrell, K. F.**, Protein transport in Archaea: Sec and twin arginine translocation pathways. *Curr. Opin. Microbiol.*, 8, 713-719, 2005.

28. **Albers, S.-V., Szabo, Z. & Driessen, A. J. M.**, Protein secretion in the Archaea: multiple paths towards a unique cell surface. *Nat. Rev. Microbiol.*, 4, 537-547, 2006.

29. **Pohlschröder, M., Dilks, K., Hand, N. J. & Wesley Rose, R.**, Translocation of proteins across archaeal cytoplasmic membranes. *FEMS Microbiol. Rev.*, 28, 3-24, 2004.

30. **Ring, G. & Eichler, J.**, Extreme secretion: protein translocation across the archaeal plasma membrane. *J. Bioenerg. Biomembr.*, 36, 35-45, 2004.

31. **Pugsley, A. P., Francetic, O., Driessen, A. J. M. & de Lorenzo, V.**, Getting out: protein traffic in prokaryotes. *Mol. Microbiol.*, 52, 3-11, 2004.

32. **Robson, A. & Collinson, I.**, The structure of the Sec complex and the problem of protein translocation. *EMBO Rep.*, 7, 1099-1103, 2006.

33. **van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A.**, X-ray structure of a protein-conducting channel. *Nature*, 427, 36-44, 2004.

34. **Bolhuis, A.**, The archaeal Sec-dependent protein translocation pathway. *Phil. Trans. R. Soc. Lond. B Biol. Sci.*, 359, 919-927, 2004.

35. **Bhuiyan, S. H., Gowda, K., Hotokezaka, H. & Zwieb, C.**, Assembly of archaeal signal recognition particle from recombinant components. *Nucleic Acids Res.*, 28, 1365-1373, 2000.

36. **Rosendal, K. R., Wild, K., Montoya, G. & Sinning, I.**, Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc. Natl. Acad. Sci. USA*, 100, 14701-14706, 2003.

37. **Paetzel, M., Karla, A., Strynadka, N. C. J. & Dalbey, R. E.**, Signal peptidases. *Chem. Rev.*, 102, 4549-4579, 2002.

38. **Hussain, M., Ichihara, S. & Mizushima, S.**, Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the *Escherichia coli* outer membrane. *J. Biol. Chem.*, 257, 5177-5182, 1982.

39. **Ichihara, S., Beppu, N. & Mizushima, S.**, Protease IV, a cytoplasmic membrane protein of *Escherichia coli*, has signal peptide peptidase activity. *J. Biol. Chem.*, 259, 9853-9857, 1984.

40. **Ichihara, S., Suzuki, T., Suzuki, M. & Mizushima, S.**, Molecular cloning and sequencing of the *sppA* gene and characterization of the encoded protease IV, a signal peptide peptidase, of *Escherichia coli*. *J. Biol. Chem.*, 261, 9405-9411, 1986.

41. **Pacaud, M.**, Purification and characterization of two novel proteolytic enzymes in membranes of *Escherichia coli*. Protease IV and protease V. *J. Biol. Chem.*, 257, 4333-4339, 1982.

42. **Novak, P., Ray, P. H. & Dev, I. K.**, Localization and purification of two enzymes from *Escherichia coli* capable of hydrolyzing a signal peptide. *J. Biol. Chem.*, 261, 420-427, 1986.

43. **Suzuki, T., Itoh, A., Ichihara, S. & Mizushima, S.**, Characterization of the *sppA* gene coding for protease IV, a signal peptide peptidase of *Escherichia coli*. *J. Bacteriol.*, 169, 2523-2528, 1987.

44. **Novak, P. & Dev, I. K.**, Degradation of a signal peptide by protease IV and oligopeptidase A. *J. Bacteriol.*, 170, 5067-5075, 1988.

45. **Braud, V. M., Allan, D. S. J., O'Callaghan, C. A., Söderström, K., D'Andrea, A., Ogg, G. S., Lazetic, S., Young, N. T., Bell, J. I., Phillips, J. H., Lanier, L. L. & McMichael, A. J.**, HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature*, 391, 795-799, 1998.

46. **Lemberg, M. K., Bland, F. A., Weihofen, A., Braud, V. M. & Martoglio, B.**, Intramembrane proteolysis of signal peptides: an essential step in the generation of HLA-E epitopes. *J. Immunol.*, 167, 6441-6446, 2001.

47. **Casso, D. J., Tanda, S., Biehs, B., Martoglio, B. & Kornberg, T. B.**, Drosophila signal peptide peptidase is an essential protease for larval development. *Genetics*, 170, 139-148, 2005.

48. **Bardy, S. L., Eichler, J. & Jarrell, K. F.**, Archaeal signal peptides-a comparative survey at the genome level. *Protein Sci.*, 12, 1833-1843, 2003.

49. **Ng, S. Y. M., Chaban, B., VanDyke, D. J. & Jarrell, K. F.**, Archaeal signal peptidases. *Microbiology*, 153, 305-314, 2007.

50. **Bardy, S. L., Ng, S. Y. M., Carnegie, D. S. & Jarrell, K. F.**, Site-directed mutagenesis analysis of amino acids critical for activity of the type I signal

peptidase of the archaeon *Methanococcus voltae*. *J. Bacteriol.*, 187, 1188-1191, 2005.

51. **Ng, S. Y. M. & Jarrell, K. F.**, Cloning and characterization of archaeal type I signal peptidase from *Methanococcus voltae*. *J. Bacteriol.*, 185, 5936-5942, 2003.

52. **Paetzel, M. & Dalbey, R. E.**, Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.*, 22, 28-31, 1997.

53. **Atomi, H.**, Recent progress towards the application of hyperthermophiles and their enzymes. *Curr. Opin. Chem. Biol.*, 9, 166-173, 2005.

54. **Gupta, R., Beg, Q. K. & Lorenz, P.**, Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.*, 59, 15-32, 2002.

55. **Ito, S., Kobayashi, T., Ara, K., Ozaki, K., Kawai, S. & Hatada, Y.**, Alkaline detergent enzymes from alkaliphiles: enzymatic properties, genetics, and structures. *Extremophiles*, 2, 185-190, 1998.

56. **Maurer, K.-H.**, Detergent proteases. *Curr. Opin. Biotechnol.*, 15, 330-334, 2004.

57. **Rawlings, N. D., Tolle, D. P. & Barrett, A. J.**, MEROPS: the peptidase database. *Nucleic Acids Res.*, 32 Database issue, D160-D164, 2004.

58. **Atomi, H., Fukui, T., Kanai, T., Morikawa, M. & Imanaka, T.**, Description of *Thermococcus kodakaraensis* sp. nov., a well studied hyperthermophilic archaeon previously reported as *Pyrococcus* sp. KOD1 *Archaea*, 1, 263-267, 2004.

59. **Morikawa, M., Izawa, Y., Rashid, N., Hoaki, T. & Imanaka, T.**, Purification

and characterization of a thermostable thiol protease from a newly isolated hyperthermophilic *Pyrococcus* sp. *Appl. Environ. Microbiol.*, 60, 4559-4566, 1994.

60. **Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T.**, Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.*, 15, 352-363, 2005.

61. **Tanaka, T., Fujiwara, S., Nishikori, S., Fukui, T., Takagi, M. & Imanaka, T.**, A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *Appl. Environ. Microbiol.*, 65, 5338-5344, 1999.

62. **Tachibana, Y., Leclere, M. M., Fujiwara, S., Takagi, M. & Imanaka, T.**, Cloning and expression of the α-amylase gene from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and characterization of the enzyme. *J. Ferment. Bioeng.*, 82, 224-232, 1996.

63. **Kannan, Y., Koga, Y., Inoue, Y., Haruki, M., Takagi, M., Imanaka, T., Morikawa, M. & Kanaya, S.**, Active subtilisin-like protease from a hyperthermophilic archaeon in a form with a putative prosequence. *Appl. Environ. Microbiol.*, 67, 2445-2452, 2001.

64. **Sato, T., Fukui, T., Atomi, H. & Imanaka, T.**, Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.*, 185, 210-220, 2003.

65. **Sato, T., Fukui, T., Atomi, H. & Imanaka, T.**, Improved and versatile transformation system allowing multiple genetic manipulations of the

hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.*, 71, 3889-3899, 2005.

66. **Worthington, P., Hoang, V., Perez-Pomares, F. & Blum, P.**, Targeted disruption of the α-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.*, 185, 482-488, 2003.

67. **Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H. & Imanaka, T.**, Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.*, 61, 898-909, 2006.

68. **Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H. & Imanaka, T.**, Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.*, 186, 5799-5807, 2004.

69. **Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M. & Blum, P.**, Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J. Bacteriol.*, 186, 427-437, 2004.

SYNOPSIS

In this study, the author has performed a biochemical examination on two membrane-anchored peptidases in the hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1 (Part I, Chapters 1-4). In Part II (Chapter 5), a method for gene disruption in this archaeon based on antibiotic resistance was developed and applied for genetic analysis of a putative sugar transporter of *T. kodakaraensis*.

In Chapter 1, the author performed a biochemical characterization of a putative signal peptide peptidase of *T. kodakaraensis*. Signal peptide peptidases had not been identified in the *Archaea*, and genes encoding proteins with particularly high similarity to bacterial or eukaryotic enzymes were not present on the archaeal genomes. The author took notice of a gene encoding a protein ($SppA_{Tk}$) of 334 amino acid residues that, although much smaller than its bacterial counterpart in *Escherichia coli* (618 residues), displayed 27% identity in primary structure. A predicted membrane-spanning domain was present in its amino-terminus, suggesting a membrane localization of the protein. Biochemical characterization of the catalytic domain of $SppA_{Tk}$ revealed that the protein indeed exhibits peptidase activity, and recognizes peptide sequences with hydrophobic/aromatic residues at the P-3 position and residues with relatively small side chains at the P-1 position. Sequences containing acidic residues were not cleaved by $SppA_{Tk}$. Taking into account that signal sequences in the *Archaea* generally contain a hydrophobic stretch of amino acid residues and do not contain acidic amino acid residues, the substrate preference of $SppA_{Tk}$ strongly suggests that the enzyme represents the signal peptide peptidase in the *Archaea*.

In Chapter 2, the author further proceeded in elucidating the amino acid

residues participating in the catalysis of $SppA_{Tk}$. $SppA_{Tk}$ belongs to a group of putative serine peptidases/proteases constituting the S49 family. The catalytic center had not been elucidated in any of the members of the S49 family. By individually replacing amino acid residues highly conserved in $SppA_{Tk}$ homologs to alanine, 16 mutant proteins were purified and analyzed for peptidase activity. As a result, $SppA_{Tk}$ was found to utilize Ser162 as the nucleophilic serine and Lys214 as the general base, comprising a Ser-Lys catalytic dyad for peptide bond hydrolysis. Intriguingly, several mutant proteins exhibited higher levels of activity than the wild-type $SppA_{Tk}$. As these proteins also displayed a broadening in substrate specificity, these residues may be present to prevent the enzyme from cleaving unintended peptide/protein substrates in the cell.

In Chapter 3, the author searched for other peptidases that may be involved in the degradation of signal peptides in *T. kodakaraensis*. Although homologs of oligopeptidase A and TepA, two peptidases involved in this process in bacteria, were not present, the author found a second gene on the *T. kodakaraensis* genome that encoded a putative membrane-bound peptidase ($SppB_{Tk}$) that was 18% identical to $SppA_{Tk}$. As in the case of $SppA_{Tk}$, $SppB_{Tk}$ harbored a predicted membrane-spanning domain in its amino-terminus. The association of the protein to the cytoplasmic membrane was confirmed through Western blot analysis. A genome database search revealed that $SppB_{Tk}$ homologs were distributed in a number of archaea and bacteria. In order to examine the possibilities of whether this protein was also involved in signal peptide degradation, the recombinant catalytic domain was biochemically examined. As in the case of $SppA_{Tk}$, $SppB_{Tk}$ showed no activity towards peptide sequences with acidic residues. However, in contrast to $SppA_{Tk}$, $SppB_{Tk}$ exhibited a strong preference for

basic amino acid residues at the P-2 position and hydrophobic residues at the P-1 site. This substrate specificity raises the possibilities that $SppB_{Tk}$ may function in cooperation with $SppA_{Tk}$ in the degradation of signal peptides, with $SppB_{Tk}$ mainly responsible for the cleavage of sequences located immediately upstream of the hydrophobic stretch in signal peptides, which in many cases harbor basic amino acid residues.

In Chapter 4, the author determined the amino acid residues that are important for the peptidase activity of $SppB_{Tk}$. No information had been available on the catalytic mechanism of $SppB_{Tk}$ homologs. A sequence comparison of over 40 archaeal and bacterial $SppB_{Tk}$ homologs was performed, and 18 conserved amino acid residues were selected as candidates for components of the catalytic center. A detailed site-directed mutagenesis study was performed by constructing and analyzing mutant proteins in which the selected residues were replaced by alanine. The study revealed that substitution of Ser130, His226 or Asp154 led to the abolishment or dramatic decrease in $SppB_{Tk}$ activity. In contrast to the catalytic center of $SppA_{Tk}$, the results suggested that $SppB_{Tk}$ relies on a Ser-His-Asp catalytic triad for proteolytic activity.

In Part II (Chapter 5), the author developed a method for gene disruption in *T. kodakaraensis* KOD1 based on antibiotic resistance. Conventional antibiotics and antibiotic resistance marker genes cannot be used in hyperthermophiles due to their lack of thermostability. Therefore a strategy based on inhibition of a particular endogenous protein by an antibiotic and relieving the inhibition by overexpressing the protein was applied. The author utilized simvastatin, a specific inhibitor of the enzyme 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, which is essential for archaeal membrane synthesis. Using an overexpression cassette of a thermostable

HMG-CoA reductase gene as a marker, the author was able to efficiently select gene disruption strains of a putative sugar transporter complex gene cluster of *T. kodakaraensis*. Phenotypic examinations clearly revealed that the transporter was the only transporter involved in maltooligosaccharide uptake in this strain. The gene disruption system developed in this chapter can be applied in nutrient-rich media, and should be helpful in developing gene disruption systems in other hyperthermophilic archaea as there is no need for the initial development of auxotrophic host strains. In addition, the methodology will surely be a powerful tool in future genetic studies on membrane transporters, as isolation of disruption mutants of microbial transporters/channels can be expected to be in some cases difficult in minimal media.

PART I

Studies on membrane-bound peptidases from *Thermococcus kodakaraensis*

CHAPTER 1

Biochemical properties of a putative signal peptide peptidase from *Thermococcus kodakaraensis*

INTRODUCTION

Compared to the wealth of studies reported on bacterial and eukaryotic protein secretion, the number of experimental examinations on archaeal secretion is still very low (1). A comparative genomics approach has been taken to examine the presence or absence of archaeal homologs structurally related to genes that function in the eukaryotic or bacterial secretion systems (2). Comparing the primary sequences of putative secretion proteins has also revealed the structural characteristics of the archaeal signal peptide (GENERAL INTRODUCTION, Fig. 4). Among the factors that comprise the archaeal secretion machinery, proteins that have been experimentally examined include those involved in flagellum formation in methanogens (3), the signal recognition particle of *Archaeoglobus fulgidus* (4) and *Sulfolobus solfataricus* (5), and the Sec protein-conducting channel complex of *Methanocaldococcus jannaschii* (6).

Signal peptide peptidases (SPPs) are enzymes considered to cleave the signal peptide chains of secreted proteins after they are removed from the precursor proteins by signal peptidases (7, 8). Eukaryotic SPPs are integral membrane enzymes, with activity dependent on two aspartate residues (9, 10). They have become a center of attention in mammalian cells due to their involvement in immune surveillance. After SPP cleaves signal peptides of the major histocompatibility complex I molecules, the peptide products are presented on the cell surface by a nonclassical major

histocompatibility complex class I molecule, HLA-E, indicating to natural killer cells that major histocompatibility complex synthesis is proceeding normally (11, 12).

The bacterial SPP was initially identified in *Escherichia coli* as a cytoplasmic membrane protein named protease IV (7, 13, 14). The enzyme, encoded by the *sppA* gene (15, 16), was found to cleave the signal peptide of outer membrane lipoprotein after its release from the precursor protein. Further studies have indicated that protease IV (SppA) carries out only the initial breakdown of the signal peptide into smaller peptide fragments, followed by complete digestion through the functions of cytoplasmic peptidases including oligopeptidase A (17, 18). The gram-positive counterpart of SppA in *Bacillus subtilis* has also been studied, and has been shown to be involved in signal peptide degradation (19). Furthermore, a cytosolic peptidase, TepA, structurally related to both SppA and ClpP has also been found to actively participate in the degradation of signal peptides in this organism (19).

In terms of signal peptidases (SPs) and SPPs from the *Archaea*, the type I SP gene from *Methanococcus voltae* has been cloned and its product characterized, confirming that the protein exhibits SP activity (20). Residues critical for the peptidase activity of the protein have been determined (21). FlaK, the SP that functions exclusively for preflagellin signal cleavage, has also been characterized from this organism and has been demonstrated to be an aspartic protease essential for preflagellin cleavage (22). In the *Crenarchaeota*, the homolog of bacterial type IV prepilin peptidases from *S. solfataricus* (PibD) has been characterized, and residues on the substrate that are important for recognition by PibD have been examined (23). In contrast to the progress on SPs, the presence or examination of SPPs have not been reported in the *Archaea*.

In this chapter, the author has identified and examined the enzymatic properties of a putative SPP from *T. kodakaraensis*, revealing that the substrate specificity of the enzyme is consistent with its presumed role as an SPP in this archaeon.

MATERIALS AND METHODS

Strains, media, and plasmids

The standard growth medium for the cultivation of *T. kodakaraensis* was a nutrient-rich ASW-YT medium supplemented with 2.0 g liter^{-1} elemental sulfur. ASW-YT medium is composed of 0.8 x artificial seawater, 5.0 g liter^{-1} yeast extract, and 5.0 g liter^{-1} tryptone. The composition of 0.8 x artificial seawater is (per liter) 20 g NaCl, 3.0 g $MgCl_2 \cdot 6H_2O$, 6.0 g $MgSO_4 \cdot 7H_2O$, 1.0 g $(NH_4)_2SO_4$, 0.2 g $NaHCO_3$, 0.3 g $CaCl_2 \cdot 2H_2O$, 0.5 g KCl, 0.42 g KH_2PO_4, 0.05 g NaBr, 0.02 g $SrCl_2 \cdot 6H_2O$, and 0.01 g $Fe(NH_4)$citrate. Resazurin was added to the medium at a concentration of 0.8 mg liter^{-1} as an oxygen indicator. Prior to cell inoculation, Na_2S was added to the medium until the color of resazurin became transparent. *T. kodakaraensis* cells were grown under anaerobic conditions at 85˚C. *E. coli* DH5α and plasmid pUC18 were used for gene cloning, sequencing, and DNA manipulation. *E. coli* BL21-CodonPlus(DE3)-RIL (Stratagene, La Jolla, CA, USA) and pET21a(+) (Novagen, Madison, WI, USA) were used for gene expression. *E. coli* strains were cultivated in LB medium (10 g of tryptone, 5.0 g of yeast extract, and 10 g of NaCl per liter) with 100 μg ml^{-1} ampicillin at 37˚C.

DNA manipulation and sequence analysis

Restriction and modification enzymes were purchased from Toyobo (Osaka, Japan) or Takara (Kyoto, Japan). KOD Plus (Toyobo) was used as a polymerase for

Polymerase Chain Reaction (PCR). Plasmid DNA was isolated and purified from *E. coli* cells with the Plasmid mini-kit from QIAGEN (Hilden, Germany). GFX PCR DNA and gel band purification kit (GE Healthcare, Little Chalfont, UK) was used to recover DNA fragments from agarose gels after electrophoresis. DNA sequencing was performed using BigDye terminator cycle sequencing kit v.3.0-3.1 and a model 3100 capillary DNA sequencer (Applied Biosystems, Foster City, CA, USA). Sequence alignments and construction of the phylogenetic tree with the neighbor-joining method were performed with the ClustalW program available at the DNA Data Bank of Japan. Bootstrap resampling was performed 1,000 times with the BSTRAP program.

Expression of the *sppA$_{Tk}$* gene in *E. coli*

Genomic DNA from *T. kodakaraensis* was isolated by methods described elsewhere (24). The *sppA$_{Tk}$* gene initiating with a Met residue preceding Gln30, omitting the transmembrane domain, was amplified from the genomic DNA of *T. kodakaraensis* using the primer set sppN1 and sppC1 (sppN1, 5'-GTTC TCCATATGCAGGTCAATCCCCCCGCTGT-3'; sppC1, 5'-CAGAATTCAACCACC CCCAATGAGGG-3'). The gene for the truncated protein initiating with a Met residue preceding Cys55 was amplified with sppN2 and sppC1 (sppN2, 5'-ACTTTACGCAT ATGTGTGAAGGCAGTGTTAAC-3'). After confirming the sequences of the DNA fragments, they were inserted into pET21a(+) at the NdeI and EcoRI sites. After introduction into *E. coli* BL21-CodonPlus(DE3)-RIL cells, gene expression was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside at the mid-exponential growth phase with further incubation for 6 h at 37˚C.

Purification of recombinant $SppA_{Tk}$

After inducing gene expression, cells were washed with 50 mM Tris-HCl (pH 8.0) and resuspended in the same buffer. Cells were sonicated on ice, and the supernatant after centrifugation (20,000 x *g*, 30 min at 4˚C) was applied to heat treatment at 85˚C for 15 min, immediately cooled on ice, and then centrifuged (20,000 x *g*, 30 min at 4˚C). The soluble protein sample was brought to 35% saturation with $(NH_4)_2SO_4$ and the precipitate which included $SppA_{Tk}$ was dissolved in 50 mM Tris-HCl (pH 8.0). This was applied to anion exchange chromatography (ResourceQ, GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 0.2 M NaCl, and proteins were eluted with a linear gradient (0.2 to 1.0 M) of NaCl. After desalting with a HiPrep26/10 column (GE Healthcare), the sample was applied to gel filtration chromatography (Superdex 200 HR 10/30, GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 0.15 M NaCl, and the fractions obtained were used for enzyme analysis.

Protein analysis of purified recombinant $SppA_{Tk}$

The native molecular mass of the purified protein was examined by gel-filtration chromatography using Superdex 200 HR 10/30 in 50 mM Tris-HCl (pH 8.0), 0.15 M NaCl. The retention time was calibrated with those of the standard proteins thyroglobulin (669 kDa), ferritin (440 kDa), catalase (232 kDa), aldolase (158 kDa), albumin (67 kDa), ovalbumin (43 kDa), chymotrypsinogen A (25 kDa), and ribonuclease A (13.7 kDa). Protein concentration was determined with the protein assay kit (Bio-Rad, Hercules, CA, USA) using bovine serum albumin as a standard. Determination of amino-terminal amino acid sequences of proteins was performed with

a protein sequencer (model 491 cLC, Applied Biosystems) after separation by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and electroblotting onto a polyvinylidene difluoride membrane (Millipore, Bedford, MA, USA).

Enzyme activity measurements

Most activity measurements were performed with peptidyl-MCA substrates [peptidyl–α-(4-methylcoumaryl-7-amide) substrates] available from Peptide Institute (Osaka, Japan). Release of 7-amino-4-methylcoumarin was monitored consecutively with a fluorescence spectrophotometer capable of maintaining the cuvette at desired temperatures between 30 and 100˚C. Excitation and emission wavelengths were 380 nm and 460 nm, respectively. Standard activity measurements were performed at 60˚C in a final volume of 1 ml with 0.1 μg of purified protein and Ala-Ala-Phe-MCA (200 μM) in 50 mM CHES (*N*-cyclohexyl-2-aminoethanesulfonic acid; pH 10.0). The final concentration of dimethyl sulfoxide used to dissolve the substrate was constant at 3% of the reaction mixture.

Effects of temperature and pH on enzyme activity

All buffers were prepared so that they would reflect accurate values at the applied temperatures. In examining the effect of temperature, the standard assay method was applied at each temperature. The effect of pH was examined in the presence of 50 mM of 2-morpholinoethanesulfonic acid (MES)-NaOH (pH 6.0 to 7.0), 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES)-NaOH (pH 7.0 to 8.0), *N*, *N*-bis(2-hydroxyethyl)glycine (Bicine)-NaOH (pH 8.0 to 9.0), CHES-NaOH (pH 9.0 to 10.0), and *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS)-NaOH (pH

10.0 to 12.0), respectively. Thermostability of the protein was analyzed by measuring the residual activity of the protein after incubation at various temperatures in 50 mM CHES-NaOH (pH 10.0). The initial activity of the enzyme incubated at 60˚C was designated as 100%. Alkaline stability was analyzed by measuring the residual activity of the protein after incubation at various pH in 50 mM CAPS-NaOH at 60˚C. The initial activity of the enzyme incubated in 50 mM CAPS-NaOH (pH 10.0) was designated as 100%. In measuring thermostability and alkaline stability, the protein concentration during incubation was 1 μg ml^{-1}. Residual activities were measured with the standard assay method described above.

Effect of protease inhibitors

The effects of various protease inhibitors at concentrations of 200 μM, 1 mM, or 10 mM were examined at 60˚C and pH 10.0. The substrate Ala-Ala-Phe-MCA was present at a concentration of 200 μM. Activity in the absence of inhibitors was defined as 100%.

Determination of substrate specificity

Substrate specificity was examined with a FRETS peptide library (25Xaa series, Peptide Institute) (25). These peptide substrates harbor a highly fluorescent 2-(*N*-methylamino)benzoyl group linked to the side chain of the amino-terminal D-2,3-diaminopropionic acid residue (D-A_2pr), along with a 2,4-dinitrophenyl group (quencher) linked to the ε-amino group of a Lys residue. In between the D-A_2pr and Lys residue lies the peptide Gly-Zaa-Yaa-Xaa-Ala-Phe-Pro, where Zaa is a mixture of Phe, Ala, Val, Glu, and Arg, Yaa is a mixture of Pro, Tyr, Lys, Ile, and Asp, and the Xaa

residue is a defined single amino acid of choice (see RESULTS section). Excitation and emission wavelengths were 340 nm and 440 nm, respectively. In the initial assay to examine the preference for residues at the P-1 position, 1 μg of purified enzyme was added to the reaction mixture with a final volume of 1 ml containing 30 μM substrate in 50 mM CHES (pH 10.0). The final concentration of dimethyl sulfoxide used to dissolve the substrate was constant at 3% of the reaction mixture. A second assay to identify the cleavage sites and the preference towards residues at the P-2 to P-4 positions was performed on selected substrates. Aliquots (100 μl) from the cleavage reactions were taken at various time intervals that corresponded to 15 to 30% cleavage of the substrates, and subjected to liquid chromatography (LC)-mass spectrometry analysis. An ODS A-302 column (YMC, Kyoto, Japan) was used for separation with 0.05% trifluoroacetic acid in H_2O as eluant A and 0.05% trifluoroacetic acid in CH_3CN as eluant B. The gradient was 5–40% of eluant B in A at a flow rate of 1.0 ml min^{-1} over a time span of 55 min. Aliquots taken from the cleavage reactions were injected and the cleaved products were monitored with absorbance at 220 nm, as well as fluorescence intensity in order to identify the amino-terminal segments. The structures of the cleaved products were deduced from the theoretical molecular weights.

RESULTS

Putative signal peptide peptidase gene on the genome of *T. kodakaraensis*

Using the primary structures of the signal peptide peptidases from *E. coli* ($SppA_{Ec}$) and human, the author performed a BLAST search against the protein sequences of *T. kodakaraensis*. Closely related homologs could not be identified for either the bacterial or eukaryotic SPP. However, although significantly smaller in size

than $SppA_{Ec}$ (618 amino acid residues), one open reading frame (TK1164) encoded a protein (334 residues) with 27% identity to $SppA_{Ec}$. The author designated the gene as *sppA*$_{Tk}$ (Fig. 1). The deduced molecular mass of the protein was 36,211 Da. A BLAST search against the complete genome sequences of various archaeal strains was performed with the $SppA_{Ec}$ and $SppA_{Tk}$ sequences. Similar homologs were found in many genera of the *Euryarchaeota*, including *Picrophilus*, *Pyrococcus*, and *Thermoplasma*, as well as the methanogens and the haloarchaea. The author also found homologs in *Nanoarchaeum* and the crenarchaeon *Pyrobaculum*. A phylogenetic analysis of these sequences along with several selected bacterial sequences is shown in Fig. 2. Although the catalytic residues have not been experimentally verified, effects of various inhibitors have suggested that $SppA_{Ec}$ is a serine protease (13). The author indeed observed multiple serine residues that were highly conserved among the archaeal and bacterial SppA sequences (Fig. 1). As in the case of $SppA_{Ec}$, $SppA_{Tk}$ is structurally categorized in the S49 family of the SK clan of serine proteases (MEROPS, the peptidase database, http://merops.sanger.ac.uk/) (26). Another common feature was that both proteins harbored a putative transmembrane region(s) near their amino-termini. Based on these similar features, the author set out to express the *sppA*$_{Tk}$ gene and examine the enzymatic properties of the recombinant protein.

```
Tko      1  MDDRIWKYVSAVLTLLLALSIVSVAVLYYQVNPPAVPANQTGFVIETPSNFTLVCEGSVN   60
Mja      1  ------------------------------------MKKIYIILLILFVILISLIGASI   23
Tac         ------------------------------------------------------------
PaI      1  -----MAKALALAGLVLAIAALALGIALWWGLLNAERTITMKIDDVRSSITAEVKTLNSS   55
Neq         ------------------------------------------------------------
Hba      1  ---------------------------------------------MTNPIDKHAKTLAT   14
Bsp      1  --------------------MNAKRWIALVIALGIFGVSIIVSISMSFFESVKGAQTDLT   40
Eco    244  AANRQIPAEQVFPGAQGLLEGLTKTGGDTAKYALENKLVDALASSAEIEKALTKEFGWSK  303

Tko     61  TTSTEVAELRDQIAFLQRLVSSNQGGTT---IAVVPIFGIIDDYTALQVIPLLRNLAMNE  117
Mja     24  LLVMSLSGENVDLFGGEKIAKVYLCNEIYFDYNQGDGIFPQQKKDARYYINLLDDLEKDD   83
Tac      1  ----------------------------MYILRTRIEGTITQQLYRSYYPIFSFAENKR   31
PaI     56  LTERLASINKTLEELTVRVKRLEKRAAAEREIVIVPVDQPIFDYYVDFLIKYVKKLQFDN  115
Neq      1  MNDIKDIIKRNSWILFLVGIILFSHIYSILTKGNVAVIYIDKPITSDFAEKVIDFLKEAK   60
Hba     15  SWTFIIVVAAIVGGVGGVALQSGSDSGPENSVAVVNIESAITGGTGDAVAKELRSIRGND   74
Bsp     41  SLTDESQEKTLENGSPSSKIAVLEVSGTIQDNGDSSSLLGADGYNHRTFLKNLERAKDDK  100
Eco    304  TDKNYRAISYYDYALKTPADTGDSIGVVFANGAIMDGEETQGNVGGDTTAAQIRDARLDP  363

Tko    118  SVGGVLLWIESPGGVVGPVINIHSEIKKLSLV--KPVVAYSGDIIASGGYYIAVGAQKIV  175
Mja     84  SVKGVLLVVNSPGGEVIASEKLARKVEELAKK--KPVVVYVEGLDASGAYMVSAPADYIV  141
Tac     32  SVAGLVLVFNSGGGDAVASQLMFEMIRKIRKK--KPVYSFIQGICASGAYWISAGSSKIY   89
PaI    116  KTAGVILLINSPGGAVGATERLYSTIKGLN----KTVYAVVAGLAASGAYYTAVAAGRIY  171
Neq     61  DYKAIVLYIDSPGGAPEPTYRIIKYLDRIN----KTKISYIAQYGTSASYWIATHTNKIF  116
Hba     75  SIDAVVLRVSSPGGAVSGSEVQYRAVKRLAQE--KPVVTSVRGPAASGGYYTIAPTDKIY  132
Bsp    101  TVKGIVLKVNSPGGGVYESAEIHKKLEEIKKETKKPIYVSMGSMAASGGYYISTAADKIF  160
Eco    364  KVKAIVLRVNSPGGSVTASEVIRAELAAARAAG-KPVVVSMGGMAASGGYWISTPANYIV  422
                  *   * ☆☆                    *         ☆  ☆       *

Tko    176  ASPLAEVGSIGVIYVHYDLEKNYEMNGIKVNVFKTGKHKDMGAEWRDLTPEEREKITEMV  235
Mja    142  AEKHSIVGSIGVRMDLMHYYGLMKKLGINVTTIKAGKYKDIGSPFRPMTKEEKEYLQKMI  201
Tac     90  SLDTSLIGSIGVISMIPYIKPLLDKIGVEMKIYKVGKYKDMLSSYREPSDEENEHYMRVL  149
PaI    172  ATPSSWVGSIGVVALLWPDEYLID---LPDYIYTTGPFKYYGMDLTEFYNDIEKTRANFV  228
Neq    117  ANELSFVGSVGVLIGKIDLSGLLSKLGVKYYSFSKGKYKEFSNPLLPLDNYTKQYYNELA  176
Hba    133  VTPSSLVGSVGVISSVSENNGVPSR-------WKSAPDKGTTGPADKARARAATFRQSFL  185
Bsp    161  ATPETLTGSLGVIMESVNYSKLADKLGISFETIKSGAHKDIMSPSREMTKEEKNIMQSMV  220
Eco    423  ANPSTLTGSIGIFGVITTVENSLDSIGVHTDGVSTSPLADVS-ITRALPPEAQLMMQLSI  481
                   ☆* ☆*                          *

Tko    236  NTYFQAFISAVSEGRNMTIDEVKNFSTGETWFAENVTG-ALVDELGGMDTAIDVLEKLMN  294
Mja    202  NETYMDFVKWVAEHRHLSINYTLKIADGKIYSGEDAKKVGLVDEVGTEEDALKKLEQLAN  261
Tac    150  NDVYRKFRDSVMTERNIPEDKMEDIAQGQIFSPSMAMENRLIDGIGTMDSMMDDMYR-QL  208
PaI    229  AAVLKGRAGRLKADP-------EVFETAKIFNAETALRLGLIDKIGGLWDAVEDMARELG  281
Neq    177  DKLYNFFLTDVLEHRDIKKECLSKVKESTIFLGIEAKKCGLIDYIGTMDDVKAYLEKTLK  236
Hba    186  DVVMNERGDDLTVDR-------ETIGRAQIYAGNKAVEIGLADEIGGLDAAIADAADRAS  238
Bsp    221  DNSYEGFVDVISKGRGMPKAEVKKIADGRVYDGRQAKKLNLVDELGFYDDTITAMKKDHK  280
Eco    482  ENGYKRFITLVADARHSTPEQIDKIAQGHVWTGQDAKANGLVDSLGDFDDAVAKAAELAK  541
                         *                          * ☆  *

Tko    295  VSGAKVVVYKDLETPEEFGVYGSTALYIDPRYLTPLIGGG--------------------  334
Mja    262  VSNPEIVEYGLEENKGLFGLTY-----YLGYGIGKGIGEVLYGMEKINGRVELLS-----  311
Tac    209  GRKYKTRDILPRRPWIMRFLGT-------------------------------------  230
PaI    282  LKNYTVVDIYEKYNATGFGIVVPLISGNKIPLQFLMNVSAPPVFYLWPGAIQIPHHPINV  341
Neq    237  IKVKFEPFREKALFYLSFPI----------------------------------------  256
Hba    239  VQNYGVVYRSPGGLGGLLSIIGGNSDTGAANAVSASEFCTNEYLAYAPQAGPDLEVIQNA  298
Bsp    281  DLKNASVISYEESFGLGSLFSMGANKMFKSEIDFLNMREILSQSGSPRMMYLYAK-----  335
Eco    542  VKQWHLEYYVDEPTFFDKVMDNMSGSVRAMLPDAFQAMLPAPLASVASTVKSESDKLAAF  601
```

Fig. 1. Sequence alignment of putative signal peptide peptidase proteins from *Escherichia coli*, *Bacillus subtilis*, and representative archaeal strains. The 19 archaeal SppA sequences described in Fig. 2 were aligned with the sequences of SppA from *E. coli* and *B. subtilis*. After alignment, representative sequences were selected. Asterisks indicate highly conserved residues that were present in at least 18 of the 21 sequences aligned. Among these, residues that were conserved in all sequences examined are indicated by stars. The bar above the alignment indicates the putative transmembrane domain of the signal peptide peptidase from *T. kodakaraensis*. Arrowheads indicate the residues that immediately follow the artificial Met residue incorporated in ΔN29SppA$_{Tk}$ and ΔN54SppA$_{Tk}$. Tko, *T. kodakaraensis*; Mja, *Methanocaldococcus jannaschii*; Tac, *Thermoplasma acidophilum*; PaI, *Pyrobaculum aerophilum* I; Neq, *Nanoarchaeum equitans*; Hba, *Halobacterium* sp. strain NRC-1; Bsp, *B. subtilis*; Eco, *E. coli*. Due to their lengths, not all residues are shown for PaI (608 residues), Hba (300 residues), and Eco (618 residues).

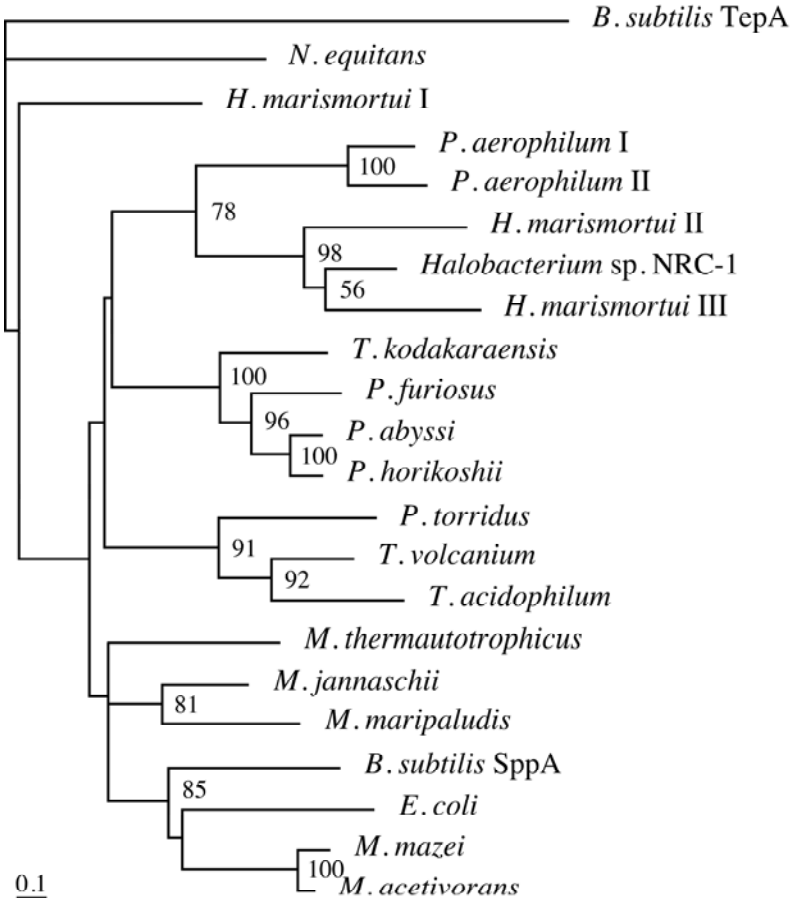


Fig. 2. Phylogenetic tree of putative archaeal SppA sequences. SppA sequences from archaea that displayed similarity to $SppA_{Ec}$ and $SppA_{Tk}$ were analyzed along with the sequences of $SppA_{Ec}$ and the SppA and TepA from *Bacillus subtilis*. The proteins used (accession numbers) were *B. subtilis* SppA (CAB14931) and TepA (CAB13552), *Escherichia coli* (BAA15557), *Haloarcula marismortui* I (AAV45638), II (AAV46904), and III (AAV47811), *Halobacterium* sp. strain NRC-1 (AAG19125), *Methanocaldococcus jannaschii* (AAB98642), *Methanococcus maripaludis* (CAF30625), *Methanosarcina acetivorans* (AAM07395), *Methanosarcina mazei* (AAM30562), *Methanothermobacter thermautotrophicus* (AAB85306), *Nanoarchaeum equitans* (AAR39164), *Picrophilus torridus* (AAT42796), *Pyrobaculum aerophilum* I (AAL65089) and II (AAL64441), *Pyrococcus abyssi* (CAB49512), *Pyrococcus furiosus* (AAL81707), *Pyrococcus horikoshii* (BAA30681), *Thermococcus kodakaraensis* (BAD85353), *Thermoplasma acidophilum* (CAC11222), and *Thermoplasma volcanium* (BAB59171). Only bootstrap values above 50 are indicated.

Expression of the *sppA$_{Tk}$* gene in *E. coli* and purification of the recombinant protein

The putative transmembrane domain of SppA$_{Tk}$ corresponds to residues Lys7 to Tyr29 (Fig. 1). In order to characterize the catalytic domain of the protein, this region was omitted when constructing the expression vector. An artificial Met residue was incorporated in the place of Tyr29, and this gene was expressed in *E. coli*. As expected, a soluble protein (ΔN29SppA$_{Tk}$) was obtained, which was resistant to heat treatment at 85˚C for 15 min. The author found that the thermostable protein exhibited peptidase activity towards peptide substrates such as Ala-Ala-Phe-MCA. The recombinant protein was purified with ammonium sulfate fractionation, anion exchange chromatography, and gel filtration chromatography. During these procedures, the author observed a gradual decrease in the molecular weight of the protein as judged by SDS-PAGE, leading to three major molecular species (Fig. 3). The amino-terminal amino acid sequences of each species was determined and revealed that in the smaller proteins, degradation at the amino-terminal region had occurred, probably due to autoproteolysis. As the smallest species harbored the Cys55 at its extreme amino-terminus, the author reconstructed an expression plasmid so that translation initiated at a Met residue adjacent to Cys55. The protein produced (ΔN54SppA$_{Tk}$) was purified with the methods mentioned above (Fig. 3), and was found to be relatively stable in terms of proteolytic degradation. This protein, ΔN54SppA$_{Tk}$, was therefore used for further biochemical examination. Although not shown in the results described below, ΔN29SppA$_{Tk}$ and ΔN54SppA$_{Tk}$ displayed similar tendencies in terms of specific activity, pH dependency, and substrate preference using the FRETS peptide library.

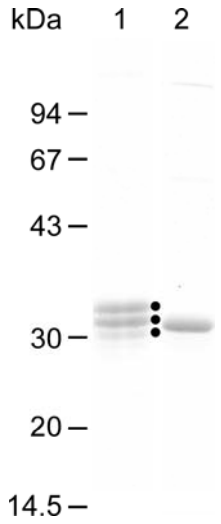


Fig. 3. SDS-PAGE analysis of purified ΔN29SppA$_{Tk}$ (lane 1) and ΔN54SppA$_{Tk}$ (lane 2). Solid dots to the right of lane 1 indicate the purified ΔN29SppA$_{Tk}$ and the major degradation products.

Oligomeric form of ΔN54SppA$_{Tk}$

The molecular mass of the purified ΔN54SppA$_{Tk}$ using gel filtration chromatography was estimated to be slightly larger than that of catalase (232 kDa). The molecular mass of a single subunit of ΔN54SppA$_{Tk}$ is 30,421 Da, suggesting that the protein formed an octamer. Besides this major peak, the author observed a second peak at 460 to 500 kDa, which may correspond to a hexadecameric form of the enzyme. The octameric form of the protein was used for further examination. It should be noted that the results obtained here are those of a truncated protein, and may not accurately reflect the oligomeric state of the native protein, which can be assumed to be associated with the membrane. It has previously been reported that the oligomeric form of the native SppA$_{Ec}$ was suggested to be tetrameric through crosslinking experiments, while the results of native PAGE raised the possibility of an even higher oligomeric form (15).

Optimal pH and temperature

The effects of pH and temperature on the activity of ΔN54SppA$_{Tk}$ were examined with the substrate Ala-Ala-Phe-MCA. As shown in Fig. 4A, the protein exhibited maximal activity at an unexpectedly high pH range of 10.0 to 10.5. High levels of activity were maintained at even higher pH values of 11.5 (58% relative to the activity observed at pH 10.0, CHES) and 12.0 (50%). The effects of temperature on activity were analyzed with the same substrate at pH 10.0. Maximum activity under the conditions examined was observed at approximately 80˚C (Fig. 4B). The Arrhenius plot gave a constant slope from 30˚C to 60˚C, and the activation energy of the reaction was calculated to be 54 kJ mol^{-1} (Fig. 4C). The thermostability of the enzyme was examined for 72 h at pH 10.0 (Fig. 4D), and alkaline stability was measured at 60˚C for 48 h (Fig. 4E). In terms of temperature, the author observed extremely high stability at 60 and 70˚C, with over 60% of the initial activity remaining after 72 h. At 80˚C, the author detected a relatively greater decrease in activity at the initial phases of incubation, which was consistently reproduced in multiple experiments. However after 30 min, the enzyme seemed to stabilize and follow the usual deactivation kinetics, where the deactivation rate ($-dNE/dt$) is proportional to the amount of enzyme (N_E), or $-dN_E/dt = kN_E$. The author still observed over 30% residual activity after 48 h at 80˚C and pH 10.0. ΔN54SppA$_{Tk}$ also exhibited high alkaline stability. Nearly 50% residual activity was observed after 24 h at pH 11.5 and 60˚C, and the calculated half-life at pH 12.0 was approximately 10 h.

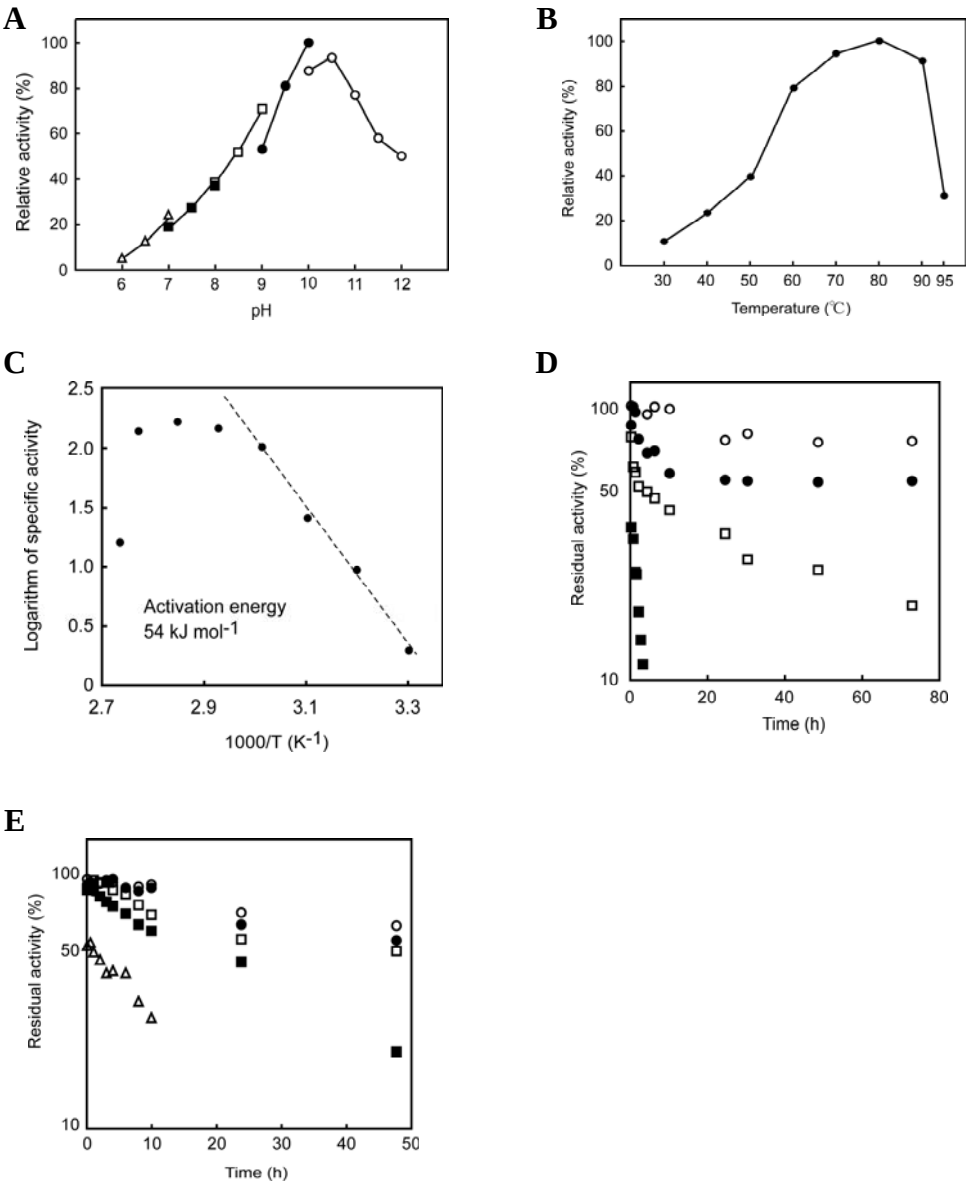


Fig. 4. (A) Effect of pH on the activity of ΔN54SppA$_{Tk}$. Measurements were performed at 60˚C in the following buffers at a concentration of 50 mM: MES-NaOH (open triangles), HEPES-NaOH (solid squares), Bicine-NaOH (open squares), CHES-NaOH (solid circles), and CAPS-NaOH (open circles). (B) Effect of temperature on the activity of ΔN54SppA$_{Tk}$. Reactions were carried out in 50 mM CHES-NaOH (pH 10.0). (C) Arrhenius plot of B, indicating the activation energy of substrate hydrolysis catalyzed by ΔN54SppA$_{Tk}$. (D) Thermostability of ΔN54SppA$_{Tk}$ at various temperatures. Incubation of the enzyme was carried out in 50 mM CHES-NaOH (pH 10.0). Symbols: 60˚C, open circles; 70˚C, solid circles; 80˚C, open squares; 90˚C, solid squares. (E) Stability of ΔN54SppA$_{Tk}$ at various pHs. Enzyme incubation was carried out at 60˚C in 50 mM CAPS-NaOH (pH 10.0, open circles; pH 10.5, solid circles; pH 11.0, open squares; pH 11.5, solid squares; pH 12.0, open triangles). All activity measurements (A to E) were carried out with 200 μM Ala-Ala-Phe-MCA.

Effects of various inhibitors

The effects of various protease inhibitors were examined at 60˚C and pH 10.0 (Fig. 5). Leupeptin, chymostatin, and antipain exhibited relatively strong inhibition, leading to an 80% or higher decrease in activity at 200 μM. Diisopropyl fluorophosphate, which specifically reacts with serine residues, also strongly inhibited the activity of $SppA_{Tk}$ at 1 mM. In contrast, addition of EDTA, a typical inhibitor of metalloproteases, and pepstatin, an inhibitor of aspartate proteases, did not result in significant decreases in activity. Although the effects of phenylmethylsulfonyl fluoride were lower than expected, the results as a whole agree with the assumption that $SppA_{Tk}$ is a serine protease.

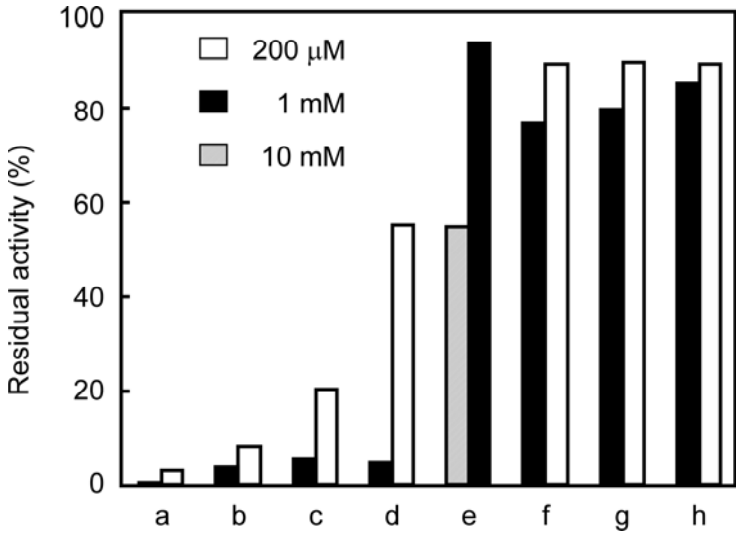


Fig. 5. Effect of various protease inhibitors on the activity of $\Delta N54SppA_{Tk}$. The inhibitors examined were leupeptin (a), antipain (b), chymostatin (c), diisopropyl fluorophosphate (d), phenylmethylsulfonyl fluoride (e), pepstatin (f), elastatinal (g), and EDTA (h). Activity measurements were performed at 60˚C with the indicated inhibitor concentrations. The concentration of the substrate Ala-Ala-Phe-MCA was 200 μM.

Preference of ΔN54SppA$_{Tk}$ for various peptidyl-MCA substrates

The author examined the activity of ΔN54SppA$_{Tk}$ on various peptidyl-MCA substrates shown in Table 1. At pH 10.0, the author found that Ala-Ala-Phe-MCA was by far the preferred substrate, followed by moderate activities towards *N*-glutaryl (Glt)-Ala-Ala-Phe-MCA and *N*-benzyloxycarbonyl (*Z*)-Val-Lys-Met-MCA. At pH 8.0, the preference became stricter, with Ala-Ala-Phe-MCA and Glt-Ala-Ala-Phe-MCA the only substrates leading to significant cleavage.

Table 1. Activity of ΔN54SppA$_{Tk}$ towards various peptidyl-MCA substrates.

Substrate	Specific activity (μmol mg protein^{-1} min^{-1})		
	pH 10.0	(%)	pH 8.0
Ala-Ala-Phe-MCA	9.80	100	4.70
Glt-Ala-Ala-Phe-MCA	2.85	29	0.98
Z-Val-Lys-Met-MCA	1.05	11	0.05
Phe-MCA	0.15	1.5	0.04
Suc-Ala-Ala-Ala-MCA	0.06	0.6	0.04
Suc-Ile-Ile-Trp-MCA	0.03	0.3	n.d.
Suc-Leu-Leu-Val-Tyr-MCA	0.02	0.2	0.03
Ala-MCA	n.d.	–	0.04
Suc(OMe)-Ala-Ala-Pro-Val-MCA	n.d.	–	0.01
Suc-Ala-Ala-Pro-Phe-MCA	n.d.	–	0.01
Z-Leu-Leu-Glu-MCA	n.d.	–	n.d.
Z-Ala-Ala-Asn-MCA	n.d.	–	n.d.
Z-Leu-Leu-Leu-MCA	n.d.	–	n.d.
Z-Leu-Arg-Gly-Gly-MCA	n.d.	–	n.d.

Activities were examined with a substrate concentration of 300 μM in 50 mM CHES (pH 10.0) or 50 mM HEPES (pH 8.0) at 60˚C.

n.d., activity not detected.

Examining the substrate preference of ΔN54SppA$_{Tk}$ with a FRETS peptide library

The author next evaluated the peptidase activity of ΔN54SppA$_{Tk}$ against a FRETS peptide library described in MATERIALS AND METHODS. This analysis provides detailed information on the preference of a peptidase towards substrate residues at the P-1, P-2, P-3, and, in some cases, the P-4 position. The P-1, P-2 and P-3 sites represent the region of the protease that recognizes the first, second and third amino acid at the amino-terminal side of the amide bond that is cleaved, respectively. An initial analysis was carried out with 19 substrates corresponding to all amino acids at the Xaa site with the exception of cysteine (Fig. 6A). The author detected a preference of ΔN54SppA$_{Tk}$ for substrates with rather small residues (Gly, Ser, Ala, and Thr) at the Xaa position (Fig. 6B). Cleavage rates of substrates with charged or aromatic residues were low. The author next selected six substrates with relatively high cleavage rates (Xaa = Gly, Ser, Ala, Thr, Asn and Val), and subjected the cleaved products to LC-mass spectrometry analysis. In order to identify products generated from the initial cleavage reaction of the substrate, reactions were stopped at various intervals corresponding to 15 to 30% substrate cleavage. The most abundant cleavage products for each of the six substrates are shown in Fig. 6C. In the case of Xaa = Gly, the author found that cleavage predominantly occurred at the carboxy-terminal amide bond of Ala, and not the expected Gly residue. In this case, the author can obtain insight on the preferred residues at the P-3 and P-4 sites. At the P-3 site, Ile led to the highest levels of cleavage, followed by Tyr and Pro. Products with Asp at the P-3 position could not be found. Concerning the P-4 position, the author found a high preference for Arg. The author also detected products with Phe or Val at the P-4 site. As in the case of the P-3 residue, the author could not find products with the negatively charged Glu residue at the P-4

site. Among the products that were actually cleaved at the Gly residue, Tyr and Lys were found at the P-2 site, and Phe and Val at the P-3 site. With Xaa = Ala, Ser, Thr, Asn or Val, the author found a common tendency with the results described above for the Xaa = Gly substrate. In all five cases, the most preferred P-3 residue was predominantly Ile or Phe, followed by Val, while cleavage products with negatively charged residues at this site were not found. A preference for the positively charged Arg at the P-4 site was also observed for all five substrates. The results also indicated a broad substrate specificity of ΔN54SppA$_{Tk}$ towards the P-2 residue.

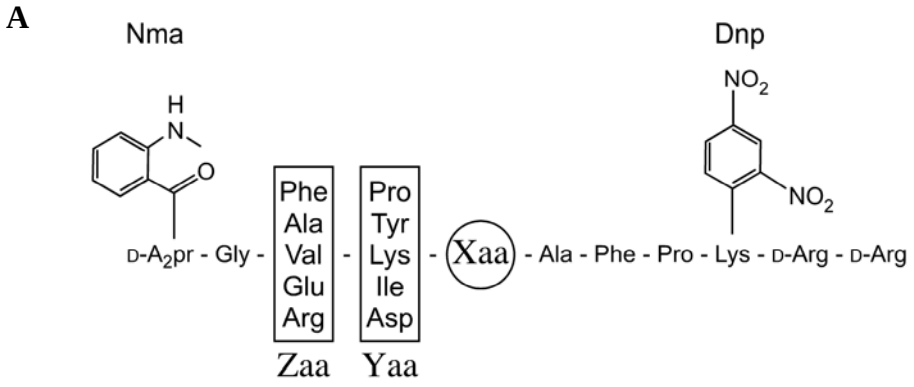
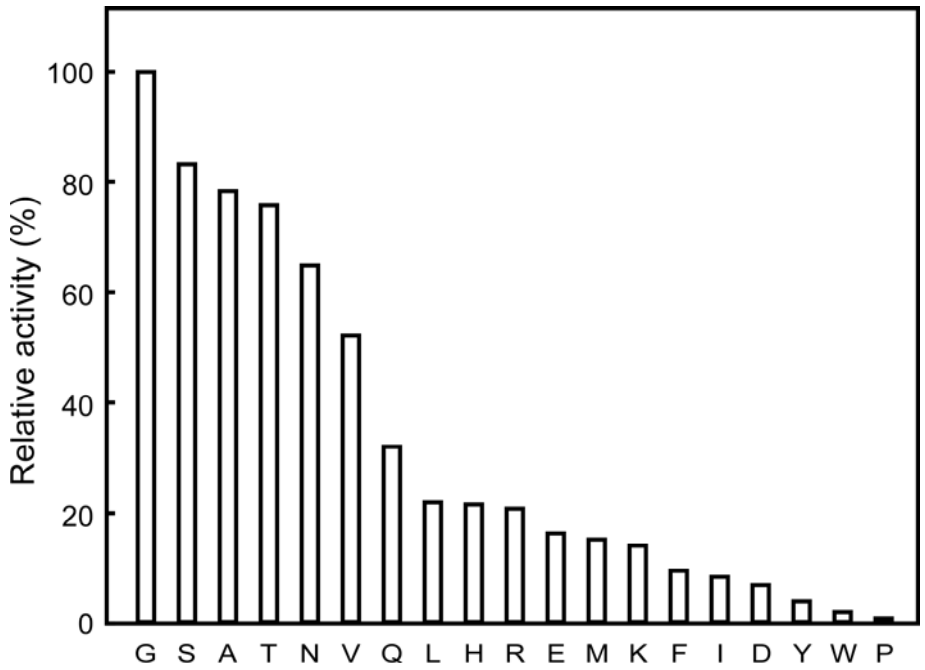


Fig. 6. (A) Structure of the peptide substrates from a FRETS peptide library utilized in B and C. (B) Peptidase activity towards various substrates from the FRETS peptide library. Amino acid residues at the Xaa position are indicated. Each substrate was examined at a concentration of 30 μM. (C) The major cleaved products of six substrates (Xaa = Gly, Ser, Ala, Thr, Asn and Val) were detected by LC-mass spectrometry. Relative quantities are indicated to the right of the sequences.

B



C

Xaa = Gly

P4	P3	P2	P1	P1'	
R	I	G	A ▼	F	100%
R	Y	G	A ▼	F	59%
R	P	G	A ▼	F	54%
R	K	G	A ▼	F	35%
F	I	G	A ▼	F	33%
G	F	Y	G ▼	A	30%

Xaa = Ser

P4	P3	P2	P1	P1'	
R	I	S	A ▼	F	100%
G	F	K	S ▼	A	75%
G	F	Y	S ▼	A	68%
G	V	Y	S ▼	A	59%
G	F	I	S ▼	A	57%
R	P	S	A ▼	F	47%

Xaa = Ala

P4	P3	P2	P1	P1'	
R	I	A	A ▼	F	100%
G	F	K	A ▼	A	90%
G	F	Y	A ▼	A	73%
G	F	I	A ▼	A	60%
F	I	G	A ▼	F	33%

Xaa = Thr

P4	P3	P2	P1	P1'	
R	I	T	A ▼	F	100%
G	F	K	T ▼	A	80%
G	F	Y	T ▼	A	69%
G	F	I	T ▼	A	45%
G	V	Y	T ▼	A	43%

Xaa = Asn

P4	P3	P2	P1	P1'	
R	I	N	A ▼	F	100%
G	F	K	N ▼	A	56%
G	F	Y	N ▼	A	49%
R	P	N	A ▼	F	38%
R	Y	N	A ▼	F	33%

Xaa = Val

P4	P3	P2	P1	P1'	
G	F	K	V ▼	A	100%
R	I	V	A ▼	F	82%
G	F	Y	V ▼	A	57%
G	V	K	V ▼	A	49%
G	F	I	V ▼	A	48%

DISCUSSION

The author has described the biochemical properties of ΔN54SppA$_{Tk}$, a truncated form of a putative signal peptide peptidase from *T. kodakaraensis*. One remarkable feature of the enzyme was its high activity at high alkaline pH. Alkaliphilic proteases have attracted much attention due to their high demand in application, particularly in the detergent industry (27-30). A number of enzymes are commercially available, including Savinase, subtilisin Carlsberg, and subtilisin BPN'. Various strategies in the protein engineering field have been applied to alkaliphilic proteases with the aim to improve their (thermo)stability, and this has led to significant improvements in terms of catalytic efficiency and stability (28). However, as ΔN54SppA$_{Tk}$ was obtained from a hyperthermophile, the stability of the enzyme can be regarded as exceptionally high in comparison with previously known alkaline proteases/peptidases. The author has detected over 50% residual activity after incubation at 70˚C and pH 10.0 for 3 days.

As the enzyme exhibits intrinsic thermophilic and alkaliphilic properties, ΔN54SppA$_{Tk}$ should be an attractive target for future protein engineering focusing on the modification of its substrate specificity. The author has examined whether ΔN54SppA$_{Tk}$ exhibits hydrolase activity towards proteins at 60˚C and pH 10.0. Using bovine serum albumin, ovalbumin, α-casein, hemoglobin, and lysozyme, the author found that protease activity of ΔN54SppA$_{Tk}$ was dependent on the particular protein substrate. Degradation of bovine serum albumin and ovalbumin was not observed, while α-casein, hemoglobin, and lysozyme were degraded to various extents (Fig. 7). The results indicate that ΔN54SppA$_{Tk}$ is able to cleave substrates longer than the artificial substrates used in this study.

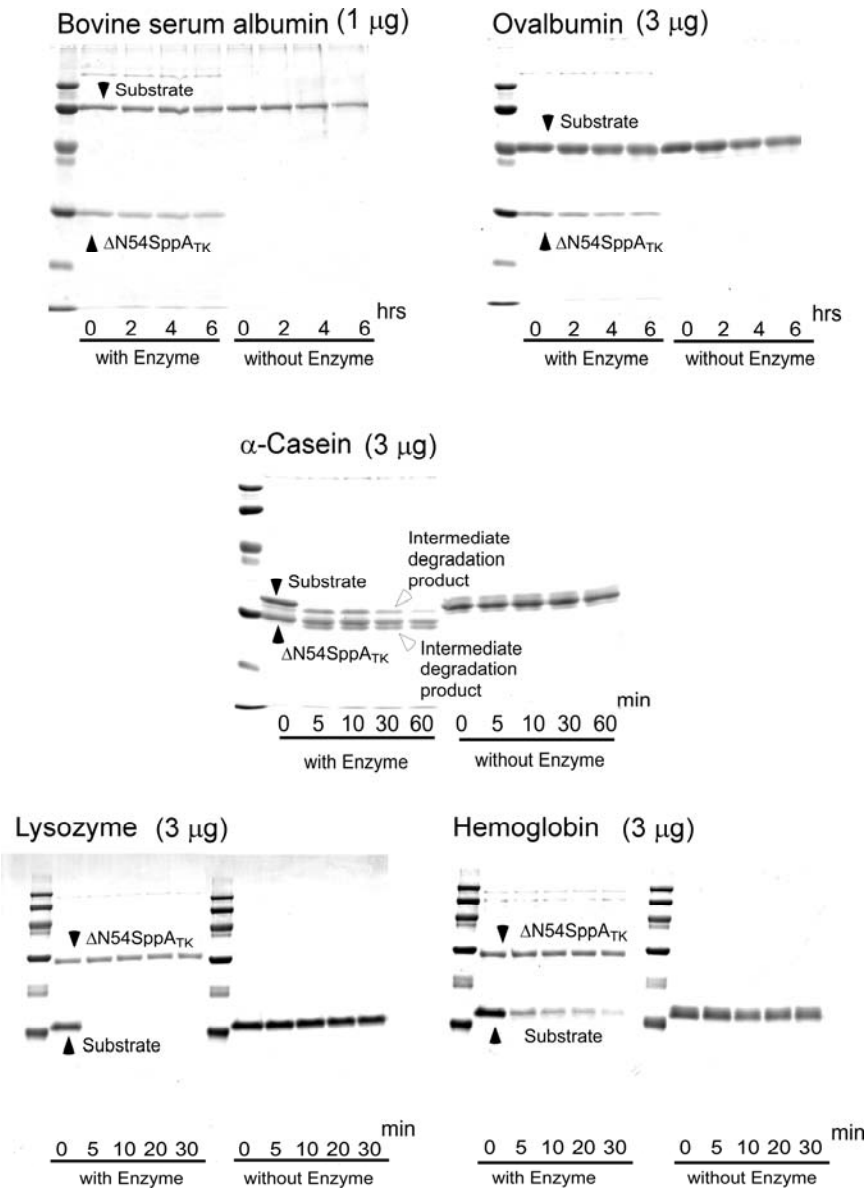


Fig. 7. Incubation was performed at 60˚C in a final volume of 1 ml with 1 μg of ΔN54SppA$_{Tk}$ and protein substrate (1 μg or 3 μg) in 50 mM CHES-NaOH (pH 10.0). After incubation, substrate degradation in the reaction mixture was confirmed by SDS-PAGE (12.5% or 15% concentration of acrylamide). As a control reaction, incubation was also performed without ΔN54SppA$_{Tk}$ in the reaction mixture.

$SppA_{Tk}$ is structurally categorized in the S49 family of serine proteases, a prokaryotic family of proteases of which still little is understood. There are no three-dimensional structures available, and moreover, the catalytic mechanism of SPPs and the residues involved have not yet been experimentally determined. Information on the enzymatic properties of previously identified SPPs is also limited. The specific activity of $SppA_{Ec}$ has been measured in several studies and ranges between 0.733 μmol mg^{-1} min^{-1} at 25˚C against Z-valine *p*-nitrophenyl ester (14) and 13.6 μmol mg^{-1} min^{-1} at 37˚C against Z-valine β-naphthyl ester (15). The specific activity of the archaeal $SppA_{Tk}$ was 2.59 μmol mg^{-1} min^{-1} at 40˚C with Ala-Ala-Phe-MCA as the substrate.

Residues preferred at the cleavage site have also been examined for $SppA_{Ec}$ with both synthetic substrates and signal peptides. The former revealed a preference of cleavage at the carboxy-terminal side of Ala, Leu, Val, Gly, and Phe (14), while the latter indicated a preference for Val, Leu, Ile, Gly, Thr, and Ala (17). As in the case of the specific activity described above, the substrates used and the conditions applied in these experiments vary greatly, making it difficult to accurately compare the two enzymes. One point that is worthy of note is that many residues (Ala, Val, Thr, and Gly) are commonly preferred at the P-1 site by both $SppA_{Ec}$ and $SppA_{Tk}$.

By performing a BLAST search against the genome sequences of archaea, the author was able to identify SppA homologs in most strains of *Euryarchaeota* (Fig. 2). As exceptions, the author could not identify highly similar homologs on the *A. fulgidus* and *Methanopyrus kandleri* genomes. Interestingly, most members of the *Crenarchaeota* do not seem to utilize structurally related SPPs. The two protein sequences from *Pyrobaculum aerophilum*, the only SppA homologs from *Crenarchaeota*, contained exceptionally long extensions in their carboxy-terminal

domains. Residues that are highly conserved among bacterial and archaeal SppA sequences are indicated in Fig. 1 and may be involved in catalysis and/or substrate specificity. In particular, Ser162 in $SppA_{Tk}$ is conserved in all SppA sequences and is also conserved in the functionally and structurally related *B. subtilis* TepA protease (19). Although this serine residue of TepA has not been experimentally analyzed, it is included in a region which displays similarity to the region containing the active site serine of *E. coli* ClpP, raising the possibilities that this residue is the nucleophile in all of these proteins. However, the His and Asp/Glu residues that constitute the conventional catalytic triad found in ClpP are not conserved among the archaeal proteins and are not even present in the SppA proteins from *E. coli* and *B. subtilis*. In order to accurately determine the active-site residues of SppA proteins, including the nucleophilic serine, site-directed mutagenesis studies (described in Chapter 2) will be necessary.

Using the FRETS peptide library, the author has been able to clarify some of the preferences of $\Delta N54SppA_{Tk}$ towards residues at the P-1, P-2, P-3, and P-4 sites. A relatively small side chain seems to be preferred at the P-1 position. The specificity at the P-2 position can be regarded as broad. Hydrophobic and/or aromatic residues are recognized most at the P-3 site, while the positively charged Arg enhances the activity of $\Delta N54SppA_{Tk}$ when present at the P-4 position. The results also indicate that the presence of acidic residues at any one of the sites from P-2 to P-4 has a negative effect on the substrate recognition of $\Delta N54SppA_{Tk}$. With the MCA substrates, a direct comparison among substrates to determine the residue preference of $\Delta N54SppA_{Tk}$ was difficult, as multiple factors differ even between two given substrates. One point that can be noted is that a negative charge at the P-4 position has a large negative effect on

substrate recognition (Ala-Ala-Phe-MCA > Glt-Ala-Ala-Phe-MCA).

Taking into account the specificity of ΔN54SppA$_{Tk}$, the author examined a vast number of putative signal sequences that were identified on the *T. kodakaraensis* genome using the SOSUI program. Some representative signal sequences, those from four proteins that have been experimentally proven to be secreted from *T. kodakaraensis* ((31-34); unpublished data), are shown in Fig. 8. As in the case of most putative signal sequences on the *T. kodakaraensis* genome, acidic residues are not found, consistent with the fact that ΔN54SppA$_{Tk}$ does not cleave peptides with acidic residues in the P-2 to P-4 sites. Further, the author found a number of candidate sequences in each signal sequence that can be presumed to be efficiently recognized and cleaved by ΔN54SppA$_{Tk}$.

Although future gene disruption studies will be necessary to confirm the physiological role of SppA$_{Tk}$, the enzymatic properties of the enzyme are in good agreement with the assumption that SppA$_{Tk}$ functions as an SPP in *T. kodakaraensis*.

Fig. 8. Predicted signal sequences of four *T. kodakaraensis* proteins that have been biochemically characterized. Prediction was carried out with the SOSUI program. The signal sequence of α-amylase has been confirmed experimentally. Scissors indicate sites that can be presumed to be cleaved by SppA$_{Tk}$ when its substrate specificity is taken into account.

SUMMARY

The author has performed the first biochemical characterization of a putative archaeal signal peptide peptidase ($SppA_{Tk}$) from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. $SppA_{Tk}$, comprised of 334 residues, was much smaller than its counterpart from *Escherichia coli* (618 residues) and harbored a single predicted transmembrane domain near its amino-terminus. A truncated mutant protein without the amino-terminal 54 amino acid residues ($\Delta N54SppA_{Tk}$) was found to be stable against autoproteolysis and was further examined. $\Delta N54SppA_{Tk}$ exhibited peptidase activity towards fluorogenic peptide substrates and was found to be highly thermostable. Moreover, the enzyme displayed a remarkable stability and preference for alkaline pH, with optimal activity detected at pH 10.0. The substrate specificity of the enzyme was examined in detail with a FRETS peptide library. By analyzing the cleavage products with liquid chromatography-mass spectrometry, $\Delta N54SppA_{Tk}$ was found to efficiently cleave peptides with a relatively small side chain at the P-1 position and a hydrophobic or aromatic residue at the P-3 position. The positively charged Arg residue was preferred at the P-4 position, while substrates with negatively charged residues at the P-2, P-3, or P-4 position were not cleaved. When predicted signal sequences from the *T. kodakaraensis* genome sequence were examined, the author found that the substrate specificity of $\Delta N54SppA_{Tk}$ was in good agreement with its presumed role as a signal peptide peptidase in this archaeon.

REFERENCES

1. **Ring, G. & Eichler, J.**, Extreme secretion: protein translocation across the archaeal plasma membrane. *J. Bioenerg. Biomembr.*, 36, 35-45, 2004.

2. **Cao, T. B. & Saier, M. H., Jr.**, The general protein secretory pathway: phylogenetic analyses leading to evolutionary conclusions. *Biochim. Biophys. Acta*, 1609, 115-125, 2003.

3. **Bardy, S. L., Ng, S. Y. M. & Jarrell, K. F.**, Recent advances in the structure and assembly of the archaeal flagellum. *J. Mol. Microbiol. Biotechnol.*, 7, 41-51, 2004.

4. **Bhuiyan, S. H., Gowda, K., Hotokezaka, H. & Zwieb, C.**, Assembly of archaeal signal recognition particle from recombinant components. *Nucleic Acids Res.*, 28, 1365-1373, 2000.

5. **Rosendal, K. R., Wild, K., Montoya, G. & Sinning, I.**, Crystal structure of the complete core of archaeal signal recognition particle and implications for interdomain communication. *Proc. Natl. Acad. Sci. USA*, 100, 14701-14706, 2003.

6. **van den Berg, B., Clemons, W. M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C. & Rapoport, T. A.**, X-ray structure of a protein-conducting channel. *Nature*, 427, 36-44, 2004.

7. **Hussain, M., Ichihara, S. & Mizushima, S.**, Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the *Escherichia coli* outer membrane. *J. Biol. Chem.*, 257, 5177-5182, 1982.

8. **Paetzel, M., Karla, A., Strynadka, N. C. J. & Dalbey, R. E.**, Signal peptidases. *Chem. Rev.*, 102, 4549-4579, 2002.

9. **Lemberg, M. K. & Martoglio, B.**, On the mechanism of SPP-catalysed intramembrane proteolysis; conformational control of peptide bond hydrolysis in the plane of the membrane. *FEBS Lett.*, 564, 213-218, 2004.

10. **Wolfe, M. S. & Kopan, R.**, Intramembrane proteolysis: theme and variations. *Science*, 305, 1119-1123, 2004.

11. **Braud, V. M., Allan, D. S. J., O'Callaghan, C. A., Söderström, K., D'Andrea, A., Ogg, G. S., Lazetic, S., Young, N. T., Bell, J. I., Phillips, J. H., Lanier, L. L. & McMichael, A. J.**, HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature*, 391, 795-799, 1998.

12. **Lemberg, M. K., Bland, F. A., Weihofen, A., Braud, V. M. & Martoglio, B.**, Intramembrane proteolysis of signal peptides: an essential step in the generation of HLA-E epitopes. *J. Immunol.*, 167, 6441-6446, 2001.

13. **Ichihara, S., Beppu, N. & Mizushima, S.**, Protease IV, a cytoplasmic membrane protein of *Escherichia coli*, has signal peptide peptidase activity. *J. Biol. Chem.*, 259, 9853-9857, 1984.

14. **Pacaud, M.**, Purification and characterization of two novel proteolytic enzymes in membranes of *Escherichia coli*. Protease IV and protease V. *J. Biol. Chem.*, 257, 4333-4339, 1982.

15. **Ichihara, S., Suzuki, T., Suzuki, M. & Mizushima, S.**, Molecular cloning and sequencing of the *sppA* gene and characterization of the encoded protease IV, a signal peptide peptidase, of *Escherichia coli*. *J. Biol. Chem.*, 261, 9405-9411, 1986.

16. **Suzuki, T., Itoh, A., Ichihara, S. & Mizushima, S.**, Characterization of the *sppA* gene coding for protease IV, a signal peptide peptidase of *Escherichia coli*. *J. Bacteriol.*, 169, 2523-2528, 1987.

17. **Novak, P. & Dev, I. K.**, Degradation of a signal peptide by protease IV and oligopeptidase A. *J. Bacteriol.*, 170, 5067-5075, 1988.

18. **Novak, P., Ray, P. H. & Dev, I. K.**, Localization and purification of two enzymes from *Escherichia coli* capable of hydrolyzing a signal peptide. *J. Biol. Chem.*, 261, 420-427, 1986.

19. **Bolhuis, A., Matzen, A., Hyyryläinen, H.-L., Kontinen, V. P., Meima, R., Chapuis, J., Venema, G., Bron, S., Freudl, R. & van Dijl, J. M.**, Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required for efficient translocation and processing of secretory proteins. *J. Biol. Chem.*, 274, 24585-24592, 1999.

20. **Ng, S. Y. M. & Jarrell, K. F.**, Cloning and characterization of archaeal type I signal peptidase from *Methanococcus voltae*. *J. Bacteriol.*, 185, 5936-5942, 2003.

21. **Bardy, S. L., Ng, S. Y. M., Carnegie, D. S. & Jarrell, K. F.**, Site-directed mutagenesis analysis of amino acids critical for activity of the type I signal peptidase of the archaeon *Methanococcus voltae*. *J. Bacteriol.*, 187, 1188-1191, 2005.

22. **Bardy, S. L. & Jarrell, K. F.**, Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. *Mol. Microbiol.*, 50, 1339-1347, 2003.

23. **Albers, S.-V., Szabó, Z. & Driessen, A. J. M.**, Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J. Bacteriol.*, 185, 3918-3925, 2003.

24. **Ramakrishnan, V. & Adams, M. W. W.**, In *Archaea-A laboratory manual*, ed. F. T. Robb & A. R. Place. Cold Spring Harbor Laboratory Press, Plainview, 1995, pp. 95-96.

25. **Tanskul, S., Oda, K., Oyama, H., Noparatnaraporn, N., Tsunemi, M. & Takada, K.**, Substrate specificity of alkaline serine proteinase isolated from photosynthetic bacterium, *Rubrivivax gelatinosus* KDDS1. *Biochem. Biophys. Res. Commun.*, 309, 547-551, 2003.

26. **Rawlings, N. D., Tolle, D. P. & Barrett, A. J.**, MEROPS: the peptidase database. *Nucleic Acids Res.*, 32 Database issue, D160-D164, 2004.

27. **Atomi, H.**, Recent progress towards the application of hyperthermophiles and their enzymes. *Curr. Opin. Chem. Biol.*, 9, 166-173, 2005.

28. **Gupta, R., Beg, Q. K. & Lorenz, P.**, Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.*, 59, 15-32, 2002.

29. **Ito, S., Kobayashi, T., Ara, K., Ozaki, K., Kawai, S. & Hatada, Y.**, Alkaline detergent enzymes from alkaliphiles: enzymatic properties, genetics, and structures. *Extremophiles*, 2, 185-190, 1998.

30. **Maurer, K.-H.**, Detergent proteases. *Curr. Opin. Biotechnol.*, 15, 330-334, 2004.

31. **Kannan, Y., Koga, Y., Inoue, Y., Haruki, M., Takagi, M., Imanaka, T., Morikawa, M. & Kanaya, S.**, Active subtilisin-like protease from a hyperthermophilic archaeon in a form with a putative prosequence. *Appl. Environ. Microbiol.*, 67, 2445-2452, 2001.

32. **Rashid, N., Cornista, J., Ezaki, S., Fukui, T., Atomi, H. & Imanaka, T.**, Characterization of an archaeal cyclodextrin glucanotransferase with a novel C-terminal domain. *J. Bacteriol.*, 184, 777-784, 2002.

33. **Tachibana, Y., Leclere, M. M., Fujiwara, S., Takagi, M. & Imanaka, T.**,

Cloning and expression of the α-amylase gene from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and characterization of the enzyme. *J. Ferment. Bioeng.*, 82, 224-232, 1996.

34. **Tanaka, T., Fujiwara, S., Nishikori, S., Fukui, T., Takagi, M. & Imanaka, T.**, A unique chitinase with dual active sites and triple substrate binding sites from the hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *Appl. Environ. Microbiol.*, 65, 5338-5344, 1999.

CHAPTER 2

Identification of the amino acid residues essential for proteolytic activity in an archaeal signal peptide peptidase

INTRODUCTION

As described in the GENERAL INTRODUCTION, secretion proteins and membrane proteins in many cases harbor signal sequences at their extreme amino-termini that are cleaved during translocation by signal peptidases (SPs) (1). The released signal peptides are subsequently cleaved into smaller fragments by signal peptide peptidases (SPPs) (2). In bacteria, SPP was first identified in *Escherichia coli*. A previously identified cytoplasmic membrane protein, protease IV (3), was found to exhibit SPP activity toward the signal peptide of the outer membrane lipoprotein (2). It is now presumed that in *E. coli*, this protein ($SppA_{Ec}$) initiates the degradation by introducing endoproteolytic cuts into the signal peptide, whereas other cytoplasmic proteases, such as oligopeptidase A, are responsible for complete degradation of the smaller fragments into free amino acids (4). SPP has also been identified and genetically characterized in the Gram-positive *Bacillus subtilis*. The enzyme, along with a cytoplasmic peptidase TepA, has been found to play an important role in signal peptide degradation in this organism (5).

In the *Archaea*, much remains to be understood on the mechanisms of protein secretion (6-8) and the fate of the signal peptide after its release from the precursor protein. In terms of SPs, the type I SP from *Methanococcus voltae* has been characterized, and a catalytic triad comprised from Ser52, His122, and Asp148 has been

determined to be critical for its peptidase activity (9, 10). FlaK, an aspartic protease essential for preflagellin signal cleavage has also been studied from this organism (11). In the *Crenarchaeota*, the homolog of bacterial type IV prepilin peptidases from *Sulfolobus solfataricus* (PibD) has been characterized, and residues on the substrate that are important for recognition by PibD have been examined (12).

As for archaeal SPPs, the author has carried out the first examination of an archaeal SPP ($SppA_{Tk}$) from the hyperthermophilic archaeon, *Thermococcus kodakaraensis* (Chapter 1). $SppA_{Tk}$ (334 residues) was much smaller in size compared with its bacterial counterpart $SppA_{Ec}$ (618 residues). A single, putative membrane-spanning domain was present in the amino-terminal region of the protein. It was found that $\Delta N54SppA_{Tk}$, a truncated protein without the amino-terminal 54 residues, was a soluble protein exhibiting peptidase activity and stable against autoproteolysis. The substrate specificity of $\Delta N54SppA_{Tk}$ examined with a FRETS peptide library was consistent with its presumed role as an SPP in *T. kodakaraensis*.

From the primary structures, archaeal and bacterial homologs of $SppA_{Tk}$ and $SppA_{Ec}$ are all members of the S49 family of peptidases included in the Clan SK (MEROPS, the peptidase database, http://merops.sanger.ac.uk/) (13). Although the eukaryotic SPPs have been determined to be aspartic proteases (14, 15), the effects of various inhibitors on $SppA_{Tk}$ (Chapter 1) and $SppA_{Ec}$ (3) strongly suggest that the enzymes are serine proteases. Although sequence comparisons among SppA homologs reveal the presence of several conserved serine residues, experimental evidence identifying the nucleophilic serine has not yet been obtained. Moreover, even with the sequence comparisons, it is still difficult to estimate what other residues might be involved in the catalytic mechanism of SPPs. In particular, His and Asp/Glu residues

that comprise the well known catalytic triad of serine proteases are not clearly conserved among the bacterial and archaeal SppA sequences.

To gain insight on the residues involved in the catalytic mechanism of prokaryotic SPPs, in this chapter the author has performed a detailed site-directed mutagenesis study on ΔN54SppA$_{Tk}$. Through the analyses of various mutant proteins, the author has been able to determine multiple residues that are essential or important for the activity of this protein. The results strongly suggest that ΔN54SppA$_{Tk}$ and other SppA homologs from the *Archaea* utilize a Ser-Lys dyad mechanism in peptide cleavage.

MATERIALS AND METHODS

Strains, media, and plasmids

As described in Chapter 1, *E. coli* BL21-CodonPlus(DE3)-RIL and the plasmid pET21a(+) were used for gene expression. *E. coli* JM109 (*dam*$^+$) was used in order to obtain methylated plasmid DNA sensitive to DpnI digestion. *E. coli* strains were cultivated in LB medium as described in Chapter 1.

DNA manipulation, sequence analysis, and site-directed mutagenesis

Isolation, purification and DNA sequencing of plasmid DNA were performed as described in Chapter 1. Sequence comparisons and alignments were performed with the ClustalW program provided by the DNA Data Bank of Japan. The expression plasmid used for production of recombinant ΔN54SppA$_{Tk}$ (Chapter 1) was amplified in *E. coli* JM109 (*dam*$^+$) so that the plasmid would be methylated. The methylated plasmid was used as a template for site-directed mutagenesis with the QuikChange XL

site-directed kit (Stratagene). The primers used to incorporate each mutation are shown in Table 1. After sequence confirmation, the plasmids were introduced into *E. coli* BL21-CodonPlus(DE3)-RIL cells.

Table 1. Primers used in this study.

Mutation	Primer sequences
S128A	5′-GTTCTCCTCTGGATTGAAGCCCCCGGTGGCGTAGTTGGG-3′
	5′-CCCAACTACGCCACCGGGGGCTTCAATCCAGAGGAGAAC-3′
G130A	5′-CTCTGGATTGAAAGTCCCGCCGGCGTAGTTGGGCCTGTT-3′
	5′-AACAGGCCCAACTACGCCGGCGGGACTTTCAATCCAGAG-3′
G131A	5′-TGGATTGAAAGTCCCGGTGCCGTAGTTGGGCCTGTTATT-3′
	5′-AATAACAGGCCCAACTACGGCACCGGGACTTTCAATCCA-3′
K150A	5′-AAAAAGTTGTCTTTAGTTGCCCCAGTCGTCGCTTACAGC-3′
	5′-GCTGTAAGCGACGACTGGGGCAACTAAAGACAACTTTTT-3′
S162A	5′-AGCGGGGATATCATAGCAGCCGGGGGATACTACATAGCA-3′
	5′-TGCTATGTAGTATCCCCCGGCTGCTATGATATCCCCGCT-3′
Y165A	5′-ATCATAGCATCAGGGGGAGCCTACATAGCAGTTGGGGCT-3′
	5′-AGCCCCAACTGCTATGTAGGCTCCCCCTGATGCTATGAT-3′
S184A	5′-CCGCTGGCTGAGGTCGGAGCCATCGGAGTTATCTACGTT-3′
	5′-AACGTAGATAACTCCGATGGCTCCGACCTCAGCCAGCGG-3′
H191A	5′-ATCGGAGTTATCTACGTTGCCTACGACCTGGAGAAGAAC-3′
	5′-GTTCTTCTCCAGGTCGTAGGCAACGTAGATAACTCCGAT-3′
K209A	5′-ATAAAGGTAAATGTATTCGCCACTGGTAAACACAAGGAC-3′
	5′-GTCCTTGTGTTTACCAGTGGCGAATACATTTACCTTTAT-3′
K214A	5′-TTCAAAACTGGTAAACACGCCGACATGGGGGCCGAGTGG-3′
	5′-CCACTCGGCCCCCATGTCGGCGTGTTTACCAGTTTTGAA-3′
D215A	5′-AAAACTGGTAAACACAAGGCCATGGGGGCCGAGTGGAGA-3′
	5′-TCTCCACTCGGCCCCCATGGCCTTGTGTTTACCAGTTTT-3′
R221A	5'-GACATGGGGGCCGAGTGGGCCGATTTAACGCCAGAAGAA-3'
	5'-TTCTTCTGGCGTTAAATCGGCCCACTCGGCCCCCATGTC-3'
E226A	5′-TGGAGAGATTTAACGCCAGCCGAACGCGAGAAGATAACG-3′
	5′-CGTTATCTTCTCGCGTTCGGCTGGCGTTAAATCTCTCCA-3′
E227A	5′-AGAGATTTAACGCCAGAAGCCCGCGAGAAGATAACGGAG-3′
	5′-CTCCGTTATCTTCTCGCGGGCTTCTGGCGTTAAATCTCT-3′
R250A	5′-AGCGCAGTGAGCGAGGGGGCCAACATGACCATTGATGAA-3′
	5′-TTCATCAATGGTCATGTTGGCCCCCTCGCTCACTGCGCT-3′
D277A	5′-GTTACTGGAGCCCTCGTCGCCGAGCTTGGGGGTATGGAC-3′
	5′-GTCCATACCCCCAAGCTCGGCGACGAGGGCTCCAGTAAC-3′

Expression and purification of wild-type and mutant ΔN54SppA$_{Tk}$ proteins

The recombinant *E. coli* cells were grown in LB medium, and gene expression was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside at the mid- to late exponential growth phase. After 6 h, the cells were collected, washed with 50 mM Tris-HCl (pH 8.0), and resuspended in the same buffer. Methods for cell disruption and protein purification were performed in the same manner as those used for the wild-type protein described in Chapter 1. The protein concentration was determined with a protein assay kit (Bio-Rad) using bovine serum albumin as a standard.

Enzyme activity measurements

Standard activity measurements using the substrate Ala-Ala-Phe-MCA are described in the previous chapter. Kinetic parameters were calculated with IGOR Pro version 5.0 (WaveMetrics, Lake Oswego, OR, USA).

Circular dichroism spectroscopy of wild-type and mutant enzymes

Each protein sample was prepared in 25 mM Tris-HCl (pH 8.0), 75 mM NaCl at a protein concentration of 0.1 mg ml^{-1}. A J-820 spectropolarimeter (Jasco, Tokyo, Japan) was used to measure ellipticity as a function of wavelength from 250 to 200 nm in 0.2-nm increments using a 0.1-cm cylindrical quartz cuvette. The samples were scanned one hundred times and averaged. The mean molar ellipticity $[\theta]$ (deg cm^2 dmol^{-1}) was calculated from the equation $[\theta]=\theta /10\ nCl$, where θ is the measured ellipticity in millidegrees, C is the molar concentration of enzyme subunits, l is the path length in centimeters, and n is the number of residues per subunit.

Examination of the substrate preference of wild-type and mutant enzymes

The substrate preferences of the wild-type, S128A, and Y165A proteins were examined with the FRETS peptide library as described in Chapter 1.

RESULTS

Highly conserved amino acid residues among the signal peptide peptidases from archaea and bacteria

To determine which residues should be selected for site-directed mutagenesis and subsequent biochemical analyses, the author aligned all SppA homologs found in the archaeal genomes, along with the biochemically and/or genetically characterized bacterial SppA from *E. coli* and *B. subtilis*. Among the 21 archaeal genomes that have been sequenced, the author could identify 19 SppA homologs in 16 organisms (see legend of Fig. 1). Although a number of other open reading frames have been annotated as putative SPPs, they displayed significantly lower degrees of similarity with $SppA_{Tk}$ and $SppA_{Ec}$ and were therefore not selected for further examination. The sequence comparison indicated that SppA homologs were not present in *Aeropyrum pernix*, *S. solfataricus*, and *Sulfolobus tokodaii* from the *Crenarchaeota* and *Methanopyrus kandleri* and *Archaeoglobus fulgidus* from the *Euryarchaeota*. A region corresponding to residues between Met115 and Lys291 in $SppA_{Tk}$ was fairly conserved among all sequences and presumably includes the catalytic core regions of these enzymes. An alignment of the sequences spanning this region, along with the mutations introduced in this study, is shown in Fig. 1. Although the eukaryotic SppA has been demonstrated to be an aspartate peptidase (14), inhibitor studies on the bacterial $SppA_{Ec}$ (3) and the archaeal $SppA_{Tk}$ (Chapter 1) have indicated that both of these enzymes are serine

peptidases. The author therefore initially focused on the presence of conserved serine residues. Ser128, Ser162, and Ser184 of $SppA_{Tk}$ were completely conserved in all sequences, with the only exceptions being a replacement of the first Ser to Thr in one of the three homologs from *Haloarcula marismortui* and a replacement of the third Ser to His in another homolog from this strain. In terms of the His residue in the well known Ser-His-Asp catalytic triad, only three His residues were found in this region of $SppA_{Tk}$ and were not highly conserved among the SppA sequences. His140 and His213 were not conserved even in the closely related *Pyrococcus* spp., whereas His191, although shared by $SppA_{Tk}$ and the three *Pyrococcus* enzymes, was not found in the other SppA sequences. As an alternative, the author found that the basic residue Arg250 was highly conserved among the SppA proteins. As for the Asp residue, Asp277 was the only acidic residue that was present without exception in all sequences (Fig. 1).

Production and purification of mutant $SppA_{Tk}$ proteins via site-directed mutagenesis

Based on the sequence alignment described above, the author first constructed the following six mutants; S128A, S162A, S184A, H191A, R250A, and D277A. The primers used to incorporate the mutations are shown in Table 1. The mutations were incorporated into the expression vector for $\Delta N54SppA_{Tk}$, a truncated protein without the amino-terminal 54 residues of $SppA_{Tk}$ (Chapter 1). Sequence analysis confirmed that only the intended mutations were introduced into the genes. Recombinant mutant proteins were produced in *E. coli* BL21-CodonPlus(DE3)-RIL cells and purified from the cell-free extracts by heat treatment at 85˚C for 15 min, followed by ammonium sulfate fractionation, anion exchange chromatography, and gel filtration

```
                     128 130 131                 150         162 165                 184     191
Tko    115  MNESVGGVLLWIESPGGVVGPVINIHSEIKKLSLV--KPVVAYSGDIIASGGYYIAVGAQKIVASPLAEVGSIGVIYVHYDLEKNYEMNG  202
Pab    117  GNDTIKGVLLWIESPGGYVGPVREIYNELKKLGYL--KPIVAYVSGYAYSGAYYIACAAREIIAEPLSEVGSIGVIYVHFNAEEYYKMNG  204
Pho    116  ENDTIKGVLVWIESPGGFVGPVREIYIELKKLDYI--KPVIAYVSGYAYSGGYYIACAAKEIVANPLADVGSIGVIYVHFNAEKYYEMNG  203
Pfu    113  SDPSIRGVLLWIDSPGGYIGPVRAIYKEVKELAYI--KPIVAYISGYATSGGYYIACGADKIIADPYAQVGSIGVIYVHFNAQKYYEMNG  200
Mja     81  KDDSVKGVLLVVNSPGGEVIASEKLARKVEELAKK--KPVVVYVEGLDASGAYMVSAPADYIVAEKHSIVGSIGVRMDLMHYYGLMKKLG  168
Mmr     74  NNDNIKAVIIKINSPGGGAVASEKLSRKIKEVSEK--KPVVAYIENMGASAAYQAASSTNYIVAERQAVVGSIGVRMELIHYYGLMEKLG  161
Mma    177  EDNNVRAIVLRVNSGGGSPSAAQEIAIEIEKAQKQG-VPVIVSMGDLAASAAYYISAPADYIFANPSTNTGSIGVIWTFENMSGFYQNEG  265
Mac    162  ADENVKAIVLRINSPGGSPAAAQEISIEIEKAQEKG-IPVVVSMGDLAASAAYYISAPADYIYANPSTSTGSIGVIWTFENMSSYYEREG  250
Mth     69  SDPSVKAIVLDINSPGGTPVASEELMDAINKSE----KPVVSWISDSGTSGAYLAASASDRIVASPSAWVGSIGVILDLTDLSQMYRQMG  254
Tvo     29  GKKSIAGLILVFNSGGGDAASSQLIHDLVKKIRKK--KPVYSLALGICASGAYWIASASTKIYAIDTSLIGSIGVISIRPNVKKLMEKIG  116
Tac     29  NKRSVAGLVLVFNSGGGDAVASQLMFEMIRKIRKK--KPVYSFIQGICASGAYWISAGSSKIYSLDTSLIGSIGVISMIPYIKPLLDKIG  116
Pto     29  SKKKIAGLILIINSGGGDANATEILYNKLLDVSKS--KPVYALIEGIGASGAYWLACAARKIYAMSTSIVGSIGVISISPDVSDFLENLG  116
Hm1    106  SDRGSEALLLKLNTPGGEIVPSEDIRIAAEQFD----GPTIAYATDVCASGGYDIAAGCDELWAREGSIVGSIGVVGSRVNAKELADRVG  191
Neq     59  EAKDYKAIVLYIDSPGGAPEPTYRIIKYLDRIN----KTKISYIAQYGTSASYWIATHTNKIFANELSFVGSVGVLIGKIDLSGLLSKLG  143
Bsu     98  DDKTVKGIVLKVNSPGGGVYESAEIHKKLEEIKKETKKPIYVSMGSMAASGGYYISTAADKIFATPETLTGSLGVIMESVNYSKLADKLG  187

Eco    361  LDPKVKAIVLRVNSPGGSVTASEVIRAELAAARAAG-KPVVVSMGGMAASGGYWISTPANYIVANPSTLTGSIGIFGVITTVENSLDSIG  449
Hba     72  GNDSIDAVVLRVSSPGGAVSGSEVQYRAVKRLAQE--KPVVTSVRGPAASGGYYTIAPTDKIYVTPSSLVGSVGVISSVSENNGVPSR--  157
Hm3     71  SNESVEAVVLRISSPGGSAASSESMYLAVKRLSAE--KPVYTSVDQYAASGAYYTAVPSDRIYVTPASLVGHVGVIG-------TAPSDG  151
Hm2     73  TNESIEAVVLRVDSPGGPVDSSEEFYLAVNRTASE--MPVVAYVEGTAASGGYYGITPADEIVVKPSSNVGSIGVIVQAP--LSLIEQVE  158
Pa1    113  FDNKTAGVILLINSPGGAVGATERLYSTIKGLN----KTVYAVVAGLAASGAYYTAVAAGRIYATPSSWVGSIGVVALLWPDEYLID---  195
Pa2     68  QRGDVAGIVLLINSPGGTVSATEALYTTVAGIN----KTKYAVVNGLAASGAYYVAVATERIYATPSSWVGSIGVIAVMWPEEYFYD---  150

                 209 214 215  221 226 227                 250                         277
Tko    203  IKVNVFKTGKHKDMGAEWRDLTPEEREKITEMVNTYFQAFISAVSEGRNMTIDEVKNFSTGETWFAENVTG-ALVDELGGMDTAIDVLEK  291
Pab    205  IEVEVFKTGPYKDMGADWRGLTPEERDIIKNEIQTYFNDFLEVVSEGRNMTINETKKFATGRAWFAKDVNG-TLVDKLGDFDVALKELLK  293
Pho    204  IEVEVFKTGPYKDMGADWRKLTPEERKIVQTQIQTYFNDFLQVVSEGRNMTVEDVKKFATGRTWFAKDVNG-TLVDKLGDLDLALKELLK  292
Pfu    201  IEVEVFKTGPYKDMGADWRGLKPEEREIIQKQIDVYFKTFLDVVMEGRNLNETKVKEYADGRAWFAYEVNG-TLIDDIGDLQYAIEETKK  289
Mja    169  INVTTIKAGKYKDIGSPFRPMTKEEKEYLQKMINETYMDFVKWVAEHRHLSINYTLKIADGKIYSGEDAKKVGLVDEVGTEEDALKKLEQ  258
Mmr    162  INTTSITGGKYKDIGTPTRSMTEEEYAMLESMVNESYYEFVSWVAENRNMTINETLEVADGKIYSGTQAQRVGLVDITGTEDDAVDMAVK  251
Mma    266  INYYVSKSGEFKDMGGTWRGLTDEEKEYADEVVMESYEEFVTQVAEGRNISRSEVKKLADGRIYTGSRAKDLGLVDDFGNLYDAIDKAAE  355
Mac    251  VEYYISKSGEFKDMGSSWRGLTDEEKEYADSVVMESYDDFVGQVAEGRNMSRSEVKELADGRIYTGTKAKELGLVDGFGNLYDAIDKAAE  340
Mth    255  INKYAIKAGEYKDMGADYRMITDEERQMLQSMVNEEYDYFIKTVAANRNLSVSYVKSLAEGRIFTGRQAVRNRLVDFTGGKDYAVEVAAK  244
Tvo    117  VDVMVYKSGKYKDMTSPFSEPNEEEKSVYQRLLDDIFEKFKRSVAEDRGIPSEKIDEIANGMVYSAKMAADNGLIDRIANYDDLVSDLTK  206
Tac    117  VEMKIYKVGKYKDMLSSYREPSDEENEHYMRVLNDVYRKFRDSVMTERNIPEDKMEDIAQGQIFSPSMAMENRLIDGIGTMDSMMDDMYR  206
Pto    117  IKMRINKIGKYKDINSPFRHMNEDENEIFKELLNDVFLRFREEVKKRRNLTDDEINDTATGLVFSARQGLERKLIDGIGTFDTVLNAIKN  206
Hm1    192  LSYEQFTAGEYKDAGVPLKEMTADEREYLQSIVDDYYDQFIDTVAEGRE-MDAEALKDTEARIFLGDEAEERGLVDRLGTRDDVEASLEQ  280
Neq    144  VKYYSFSKGKYKEFSNPLLPLDNYTKQYYNELADKLYNFFLTDVLEHRDIKKECLSKVKESTIFLGIEAKKCGLIDYIGTMDDVKAYLEK  233
Bsu    188  ISFETIKSGAHKDIMSPSREMTKEEKNIMQSMVDNSYEGFVDVISKGRGMPKAEVKKIADGRVYDGRQAKKLNLVDELGFYDDTITAMKK  277

Eco    450  VHTDGVSTSPLADVS-ITRALPPEAQLMMQLSIENGYKRFITLVADARHSTPEQIDKIAQGHVWTGQDAKANGLVDSLGDFDDAVAKAAE  538
Hba    158  -----WKSAPDKGTTGPADKARARAATFRQSFLDVVMNERGDDLTVDR-------ETIGRAQIYAGNKAVEIGLADEIGGLDAAIADAAD  235
Hm3    152  LSPAATTGPDKAHRGMTRDQYYASLESMKRAFVGAVMAERGDRLNVSR-------ETVAEASAYQGGRAVQTGYADEVGGLEAAIAGAAG  234
Hm2    159  QQGETFVRSGPDKAQISKDSLREDIEVLQRSFVGTVMRHRGEQLTVSR-------EEVANGGTYLGAQATENGFADRIGDTGLAIERAAA  241
Pa1    196  LPDYIYTTGPFKYYGMDLTEFYNDIEKTRANFVAAVLKGRAGRLKADP-------EVFETAKIFNAETALRLGLIDKIGGLWDAVEDMAR  278
Pa2    151  IPDYIYTTGPLKYYGRELTDYYNDIERVRLNFVQAVLKGRAGRIKASP-------DVFETASIFTADQAKELGLVDEIGGIFDAVRDMAQ  233
```

Fig. 1. An amino acid sequence alignment of the core regions of archaeal and bacterial homologs of $SppA_{Tk}$ and $SppA_{Ec}$. Nineteen archaeal SppA sequences were aligned along with the sequences of SppA from *Escherichia coli* and *Bacillus subtilis*. The conserved residues selected for site-directed mutagenesis are indicated with arrowheads and numbered. Residues identical with those of $SppA_{Tk}$ are indicated in red. The abbreviations of the proteins (*italicized*) and accession numbers of all sequences used for the alignment are as follows: *B. subtilis* SppA (CAB14931, *Bsu*), *E. coli* (BAA15557, *Eco*), *H. marismortui* (I, AAV45638, *Hm1*; II, AAV46904, *Hm2*; III, AAV47811, *Hm3*), *Halobacterium* sp. NRC-1 (AAG19125, *Hba*), *Methanocaldococcus jannaschii* (AAB98642, *Mja*), *Methanococcus maripaludis* (CAF30625, *Mmr*), *Methanosarcina acetivorans* (AAM07395, *Mac*), *Methanosarcina mazei* (AAM30562, *Mma*), *Methanothermobacter thermautotrophicus* (AAB85306, *Mth*), *Nanoarchaeum equitans* (AAR39164, *Neq*), *Picrophilus torridus* (AAT42796, *Pto*), *Pyrobaculum aerophilum* (I, AAL65089, *Pa1*; II, AAL64441, *Pa2*), *Pyrococcus abyssi* (CAB49512, *Pab*), *Pyrococcus furiosus* (AAL81707, *Pfu*), *Pyrococcus horikoshii* (BAA30681, *Pho*), *T. kodakaraensis* (BAD85353, *Tko*), *Thermoplasma acidophilum* (CAC11222, *Tac*), and *Thermoplasma volcanium* (BAB59171, *Tvo*). The division of the sequences into two groups is described under "DISCUSSION."

chromatography. Although the author was able to obtain each mutant protein in a soluble form, the amount of protein produced in the *E. coli* cells varied with each mutant. As described in the previous chapter, the recombinant ΔN54SppA$_{Tk}$ was obtained in an octameric or hexadecameric form. The author also observed various quaternary structures in the mutant proteins, but only the fraction corresponding to the octameric form of the protein was used for further analysis. The apparent homogeneity of each protein after the purification procedure was examined by SDS-PAGE (Fig. 2).

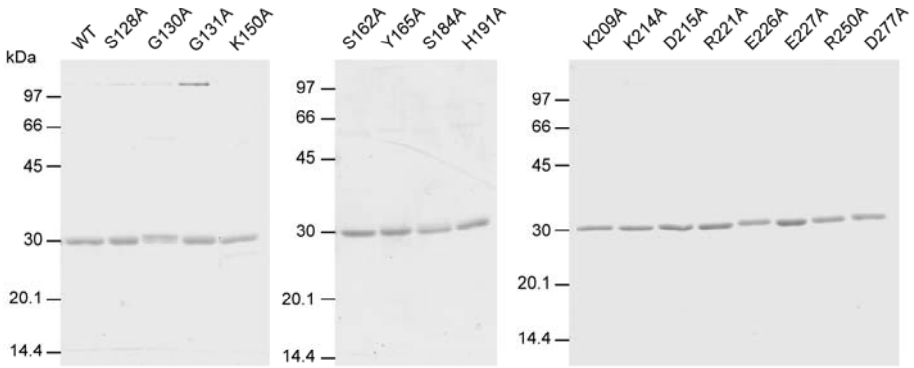


Fig. 2. SDS-PAGE of the wild-type and mutant proteins. The wild-type (WT) ΔN54SppA$_{Tk}$ and 16 mutant proteins with single amino acid replacements were subjected to SDS-PAGE after the purification procedures described under "MATERIALS AND METHODS." Mobility of molecular mass markers is indicated on the left side of each gel.

Peptidase activity of the wild-type and mutant ΔN54SppA$_{Tk}$ proteins

As described in the previous chapter, the synthetic peptide Ala-Ala-Phe-MCA is a good substrate for the wild-type ΔN54SppA$_{Tk}$. The activity levels of each mutant protein were measured at 60˚C in the presence of a fixed concentration of this substrate (200 μM) and compared with that of the wild-type enzyme (Fig. 3). The author found that the mutation of Ser162 had the most detrimental effect. No activity could be

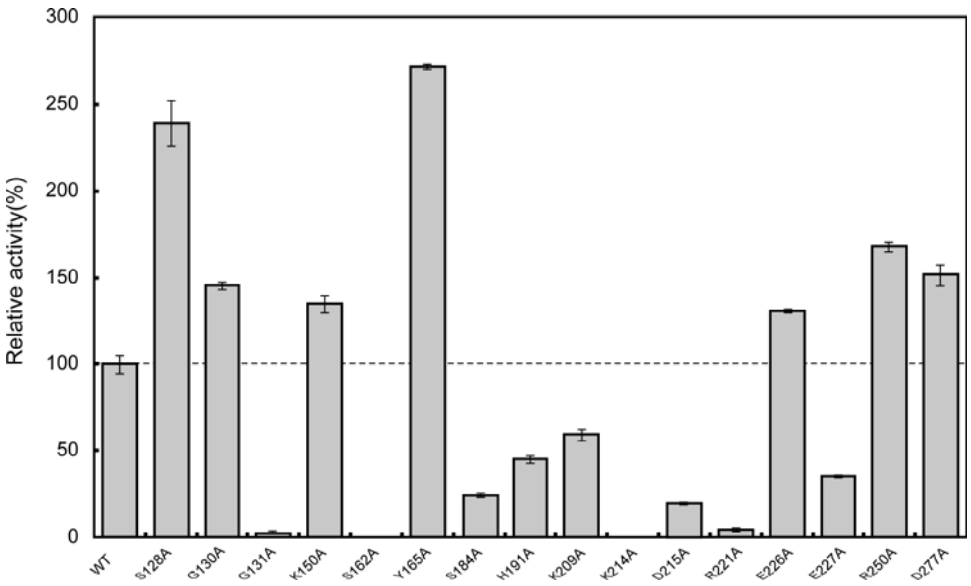


Fig. 3. Relative activity level of each mutant protein toward Ala-Ala-Phe-MCA. Substrate concentration was fixed at 200 μM, and the reaction temperature was 60˚C. The activity level of the wild-type (WT) ΔN54SppA$_{Tk}$ was designated as 100%. The error bars indicate standard deviation.

observed with the standard assay procedures. The author could estimate that the activity of the protein was only 0.004% of that of the wild-type enzyme by using a 100-fold higher protein concentration. The S184A and H191A mutations also brought about a decrease in enzyme activity, but significant levels of activity, 24 and 45% of that of the wild-type enzyme, respectively, were still observed in each mutant protein. Interestingly, the S128A, R250A, and D277A mutations did not have any negative effect but rather led to an unexpected increase in peptidase activity. It has been shown that mutations in the catalytic residues of serine proteases with a catalytic triad result in 10^4–10^6-fold reduction in activity (16). The results clearly indicate that Ser162 is essential for the peptidase activity of SppA$_{Tk}$ and is most likely the nucleophilic serine of the enzyme. The results also reveal that Arg250 and Asp277, two highly conserved residues in both archaeal and bacterial SPPs, have little, if any, role in the peptide hydrolyzing mechanism of SppA$_{Tk}$.

Mutations to identify the general base residue of $SppA_{Tk}$

The absence of conserved histidine residues among SppA proteins and the fact that His191 was not essential for the catalytic activity suggested that $SppA_{Tk}$ was not dependent on the well known Ser-His-Asp catalytic triad. The author therefore searched for other basic and acidic residues as targets for site-directed mutagenesis. Intriguingly, there were no other charged residues that were completely conserved in all of the sequences. The author therefore chose those that were the most highly conserved among the sequences, Lys150, Lys209, Lys214, Asp215, Arg221, Glu226, and Glu227. Each residue was replaced by Ala with appropriate primers (Table 1). Expression and purification were carried out as described for the initial mutant proteins, and the apparent homogeneity of each mutant protein is displayed in Fig. 2. The K150A and E226A mutations did not have a negative effect on the activity of the enzyme. All other residue replacements led to mutant proteins with lower levels of activity than that of the wild-type enzyme. Moderate effects were observed in K209A (59% activity retained), E227A (35%), and D215A (20%), whereas the effect of the R221A (4%) was much more significant. The K214A mutation had the most dramatic effect, and as in the case of the S162A mutation, activity could not be detected under standard procedures. By applying increased protein concentration, activity of K214A was estimated at 0.01% of that of the wild-type protein. The results indicate that Lys214 is essential for activity of $SppA_{Tk}$ and that the protein most likely utilizes a Ser-Lys catalytic dyad for peptide cleavage. The results also indicate that Arg221, although not essential, plays an important role in the peptidase activity of $SppA_{Tk}$.

Effects of mutations on other highly conserved residues

The author next focused on conserved, noncharged residues. Several glycine residues were highly conserved among the homologs, and in particular, a consecutive Gly-Gly sequence was conserved in all SppA proteins. In some lipases and esterases, a consecutive glycine sequence provides the main chain amino group(s) that stabilize the oxyanion intermediate that is formed after the nucleophilic attack on the carbonyl group by serine (17). The author therefore constructed the mutant proteins G130A and G131A. The author also selected a Tyr residue in the near vicinity of the nucleophilic Ser162 that is completely conserved among Spp proteins and introduced a Y165A mutation. As a result, the author observed a moderately positive effect with the G130A mutation and a surprisingly high increase in activity in the Y165A mutant protein. The G131A mutation had a drastic effect on activity, and the protein exhibited only 2% of the activity levels of the wild-type enzyme, indicating an important role for this glycine residue in the peptidase activity of $SppA_{Tk}$.

Circular dichroism spectra of wild-type and mutant ΔN54$SppA_{Tk}$ proteins

To examine whether the single-residue replacements had unintended broad effects on the protein structure, the author analyzed the CD spectrum of each mutant protein. In all mutant proteins, including the two that hardly exhibited activity (S162A and K214A), the CD spectra were indistinguishable from that of the wild-type protein (Fig. 4). The author would like to note that in the case of G130A and G131A, the author observed a slight increase in ellipticity between 225 and 230 nm. Overall, the results indicate that the single-residue replacements introduced in this study did not lead to significant changes in the secondary structures of the proteins. This allows the author to

interpret the changes in activity as direct consequences brought about by residue exchange.

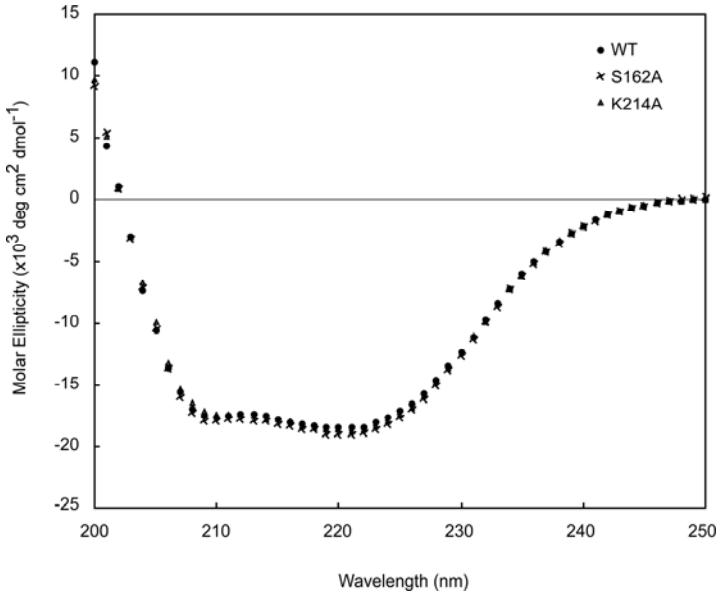


Fig. 4. Circular dichroism spectra of wild-type (WT), S162A and K214A proteins.

Kinetic analysis of the mutant proteins

The author next performed kinetic analyses on the wild-type ΔN54SppA$_{Tk}$ and all mutant proteins that exhibited sufficient levels of activity. Most mutant proteins were analyzed by standard procedures with various concentrations of Ala-Ala-Phe-MCA. The S162A and K214A mutant proteins could not be examined because of their extremely low levels of activity. Analyses of the G131A and R221A proteins were carried out with 20- and 10-fold higher amounts of enzyme, respectively. A decrease in activity levels was observed for all proteins at high substrate concentrations, indicating the occurrence of substrate inhibition. By considering several equations, the author found that the data fit very well (R^2 > 99.4) to one of the typical substrate inhibition models, expressed as $v = V_{max}[S]/(K_{s1} + [S] + 1/K_{s2}\,[S]^2)$, where v is initial velocity, V_{max}

is maximum velocity, [S] is substrate concentration, K_{s1} is the dissociation constant between enzyme and the first substrate, and K_{s2} is the dissociation constant between the enzyme-substrate complex and the second, inhibitory substrate. Representative [S]-*v* plots with the respective curves are shown for the wild-type and K150A, Y165A, and H191A mutant proteins in Fig. 5. The kinetic parameters of each protein are indicated in Table 2.

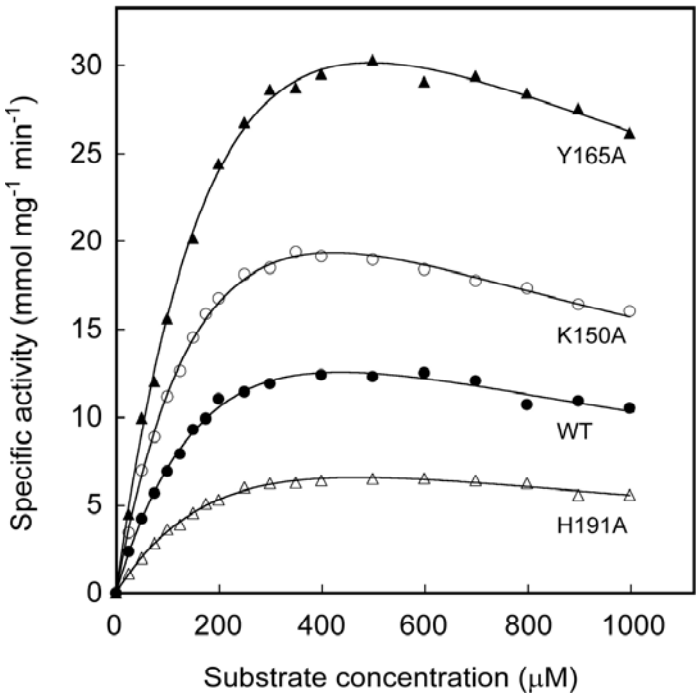


Fig. 5. Substrate concentration-initial velocity plots for wild-type (WT) ΔN54SppA$_{Tk}$ and its proteins K150A, Y165A, and H191A. The measurements were performed at 60˚C. The data for each protein were fitted with the equation $v=V_{max}[S]/(K_{s1}+[S]+1/K_{s2}[S]^2)$.

Table 2. Kinetic parameters for wild-type and mutant proteins.

Protein	V_{max} (mmol mg^{-1} min^{-1})	k_{cat} (s^{-1})	K_{s1} (μM)	k_{cat}/K_{s1}(x100) (s^{-1} μM^{-1})	K_{s2} (μM)
Wild-type	31.1 ± 2.4	16	325 ± 36	4.9	587
S128A	215 ± 23	110	1430 ± 192	7.6	897
G130A[a]	–	–	>10^4	–	–
G131A[a]	–	–	>10^6	–	–
K150A	45.9 ± 2.4	23	293 ± 22	8.0	609
Y165A	72.1 ± 5.0	37	346 ± 35	11	710
S184A	7.78 ±1.06	4.0	366 ± 69	1.1	476
H191A	16.9 ± 1.5	8.6	373 ± 45	2.3	594
K209A	18.0 ± 1.4	9.1	307 ± 34	3.0	709
D215A	17.4 ± 3.8	8.8	1190 ± 305	0.74	411
R221A	3.99 ±1.74	2.0	1760 ± 858	0.12	228
E226A	42.1 ± 3.2	21	326 ± 36	6.6	555
E227A	31.8 ± 7.0	16	1420 ± 359	1.1	359
R250A	45.4 ± 3.8	23	278 ± 34	8.3	581
D277A	49.5 ± 4.2	25	440 ± 49	5.7	438

[a]Accurate estimations of the kinetic parameters were not possible due to reasons described in the RESULTS section.

When compared with the wild-type protein, a number of mutant proteins exhibited similar K_{s1} values, indicating that these residue replacements did not greatly alter the affinity of the enzyme toward the substrate Ala-Ala-Phe-MCA. Significant increases in the K_{s1} values were observed for S128A, G130A, G131A, D215A, R221A, and E227A. Because these mutations, other than G130A and G131A, result in a decrease in the size of the side chain, the increase in K_{s1} values may be due to the

abolishment of favorable interactions between the binding pocket and substrate. The increase may also be due to the hydrophobic property of the Ala residue, which may lead to a repulsive effect on a hydrophilic region of the substrate. As for G131A, initial velocity continued to increase with increases in substrate concentration up to at least 3,000 μM, and therefore reliable curve fitting could not be performed. Fitting the equation with the limited data indicated that the K_{s1} value was higher than 106 μM. The G130A mutant also displayed K_{s1} values>10^4 μM, and reliable values for the kinetic parameters could not be obtained. In these two proteins, the additional methyl group may directly hinder the binding between enzyme and substrate. On the other hand, the glycine residues may be important in maintaining the optimized structure of the protein, because consecutive glycine residues would allow dramatic bends in the main chain backbone. As mentioned above, the author did observe subtle changes in the CD spectra of these two glycine mutant proteins. On the SDS-PAGE gels, the author also noticed a slight difference in the mobility of G130A along with the appearance of a high molecular weight band in the case of G131A, both of which are most likely due to incomplete denaturation or dissociation of the oligomeric protein (Fig. 2). Although the CD spectra rule out drastic changes in protein conformation, mutations of these two residues may result in subtle changes in the enzyme conformation. A rather unexpected result was the relatively large number of mutations that led to increases in the V_{max} or k_{cat}/K_{s1} values compared with the wild-type enzyme. Significant increases in the V_{max} value were observed in S128A (690%), K150A (148%), Y165A (232%), E226A (135%), R250A (146%), and D277A (159%). Even in the k_{cat}/K_{s1} values, which should better represent the functional capacity of the enzyme in the cell, notable increases were observed in S128A (155%), K150A (163%), Y165A (222%), E226A (135%), and

R250A (169%). The fact that these conserved residues seem to suppress the activity levels of $SppA_{Tk}$ may be important for the proper function of the enzyme *in vivo* (see DISCUSSION).

Examination of the activity of S128A and Y165A toward various substrates

Two mutant enzymes, S128A and Y165A, which displayed particularly large increases in both V_{max} and k_{cat}/K_{s1} values, were selected for examination of their activities against a FRETS peptide library as described in Chapter 1. The previous examination on the wild-type enzyme against these substrates revealed higher cleavage rates toward peptides with small amino acid residues at the Xaa position, corresponding to the P-1 position. Activity toward substrates with charged residues or aromatic residues at the Xaa position was low. When the same activity measurements were performed once again with the wild-type enzyme along with the S128A and Y165A mutant proteins, a significant increase in activity levels was observed in the mutant proteins for substrates with positively charged His, Arg, or Lys residues at the Xaa position (Fig. 6). An increase in activity levels against substrates with the aromatic residues Phe and Tyr was also observed. Even when normalizing the activity levels of each protein by designating their activities against the glycine substrate as 100, the author observed a nearly 3-fold increase in preference toward the Lys substrate in the S128A protein and a more than 2-fold increase in preference toward the Arg substrate in Y165A. Preferences toward the Tyr and Phe substrates also increased 3-fold in both enzymes. The results clearly indicate that in addition to limiting the activity levels of $SppA_{Tk}$, Ser128 and Tyr165 also play a role in restricting the substrate preference of the enzyme toward peptide substrates.

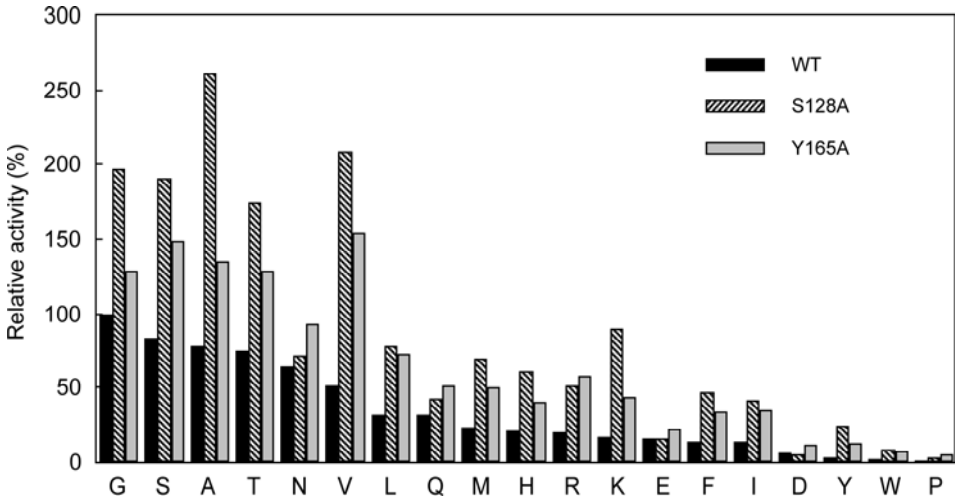


Fig. 6. Peptidase activity of wild-type ΔN54SppA$_{Tk}$ and its mutants S128A and Y165A toward various substrates from a FRETS peptide library. Amino acid residues at the Xaa position are indicated with single-letter abbreviations. Each substrate was examined at a concentration of 30 μM. The activity of the wild-type (WT) protein toward the Xaa=Gly substrate was designated as 100%.

DISCUSSION

In this chapter, the author has performed a detailed site-directed mutagenesis study on an archaeal SPP from *T. kodakaraensis*. The results have revealed multiple residues that are critical for the peptidase activity of SppA$_{Tk}$ and provide the first insight into the catalytic mechanism of prokaryotic SPPs. The analyses strongly indicate that the catalytic center of SppA$_{Tk}$ is comprised of a Ser-Lys dyad and not the Ser-His-Asp catalytic triad that is present in the majority of serine proteases.

The nucleophilic serine residue of SppA$_{Tk}$ has been clarified to be Ser162. Along with the inhibitor studies that indicated the protein was a serine peptidase (Chapter 1), activity was completely abolished in the S162A protein, whereas mutations in the other well conserved serine residues (S128A and S184A) had relatively smaller or no effects on the peptidase activity. Because Ser162 is completely conserved in all of the archaeal SppA homologs as well as in the SppA from *E. coli* and *B. subtilis*, this

serine residue most likely serves as the nucleophile in all prokaryotic SPPs. Besides the three residues examined, no other Ser residue is conserved in the core region of the enzyme, even among the closely related proteins from Thermococcales. The results also indicate that the residue acting as the general base in $SppA_{Tk}$ is Lys214. This residue is clearly conserved in the archaeal enzymes from *Pyrococcus* spp., the methanogens, *Thermoplasma* spp., *Picrophilus*, and one homolog from *Haloarcula*. The residues surrounding the Lys residues are also highly conserved among these proteins, and the sequences can be clearly aligned without any gaps from the nucleophilic Ser162 to a conserved Glu227. Although to a lower extent, the enzyme from *Nanoarchaeum equitans* also displays similarity and harbors most of these residues. It can therefore be presumed that the residue acting as the general base in these enzymes (Fig. 1, *upper group* of sequences), corresponds to the Lys214 in $SppA_{Tk}$. Although a lysine residue is found in some of the remaining homologs from the halophiles and the Crenarchaeon *Pyrobaculum* (Fig. 1, *lower group* of sequences), the relatively lower sequence similarity makes it difficult to estimate the general base in these enzymes. Interestingly, the bacterial SppA from *B. subtilis* also harbors the basic Lys residue, and surrounding sequences are particularly well conserved between the enzyme and $SppA_{Tk}$, including the charged residues Lys209, Lys214, Asp215, Arg221, Glu226, and Glu227. This strongly suggests that the bacterial SppA from *B. subtilis* also utilizes Lys214 as the general base. On the other hand, Lys214 is not conserved in $SppA_{Ec}$, and Lys residues are not found in the near vicinity. This indicates that the enzyme from *E. coli* utilizes distinct residues for catalysis.

Besides Ser162 and Lys214, which were essential for activity, the author found that the S184A, H191A, K209A, D215A, and R221A mutations led to decreases in V_{max}

or k_{cat} values. The results suggest that these residues, although not essential for activity, may also play a role in the catalytic mechanism. In particular, the mutation of Arg221 had severe effects. Because our results have clearly indicated the presence of a Ser-Lys catalytic dyad in $SppA_{Tk}$, one important question to be solved would be how the lysine residue is maintained in a deprotonated state, which would be necessary to increase the nucleophilicity of the serine side chain Oγ. The p*K*' value of the lysine side chain is 10.8, and the microenvironment surrounding the catalytic center would have to be extremely alkaline. One feasible explanation taking the results into consideration would be that the side chain of Arg221 is positioned in the very near vicinity of the Lys214 side chain. Because the p*K*' value of an arginine side chain is much higher (12.5), Arg221 may act to sequester protons that would otherwise tend to protonate Lys214. The positive charge of the protonated Arg side chain may also act to repulse other protons from approaching the Lys214 ε-amino group. Another possibility would be that Lys214 is surrounded by a neutral or hydrophobic environment, which would also result in an apparent decrease in the p*K*' value of the ε-amino group, as reported in LexA (18) and the SP from *E. coli* (16, 19).

The S184A mutation also had significant effects on the activity of $SppA_{Tk}$. From a comparison of the three-dimensional structures of the bacterial type I SP (19), Lon protease (20), LexA (18), UmuD' (21), and the γ repressor carboxy-terminal domain (22), which are all Ser-Lys proteases, possibilities of a third residue interacting with the Ser-Lys dyad have been raised (20). The residue is conserved as either a serine or threonine residue, with the side chain Oγ forming a hydrogen bond with the catalytic Lys ε-amino group. Because no other serine or threonine residues are conserved among the SPPs and because the S128A mutation led to an increase in activity, Ser184 may

represent the corresponding residue in $SppA_{Tk}$. Future studies on the three-dimensional structure of the protein should clarify the roles of Ser184 and Arg221.

As mentioned in the RESULTS section, it is intriguing that the replacements of so many highly conserved residues, usually presumed to contribute in maintaining enzyme function, actually lead to increases in peptidase activity. One possibility is that the substrate or the reaction conditions the author has applied in the activity measurements differ so much from the actual environment of the enzyme that the results do not reflect at all the actual signal peptide degrading activity of the individual proteins. However, another tempting possibility is that these residues are deliberately present to limit the functional capacity of $SppA_{Tk}$. In bacteria, $SppA_{Ec}$ is presumed to be anchored to the cell membrane with its catalytic core facing the cytoplasm. Although peptidase activity is necessary for the breakdown of free signal peptides, it should also be important for the cell that the enzyme does not cleave other proteins, whether they are soluble proteins in the cytoplasm or proteins integrated or anchored to the cell membrane. This can be achieved by a variety of strategies on the protein itself or toward its environment. A strict confinement of the substrate specificity of the enzyme would surely lead to a decrease in SPP acting on proteins other than free signal peptides. Limiting the activity levels of the enzyme would also contribute to prevent unintended protein degradation. Preventing the access of SPPs to proteins other than signal peptides would be another practical strategy. Because the author has found that the conserved Ser128 and Tyr165 not only place a limitation on activity levels but also act to restrict the substrate specificity of $SppA_{Tk}$, the presence of these residues may well be to suppress activity of the enzyme toward unintended substrates.

Although still a minority among the serine proteases, the number of enzymes

known to utilize the Ser-Lys catalytic dyad is steadily increasing. In addition to the bacterial type I SP, Lon protease, LexA, UmuD', and the λ repressor carboxy-terminal domain mentioned above, other examples include the archaeal Lon protease from *Methanocaldococcus jannaschii* (23), PH1510 from *Pyrococcus horikoshii* (24), the periplasmic tail-specific protease from *E. coli* (25), and the carboxy-terminal endoprotease from cyanobacteria (26). This study has revealed that the SPPs from *Euryarchaeota* are also members of this enzyme family. There are still no indications for a reason why certain serine proteases utilize the Ser-His-Asp catalytic triad, whereas others use the Ser-Lys dyad. Although presumed to execute identical functions, the SP from the archaeon *M. voltae* utilizes a catalytic triad (9), whereas the SP from *E. coli* harbors a dyad (19). The results in this study also raise the possibility that the SPPs from *E. coli* and *T. kodakaraensis* function through distinct catalytic mechanisms. Discovery of novel peptidases/proteases that utilize the Ser-Lys dyad and further structure-function studies on both families of enzymes will be necessary for an understanding of this intriguing distinction in catalytic mechanisms.

SUMMARY

Signal peptide peptidases (SPPs) are enzymes involved in the initial degradation of signal peptides after they are released from the precursor proteins by signal peptidases. In contrast to the eukaryotic enzymes that are aspartate peptidases, the catalytic mechanisms of prokaryotic SPPs had not been known. In this study on the SPP from the hyperthermophilic archaeon *Thermococcus kodakaraensis* ($SppA_{Tk}$), the author has identified amino acid residues that are essential for the peptidase activity of the enzyme. $\Delta N54SppA_{Tk}$, a truncated protein without the amino-terminal 54 residues

and putative transmembrane domain, exhibits high peptidase activity, and was used as the wild-type protein. Sixteen residues, highly conserved among archaeal SPP homolog sequences, were selected and replaced by alanine residues. The mutations S162A and K214A were found to abolish peptidase activity of the protein, whereas all other mutant proteins displayed activity to various extents. The results indicated the function of Ser162 as the nucleophilic serine and that of Lys214 as the general base, comprising a Ser-Lys catalytic dyad in $SppA_{Tk}$. Kinetic analyses indicated that Ser184, His191, Lys209, Asp215, and Arg221 supported peptidase activity. Intriguingly, a large number of mutations led to an increase in activity levels of the enzyme. In particular, mutations in Ser128 and Tyr165 not only increased activity levels but also broadened the substrate specificity of $SppA_{Tk}$, suggesting that these residues may be present to prevent the enzyme from cleaving unintended peptide/protein substrates in the cell. A detailed alignment of prokaryotic SPP sequences strongly suggested that the majority of archaeal enzymes, along with the bacterial enzyme from *Bacillus subtilis*, adopt the same catalytic mechanism for peptide hydrolysis.

REFERENCES

1. **Paetzel, M., Karla, A., Strynadka, N. C. J. & Dalbey, R. E.**, Signal peptidases. *Chem. Rev.*, 102, 4549-4579, 2002.
2. **Ichihara, S., Beppu, N. & Mizushima, S.**, Protease IV, a cytoplasmic membrane protein of *Escherichia coli*, has signal peptide peptidase activity. *J. Biol. Chem.*, 259, 9853-9857, 1984.
3. **Pacaud, M.**, Purification and characterization of two novel proteolytic enzymes in membranes of *Escherichia coli*. Protease IV and protease V. *J. Biol. Chem.*,

257, 4333-4339, 1982.

4. **Novak, P. & Dev, I. K.**, Degradation of a signal peptide by protease IV and oligopeptidase A. *J. Bacteriol.*, 170, 5067-5075, 1988.

5. **Bolhuis, A., Matzen, A., Hyyryläinen, H.-L., Kontinen, V. P., Meima, R., Chapuis, J., Venema, G., Bron, S., Freudl, R. & van Dijl, J. M.**, Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required for efficient translocation and processing of secretory proteins. *J. Biol. Chem.*, 274, 24585-24592, 1999.

6. **Pohlschröder, M., Dilks, K., Hand, N. J. & Wesley Rose, R.**, Translocation of proteins across archaeal cytoplasmic membranes. *FEMS Microbiol. Rev.*, 28, 3-24, 2004.

7. **Pugsley, A. P., Francetic, O., Driessen, A. J. M. & de Lorenzo, V.**, Getting out: protein traffic in prokaryotes. *Mol. Microbiol.*, 52, 3-11, 2004.

8. **Ring, G. & Eichler, J.**, Extreme secretion: protein translocation across the archael plasma membrane. *J. Bioenerg. Biomembr.*, 36, 35-45, 2004.

9. **Bardy, S. L., Ng, S. Y. M., Carnegie, D. S. & Jarrell, K. F.**, Site-directed mutagenesis analysis of amino acids critical for activity of the type I signal peptidase of the archaeon *Methanococcus voltae*. *J. Bacteriol.*, 187, 1188-1191, 2005.

10. **Ng, S. Y. M. & Jarrell, K. F.**, Cloning and characterization of archaeal type I signal peptidase from *Methanococcus voltae*. *J. Bacteriol.*, 185, 5936-5942, 2003.

11. **Bardy, S. L. & Jarrell, K. F.**, Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus*

voltae. *Mol. Microbiol.*, 50, 1339-1347, 2003.

12. **Albers, S.-V., Szabó, Z. & Driessen, A. J. M.**, Archaeal homolog of bacterial type IV prepilin signal peptidases with broad substrate specificity. *J. Bacteriol.*, 185, 3918-3925, 2003.

13. **Rawlings, N. D., Tolle, D. P. & Barrett, A. J.**, MEROPS: the peptidase database. *Nucleic Acids Res.*, 32 Database issue, D160-D164, 2004.

14. **Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B.**, Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science*, 296, 2215-2218, 2002.

15. **Wolfe, M. S. & Kopan, R.**, Intramembrane proteolysis: theme and variations. *Science*, 305, 1119-1123, 2004.

16. **Paetzel, M. & Dalbey, R. E.**, Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.*, 22, 28-31, 1997.

17. **Henke, E., Pleiss, J. & Bornscheuer, U. T.**, Activity of lipases and esterases towards tertiary alcohols: insights into structure-function relationships. *Angew. Chem. Int. Ed. Engl.*, 41, 3211-3213, 2002.

18. **Luo, Y., Pfuetzner, R. A., Mosimann, S., Paetzel, M., Frey, E. A., Cherney, M., Kim, B., Little, J. W. & Strynadka, N. C. J.**, Crystal structure of LexA: a conformational switch for regulation of self-cleavage. *Cell*, 106, 585-594, 2001.

19. **Paetzel, M., Dalbey, R. E. & Strynadka, N. C. J.**, Crystal structure of a bacterial signal peptidase in complex with a β-lactam inhibitor. *Nature*, 396, 186-190, 1998.

20. **Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A. &**

Gustchina, A., The catalytic domain of *Escherichia coli* Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J. Biol. Chem.*, 279, 8140-8148, 2004.

21. **Peat, T. S., Frank, E. G., McDonald, J. P., Levine, A. S., Woodgate, R. & Hendrickson, W. A.**, Structure of the UmuD' protein and its regulation in response to DNA damage. *Nature*, 380, 727-730, 1996.

22. **Bell, C. E., Frescura, P., Hochschild, A. & Lewis, M.**, Crystal structure of the λ repressor C-terminal domain provides a model for cooperative operator binding. *Cell*, 101, 801-811, 2000.

23. **Im, Y. J., Na, Y., Kang, G. B., Rho, S.-H., Kim, M.-K., Lee, J. H., Chung, C. H. & Eom, S. H.**, The active site of a lon protease from *Methanococcus jannaschii* distinctly differs from the canonical catalytic dyad of Lon proteases. *J. Biol. Chem.*, 279, 53451-53457, 2004.

24. **Yokoyama, H. & Matsui, I.**, A novel thermostable membrane protease forming an operon with a stomatin homolog from the hyperthermophilic archaebacterium *Pyrococcus horikoshii*. *J. Biol. Chem.*, 280, 6588-6594, 2005.

25. **Keiler, K. C. & Sauer, R. T.**, Identification of active site residues of the Tsp protease. *J. Biol. Chem.*, 270, 28864-28868, 1995.

26. **Inagaki, N., Maitra, R., Satoh, K. & Pakrasi, H. B.**, Amino acid residues that are critical for *in vivo* catalytic activity of CtpA, the carboxyl-terminal processing protease for the D1 protein of photosystem II. *J. Biol. Chem.*, 276, 30099-30105, 2001.

CHAPTER 3

Biochemical properties of a novel membrane-bound peptidase from *Thermococcus kodakaraensis*

INTRODUCTION

As described in Chapters 1 and 2, archaeal genomes do not harbor highly similar bacterial or eukaryotic signal peptide peptidase (SPP) homologs. However, the author noticed an open reading frame on the genome of *Thermococcus kodakaraensis* that, although much smaller in size (334 residues), encoded a protein ($SppA_{Tk}$) that was 27% identical to the SPP from *Escherichia coli* ($SppA_{Ec}$, 618 residues). The author examined $SppA_{Tk}$ and found that the peptidase utilized a Ser-Lys catalytic dyad for hydrolytic activity, and recognized amino acid sequences with a small side chain at the P-1 position and hydrophobic residues at the P-3 position. $SppA_{Tk}$ did not at all cleave peptide chains with an acidic residue in the vicinity of the cleavage site. These biochemical characteristics, together with the absence of highly similar bacterial or eukaryotic SPP homologs on the archaeal genomes, suggested that $SppA_{Tk}$ represents the archaeal SPP.

In bacteria, several other proteases/peptidases that are involved in signal peptide degradation have been identified. In *E. coli*, it has been found that the membrane fraction, where SPP resides, is responsible only for the initial fragmentation of signal peptides (1-3). Complete degradation to free amino acids was found to occur in the cytoplasm, and a peptidase (oligopeptidase A) involved in this process was identified and characterized (4, 5). SPP has also been studied in the gram-positive

Bacillus subtilis. This organism also harbors an additional cytosolic peptidase (TepA) whose structure is related to both bacterial SPP and ClpP. A genetic examination of TepA has shown that the protein actively participates in the degradation of signal peptides in *B. subtilis* (6).

In this Chapter, the author examined the presence of other peptidases that may be involved in signal peptide degradation in *T. kodakaraensis*. A BLAST search using the bacterial oligopeptidase A and TepA sequences against the *T. kodakaraensis* genome database revealed that there were no open reading frames with notable similarity to the bacterial sequences. However, the author identified an open reading frame (TK0130) that encoded a protein with slight similarity to $SppA_{Tk}$. As in the case of $SppA_{Tk}$, the TK0130 protein product also harbored a membrane-spanning domain in its amino terminus followed by a putative peptidase domain. In order to gain insight on the involvement of this protein ($SppB_{Tk}$) towards signal peptide degradation, the author examined the biochemical properties of $SppB_{Tk}$ and compared them with those of $SppA_{Tk}$.

MATERIALS AND METHODS

Strains, media, and plasmids

Unless otherwise mentioned, *T. kodakaraensis* KOD1 was cultivated in ASW-YT medium supplemented with elemental sulfur as described in Chapter 1. *E. coli* strains and plasmids used for DNA manipulation and gene expression are described in Chapter 1. *E. coli* strains were cultivated in LB medium with 100 μg ml^{-1} ampicillin at 37˚C.

DNA manipulation and sequence analysis

All enzymes, reagents and apparatus used for DNA purification, plasmid construction and sequence confirmation are described in the MATERIALS AND METHODS section of Chapter 1.

Expression of the *sppB$_{Tk}$* gene in *E. coli*

The *sppB$_{Tk}$* gene initiating with a Met residue replacing Phe28, omitting the transmembrane domain, was amplified from the genomic DNA of *T. kodakaraensis* using the primer set sppBN and sppBC (sppBN, 5'-CAGCTCCATATGAGGGC CCTTCAGGCGGCCAG-3'; sppBC, 5'-TAAGAATTCATGAAAGGTCACTTCTT-3'). After confirming the sequence of the amplified DNA fragment, it was inserted into pET21a(+) at the NdeI and EcoRI sites. After introduction into *E. coli* BL21-CodonPlus(DE3)-RIL cells (Stratagene), gene expression was induced with 0.1 mM isopropyl-β-D-thiogalactopyranoside at the late-exponential growth phase with further incubation for 4 h at 37˚C. The protein product was designated ΔN28SppB$_{Tk}$.

Purification of recombinant ΔN28SppB$_{Tk}$

After inducing gene expression, cells were washed with 50 mM Tris-HCl (pH 8.0) and resuspended in the same buffer. Cells were sonicated on ice, and the supernatant after centrifugation (20,000 x *g*, 30 min at 4˚C) was applied to heat treatment at 80˚C for 10 min, immediately cooled on ice, and then centrifuged (20,000 x *g*, 30 min at 4˚C). The soluble protein sample was brought to 40-70% saturation with $(NH_4)_2SO_4$ and the precipitate, which included ΔN28SppB$_{Tk}$, was dissolved in 50 mM Tris-HCl (pH 8.0). After desalting with a HiPrep26/10 column (GE Healthcare), the

sample was brought to 1.5 M $(NH_4)_2SO_4$ and applied to hydrophobic chromatography (ResourcePHE, GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 1.5 M $(NH_4)_2SO_4$, and proteins were eluted with a linear gradient (1.5 to 0 M) of $(NH_4)_2SO_4$. After desalting with a HiPrep26/10 column (GE Healthcare), the sample was applied to anion exchange chromatography (ResourceQ, GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 0.15 M NaCl, and proteins were eluted with a linear gradient (0.15 to 0.5 M) of NaCl. The sample was applied to gel filtration chromatography (Superdex 200 HR 10/30, GE Healthcare) equilibrated with 50 mM Tris-HCl (pH 8.0), 0.15 M NaCl, and the fractions obtained were used for enzyme analysis.

Protein analysis of purified recombinant ΔN28SppB$_{Tk}$

The molecular mass of the purified protein was examined by gel-filtration chromatography using Superdex 200 HR 10/30 in 50 mM Tris-HCl (pH 8.0), 0.15 M NaCl. The retention time was calibrated with those of the standard proteins ferritin (440 kDa), aldolase (158 kDa), conalbumin (75 kDa), ovalbumin (43 kDa), carbonic anhydrase (29 kDa), ribonuclease A (13.7 kDa) and aprotinin (6.5 kDa). Protein concentration was determined with the protein assay kit (Bio-Rad) using bovine serum albumin as a standard.

Western blot analysis

T. kodakaraensis KOD1 was cultivated in ASW-YT medium supplemented with 1% sodium pyruvate. Cells grown at 85˚C for 12 h were harvested, washed with 0.1 M potassium phosphate buffer (pH 7.0) containing 0.4 M NaCl, and then disrupted by French press (96 MPa). The resulting cell extract was ultracentrifuged (110,000 x *g*,

2 h at 4˚C) to separate the cytosol and membrane fractions. Each fraction (32 μg of protein) was subjected to SDS-PAGE and followed by Western blot analyses with specific antisera against the purified ΔN28SppB$_{Tk}$, glutamate dehydrogenase (7) and Lon protease (8) of *T. kodakaraensis*, respectively. A protein A-peroxidase conjugate was used for visualization together with 4-chloro-1-naphthol and hydrogen peroxide.

Enzyme activity measurements

Most activity measurements were performed with peptidyl-MCA substrates [peptidyl–α-(4-methylcoumaryl-7-amide) substrates] available from Peptide Institute. Release of 7-amino-4-methylcoumarin was monitored consecutively with a fluorescence spectrophotometer capable of maintaining the cuvette at desired temperatures between 30 and 100˚C. Excitation and emission wavelengths were 380 nm and 460 nm, respectively. Standard activity measurements were performed at 50˚C in a final volume of 1 ml with 0.5 μg of purified protein and *Z*-Val-Lys-Met-MCA (50 μM) in 50 mM CHES (pH 10.0). The final concentration of dimethyl sulfoxide used to dissolve the substrate was constant at 10% of the reaction mixture.

Effects of temperature and pH on enzyme activity

All buffers were prepared so that they would reflect accurate values at the applied temperatures. In examining the effect of temperature, the standard assay method was applied at each temperature. The effect of pH was examined in the presence of 50 mM of MES-NaOH (pH 6.0 to 7.0), HEPES-NaOH (pH 7.0 to 8.0), Bicine-NaOH (pH 8.0 to 9.0), CHES-NaOH (pH 9.0 to 10.0), and CAPS-NaOH (pH 10.0 to 12.0), respectively. Thermostability of the protein was analyzed by measuring the residual

activity of the protein after incubation at various temperatures in 50 mM CHES-NaOH (pH 10.0). The initial activity of the enzyme incubated at each respective temperature was designated as 100%. Alkaline stability was analyzed by measuring the residual activity of the protein after incubation at various pH in 50 mM CAPS-NaOH at 50˚C. The initial activity of the enzyme incubated in each buffer was designated as 100%. In measuring thermostability and alkaline stability, the protein concentration during incubation was 1 μg ml^{-1}. Residual activities were measured with the standard assay method described above.

Kinetic analysis

Kinetic analysis of the $\Delta N28SppB_{Tk}$ reaction was carried out with *Z*-Val-Lys-Met-MCA in 50 mM MES-NaOH (pH 7.0) at 50˚C. Regardless of the substrate concentration, the final concentration of dimethyl sulfoxide used to dissolve the substrate was constant at 10% of the reaction mixture.

Effect of protease inhibitors

The effects of various protease inhibitors at concentrations of 2 μM, 20 μM, 200 μM, 1 mM, 5 mM or 10 mM were examined at 50˚C and pH 10.0. The substrate *Z*-Val-Lys-Met-MCA was present at a concentration of 50 μM. Activity in the absence of inhibitors was defined as 100%.

Determination of substrate specificity

Substrate specificity was examined with a FRETS peptide library (25Xaa series, Peptide Institute) (9) as described in Chapter 1. In the initial assay to examine the

preference for residues at the P-1 position, 0.4 μg of purified enzyme was added to the reaction mixture with a final volume of 1 ml containing 30 μM substrate in 50 mM CHES (pH 10.0). The final concentration of dimethyl sulfoxide used to dissolve the substrate was constant at 3% of the reaction mixture. A second assay to identify the cleavage sites and the preference towards residues at the P-2 to P-4 positions was performed on selected substrates. Aliquots (100 μl) from the cleavage reactions were taken at various time intervals that corresponded to 15 to 25% cleavage of the substrates, and subjected to liquid chromatography (LC)-mass spectrometry analysis. An ODS A-302 column (YMC) was used for separation with 0.05% trifluoroacetic acid in H_2O as eluant A and 0.05% trifluoroacetic acid in CH_3CN as eluant B. The gradient was 5–40% of eluant B in A at a flow rate of 1.0 ml min^{-1} over a time span of 55 min. Aliquots taken from the cleavage reactions were injected and the cleaved products were monitored with absorbance at 220 nm, as well as fluorescence intensity in order to identify the amino-terminal segments. The structures of the cleaved products were deduced from the theoretical molecular weights.

RESULTS

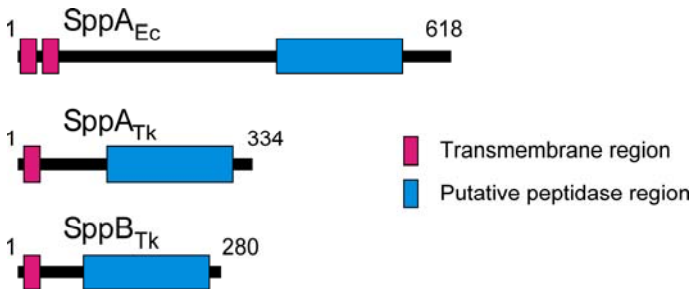
A homolog search for bacterial signal peptide degrading peptidases on the *T. kodakaraensis* genome

Oligopeptidase A and TepA are the cytosolic peptidases involved in signal peptide degradation in *E. coli* and *B. subtilis*, respectively. The two enzymes have been suggested to contribute in the complete breakdown of signal peptides to free amino acids (4-6). The substrate specificity of $SppA_{Tk}$ suggested that the enzyme is involved only in the partial degradation of signal peptides into smaller peptide fragments

(Chapter 1). One would thus expect the presence of other peptidases that exhibit complementary substrate specificities or exoproteolytic activity in order to achieve complete breakdown. Using the primary structure of the oligopeptidase A of *E. coli* and TepA of *B. subtilis*, the author performed a BLAST search against the protein sequences of *T. kodakaraensis*. Intriguingly, no open reading frames with notable similarity to either of the proteins were found. The author then performed a BLAST search on the *T. kodakaraensis* genome with the primary structure of $SppA_{Tk}$. This led to the identification of an open reading frame (TK0130) that encoded a protein with a putative membrane-spanning domain, followed by a putative peptidase domain, similar to the architecture of $SppA_{Tk}$ (Fig. 1). The protein, designated here as $SppB_{Tk}$, was 280 residues in length with a deduced molecular mass of 31,519 Da and was 18% identical to $SppA_{Tk}$. Due to its low similarity with the SPP sequence of *E. coli*, the protein cannot be identified in a BLAST search using the primary structure of $SppA_{Ec}$.

A BLAST search against the complete genome sequences of various archaeal and bacterial strains was performed with the $SppB_{Tk}$ sequence. As in the case of $SppA_{Tk}$, highly related homologs of $SppB_{Tk}$ were present on a number of archaeal genomes, including those of the three *Pyrococcus* strains, *Archaeoglobus fulgidus*, *Methanocaldococcus jannaschii* and many other methanogens. Although homologs were absent in most Crenarchaeota genomes, one was present on the genome of *Aeropyrum pernix*. Moreover, $SppB_{Tk}$ homologs with high similarity were present on the genomes of a range of bacteria, from the (hyper)thermophilic *Aquifex aeolicus*, *Thermotoga maritima*, and *Thermus thermophilus* to the green-nonsulfur *Chloroflexus*, the α-proteobacterium *Bradyrhizobium* and cyanobacteria, such as *Synechococcus* and

A



B

```
                                                                M
SppA   MDDRIWKYVSAVLTLLLALSIVSVAVLYYQVNPPAVPANQTGFVIETPSNFTLVCEGSVN
       ::        :    :
SppB   MDPLSGFLGSLLWWLFFLYLLMWPQLQFRALQAARARLMAQLARKRNSTVIAMIHR----
                                  M

SppA   TTSTEVAELRDQIAFLQRLVSSNQGGTTIAVVPIFGIIDDYTALQVIPLLRNLAMNESVG
                   :                  :    :        :    :
SppB   ---------QESIGLFG--------------IPVYKFISIEDSEEVLRAIRSAPKDKPID

                                                 ▼
SppA   GVLLWIESPGGVVGPVINIHSEIKKLSLVKPVVAYSGDIIASGGYYIAVGAQKIVASPLA
          : :  ::: :     :    :        :         :::  ::  : ::   : :
SppB   ---LIIHTPGGLVLAATQIARALKEHPAETRVIVPH--YAMSGGTLIALAADKIIMDPNA

                                          ▼
SppA   EVGSIGVIYVHYDLEKNYEMNGIKVNVFKTGKHKDMGAEWRDLTPEEREKITEMVNTYFQ
         :         :          :   : : :                  ::    :
SppB   VLGPVDPQLGQYPAPS------ILRAVEKKG--PEKVDDQTLILADVAEKAIKQVQDFVF

SppA   AFISAVSEGRNMTIDEVKNFSTGETWFAENVTGALVDELGGMDTAIDVLEKLMNVSGAKV
                 :             :        :     :::      :: :
SppB   SLLKDKY-GEEKARELAQILTEGRWTHDYPITVDHARELG-LNVETDVPEEVYALMELYK

SppA   VVYKDLETPEEFGVYGSTALYIDPRYLTPLIGGG-
          :                :
SppB   QPVK----------QRGTVEFMPYPVKQESKK---
```

Fig. 1. Structural comparison among SppA$_{Ec}$, SppA$_{Tk}$ and SppB$_{Tk}$. (A) A diagram illustrating the basic structural features and sizes of the three proteins. Transmembrane regions and peptidase regions are indicated with pink and blue boxes, respectively. (B) A sequence alignment between SppA$_{Tk}$ and SppB$_{Tk}$. Identical residues are indicated with colons. The letters M in red indicate the positions at which an artificial methionine residue was incorporated to produce ΔN54SppA$_{Tk}$ and ΔN28SppB$_{Tk}$. Red arrowheads above the SppA$_{Tk}$ sequence indicate the catalytic residues of the enzyme elucidated in Chapter 2.

Nostoc species. In the MEROPS database (the peptidase database, http://merops.sanger.ac.uk/) (10), the $SppB_{Tk}$ homologs are classified as the unassigned peptidases of the S49 family.

Expression of the *$sppB_{Tk}$* gene in *E. coli* and purification of the recombinant protein

The putative membrane-spanning domain of $SppB_{Tk}$ corresponds to residues Ser5 to Gln27 (Fig. 1B). As in the case of $SppA_{Tk}$, the author omitted the membrane-spanning region when constructing the expression vector. An artificial Met residue was incorporated in the place of Phe28, and the gene was expressed in *E. coli*. A soluble protein, designated as $\Delta N28SppB_{Tk}$, which was resistant to heat treatment at 80˚C for 10 min, was obtained. The author found that the thermostable protein exhibited peptidase activity toward the peptide substrate *Z*-Val-Lys-Met-MCA (see below). The recombinant protein was purified with ammonium sulfate fractionation, hydrophobic chromatography, anion exchange chromatography, and gel filtration chromatography. SDS-PAGE analysis of the sample after gel filtration chromatography indicated that the recombinant $\Delta N28SppB_{Tk}$ was purified to apparent homogeneity. The molecular mass of the purified protein estimated by SDS-PAGE was in good agreement with the calculated value of the amino acid sequence (Fig. 2).

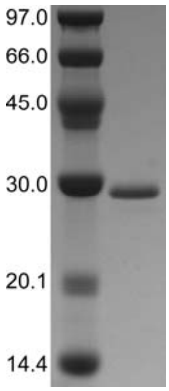


Fig. 2. SDS-PAGE analysis of purified $\Delta N28SppB_{Tk}$. The gel was stained with Coomassie Brilliant Blue.

Oligomeric form of ΔN28SppB$_{Tk}$

The molecular mass of the purified ΔN28SppB$_{Tk}$ was estimated by gel filtration chromatography. This indicated a molecular mass of 350 kDa. The molecular mass of a single subunit of ΔN28SppB$_{Tk}$ is 28,205 Da, suggesting that the protein formed a dodecamer. In contrast to SppA$_{Tk}$, no other peaks corresponding to other oligomeric states could be observed.

Intracellular localization of SppB$_{Tk}$

As the author was able to purify the recombinant enzyme, polyclonal antibodies were raised against the protein and used to examine the subcellular localization of SppB$_{Tk}$. Wild-type *T. kodakaraensis* cells were disrupted by French press, and extracts were separated by ultracentrifugation into the soluble fraction (cytosolic fraction) and pellet (membrane fraction). Proper separation was confirmed by performing Western blot analyses (Fig. 3) with antibodies against glutamate dehydrogenase (cytosolic localization) and Lon protease (membrane-bound) (8). Analysis with the SppB$_{Tk}$ antibodies revealed signal ratios similar to Lon protease, confirming that SppB$_{Tk}$, most likely by means of its amino-terminal membrane-spanning domain, is physically associated with the cytoplasmic membrane of *T. kodakaraensis*.

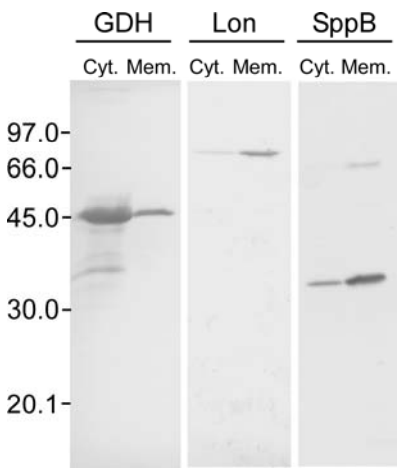


Fig. 3. Western blot analyses of glutamate dehydrogenase (GDH), Lon protease (Lon) and SppB$_{Tk}$ (SppB) with the cytosol (Cyt.) and membrane (Mem.) fractions of *T. kodakaraensis* cells.

Proteolytic activity towards various MCA substrates

The author could not detect peptidase activity of ΔN28SppB$_{Tk}$ towards Ala-Ala-Phe-MCA, a substrate that served as a good substrate for SppA$_{Tk}$. The author examined a large number of MCA substrates, listed in Table 1, for their potential to be recognized as substrates of ΔN28SppB$_{Tk}$. As a result, the author found that *Z*-Val-Lys-Met-MCA was the substrate which led to the highest levels of proteolytic activity. *Z*-Leu-Leu-Leu-MCA was also well recognized by ΔN28SppB$_{Tk}$, with a relative activity of 61%. All other substrates were cleaved at very low levels compared to *Z*-Val-Lys-Met-MCA and *Z*-Leu-Leu-Leu-MCA. As the latter substrate had a tendency to aggregate, *Z*-Val-Lys-Met-MCA was used for standard activity measurements in all further experiments.

Table 1. MCA substrates examined in this study.

Substrate	Relative activity (%)
Z-Val-Lys-Met-MCA	100
Z-Leu-Leu-Leu-MCA	61
Z-Leu-Leu-Glu-MCA	1.36
Suc-Ala-Ala-Ala-MCA	0.53
Z-Ala-Ala-Asn-MCA	0.31
Z-Leu-Arg-Gly-Gly-MCA	0.21
Ala-Ala-Phe-MCA	0.13
Z-Val-Val-Arg-MCA	n.d.
Suc-Ile-Ile-Trp-MCA	n.d.
Suc(OMe)-Ala-Ala-Pro-Val-MCA	n.d.
Suc-Ala-Ala-Pro-Phe-MCA	n.d.
Suc-Arg-Pro-Phe-His-Leu-Leu-Val-Tyr-MCA	n.d.
Boc-Arg-Val-Arg-Arg-MCA	n.d.
Boc-Val-Leu-Lys-MCA	n.d.
Ac-Val-Asp-Val-Ala-Asp-MCA	n.d.
Ac-Ile-Glu-Thr-Asp-MCA	n.d.

Activities were examined with a substrate concentration of 50 μM in 50 mM MES-NaOH (pH 7.0) at 50˚C.

n.d., activity not detected.

Optimal pH and temperature

The effects of pH and temperature on the activity of ΔN28SppB$_{Tk}$ were examined with *Z*-Val-Lys-Met-MCA. As shown in Fig. 4A, the protein exhibited maximal activity at pH 10.0. High levels of activity were maintained at even higher pH values of 10.5 (80%), 11.0 (70%) and 11.5 (52%). The effects of temperature on activity were analyzed with the same substrate at pH 10.0. Maximum activity under the conditions examined was observed at approximately 75˚C (Fig. 4B). The Arrhenius plot gave two constant slopes from 30˚C to 55˚C, and from 60˚C to 75˚C, corresponding to activation energies of 79.7 kJ mol$^{-1,}$ and 14.8 kJ mol^{-1}, respectively (Fig. 4C). This indicates that the enzyme takes an optimal catalytic structure at temperatures between 60˚C and 75˚C. The thermostability of the enzyme was examined for 2 h at pH 10.0 (Fig. 4D), and alkaline stability was measured at 50˚C for 4 h (Fig. 4E). The enzyme retained 26% activity after 2 h incubation at 70˚C and pH 10.0. The stability against temperature and alkaline pH were much lower than those observed for SppA$_{Tk}$.

Kinetic analysis of ΔN28SppB$_{Tk}$

The author performed a kinetic analysis of ΔN28SppB$_{Tk}$ using the substrate *Z*-Val-Lys-Met-MCA. Although optimal reaction conditions (70˚C, pH 10.0) were preferred, aggregation/precipitation of the substrate was observed at high substrate concentrations. By examining several reaction conditions, the author found that substrate solubility was sufficient for kinetic analysis at 50˚C, pH 7.0. Under these conditions, the author obtained a substrate concentration [S]-initial velocity (v) plot as shown in Fig. 5A. The kinetics did not at all follow Michaelis-Menten kinetics, and negative cooperativity with respect to substrate binding was apparent. A Hill plot of the

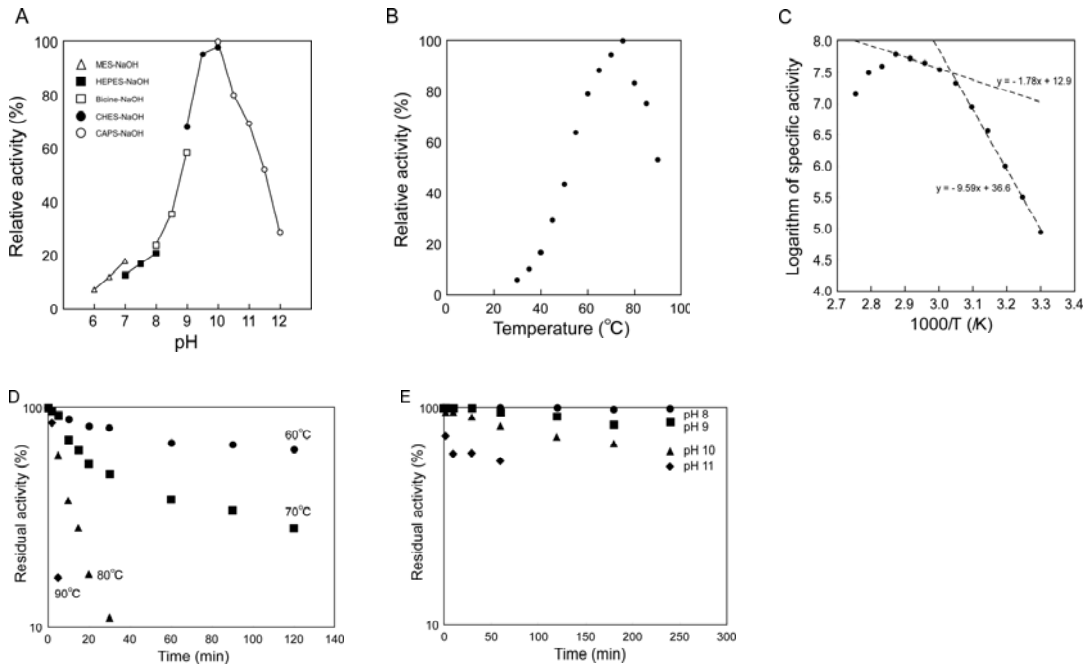


Fig. 4. (A) Effect of pH on the activity of ΔN28SppB$_{Tk}$. Measurements were performed at 50˚C in the following buffers at a concentration of 50 mM: MES-NaOH (open triangles), HEPES-NaOH (solid squares), Bicine-NaOH (open squares), CHES-NaOH (solid circles), and CAPS-NaOH (open circles). (B) Effect of temperature on the activity of ΔN28SppB$_{Tk}$. Reactions were carried out in 50 mM CHES-NaOH (pH 10.0). (C) Arrhenius plot of B, indicating the activation energy of substrate hydrolysis catalyzed by ΔN28SppB$_{Tk}$. (D) Thermostability of ΔN28SppB$_{Tk}$ at various temperatures. Incubation of the enzyme was carried out in 50 mM CHES-NaOH (pH 10.0). Symbols: 60˚C, circles; 70˚C, squares; 80˚C, triangles; 90˚C, diamonds. (E) Stability of ΔN28SppB$_{Tk}$ at various pHs. Enzyme incubation was carried out at 60˚C in 50 mM CAPS-NaOH (pH 8.0, circles; pH 9.0, squares, pH 10.0, triangles, pH 11.0, diamonds). All activity measurements (A to E) were carried out with 50 μM *Z*-Val-Lys-Met-MCA.

kinetic data revealed linear relationships between log[S] and log{$v/(V_{max}-v)$} (V_{max}, maximum velocity of the enzyme) with different slopes (apparent Hill constants, n_H) depending on the substrate concentration range (Fig. 5B). As expected, the apparent Hill constant was near 1 (n_H=1.1) at low substrate concentrations (10–50 μM), suggesting that only one active site of the enzyme is occupied by a substrate. In contrast, at concentrations between 100 and 170 μM, n_H was dramatically higher at 5.7, indicating that the enzyme possesses at least 6 substrate binding sites with strong cooperativity. As

gel filtration of ΔN28SppB$_{Tk}$ suggested a dodecameric oligomerization, the enzyme may well exhibit its allosteric properties via the interaction within hexameric units. The V_{max} value of the reaction was 593 μmole min^{-1} mg^{-1}, and the $[S]_{0.5}$ value calculated from the Hill plot was 88 μM.

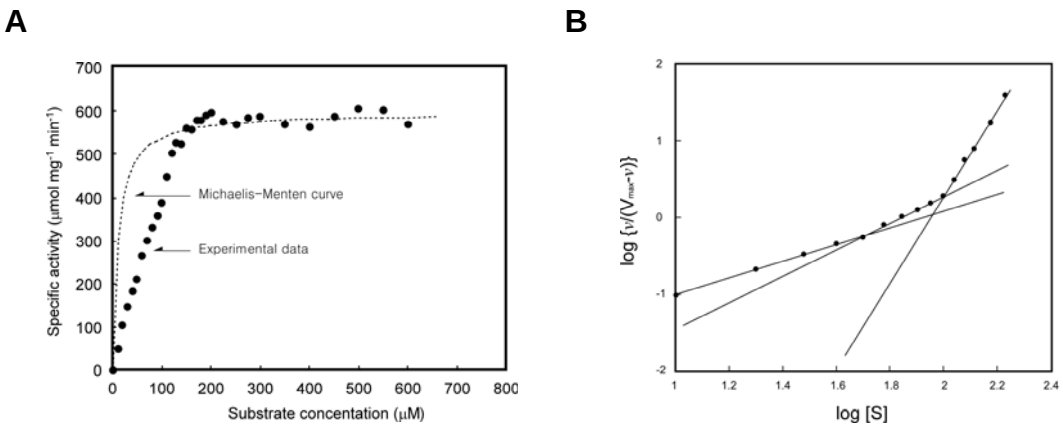
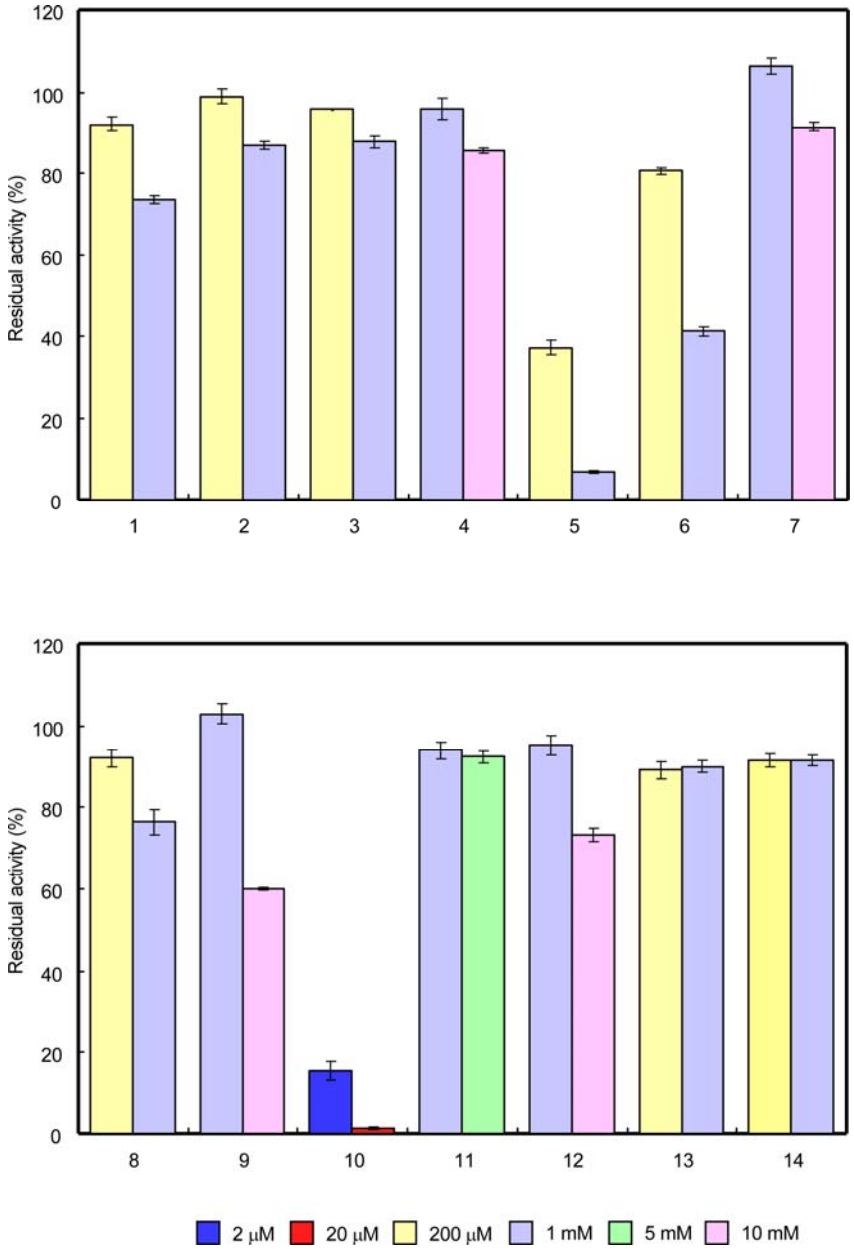


Fig. 5. (A) Substrate concentration-initial velocity plots for ΔN28SppB$_{Tk}$ using substrate Z-Val-Lys-Met-MCA. (B) A Hill plot of the kinetic data. Lines are fitted to the three ranges in substrate concentration in which log [S] and log {v/(V_{max}-v)} displayed a linear correlation.

Effects of various inhibitors

The effects of various protease inhibitors were examined at 50˚C and pH 10.0 (Fig. 6). Interestingly, many of the conventional protease inhibitors did not have notable effects on ΔN28SppB$_{Tk}$ activity. Chymostatin and pepstatin A exhibited relatively moderate levels of inhibition. Intriguingly, we found that (Z-Leu-Leu-NH-CH_2)$_2$-CO, an inhibitor specific to the eukaryotic SPP, displayed strong inhibition (11-13). Activity decreased 99% in the presence of 20 μM inhibitor, while an 84% decrease was observed at 2 μM.



Inhibitors

1) Leupeptin	Trypsin/Papain inhibitor
2) Antipain	Trypsin/Papain inhibitor
3) Elastatinal	Elastase inhibitor
4) EDTA	Metal protease inhibitor
5) Chymostatin	Chymotrypsin/Papain inhibitor
6) Pepstatin A	Pepsin/Cathepsin inhibitor
7) Phenylmethanesulfonyl fluoride	Serine protease inhibitor
8) Diisopropyl fluorophosphate	Serine protease inhibitor
9) Phebestin	Aminopeptidase N inhibitor
10) (Z-Leu-Leu-$NHCH_2)_2CO$	Signal peptide peptidase inhibitor
11) E-64	Thiol protease inhibitor
12) Arphamenine A	Aminopeptidase B inhibitor
13) Leuhistin	Aminopeptidase M inhibitor
14) Phosphoramidon	Thermolysin inhibitor

Fig. 6. Effects of various protease inhibitors on the activity of $\Delta N28SppB_{Tk}$.

Examining the substrate preference of ΔN28SppB$_{Tk}$ with a FRETS peptide library

The author next evaluated the peptidase activity of ΔN28SppB$_{Tk}$ against a FRETS peptide library, which was applied in the analysis of SppA$_{Tk}$. An initial analysis was carried out with 19 substrates corresponding to all amino acids at the Xaa site with the exception of cysteine (Fig. 7A). The author detected a preference of ΔN28SppB$_{Tk}$ for substrates with neutral or hydrophobic residues (Ala, Met, Thr, Ser, Leu, Val and Ile) at the Xaa position (Fig. 7B). However, compared to the results obtained with SppA$_{Tk}$, the differences in activity towards the 19 substrates seemed to be much smaller, suggesting a broader substrate specificity. The author next performed LC-mass spectrometry analysis on the cleaved products of ten selected substrates (Xaa = Ala, Met, Thr, Ser, Leu, Val, Ile, Phe, Lys and Gly). In order to identify products generated from the initial cleavage reaction of the substrate, reactions were stopped at various intervals corresponding to 15 to 25% substrate cleavage. The most abundant cleavage products for each of the ten substrates are shown in Fig. 7C. The results were striking; in almost all cases, the most preferred mode of cleavage was with a basic residue at the P-2 position. When Arg (Zaa) was positioned at the P-2 position, the P-1 position was occupied exclusively by Ile (Yaa), indicating that Pro/Tyr/Lys/Asp were not preferred. The degradation products of Xaa=Lys confirm that Lys is not accepted at the P-1 position. The results of Xaa=Phe, along with the fact that Tyr is not preferred at the P-1 position suggest that ΔN28SppB$_{Tk}$ does not cleave peptides with an aromatic residue at the P-1 position. Degradation products with Glu or Asp at any one of the sites from P-2 to P-4 were not observed, and considering that the Xaa=Glu/Asp substrates were the least cleaved, the results indicate that ΔN28SppB$_{Tk}$ does not recognize sequences with acidic residues. In summary, ΔN28SppB$_{Tk}$ strongly prefers substrates with a basic

residue at the P-2 position, and to a lesser extent, aromatic residues. At the P-1 position, hydrophobic, but not aromatic residues, are favored. The enzyme does not cleave substrates with acidic residues.

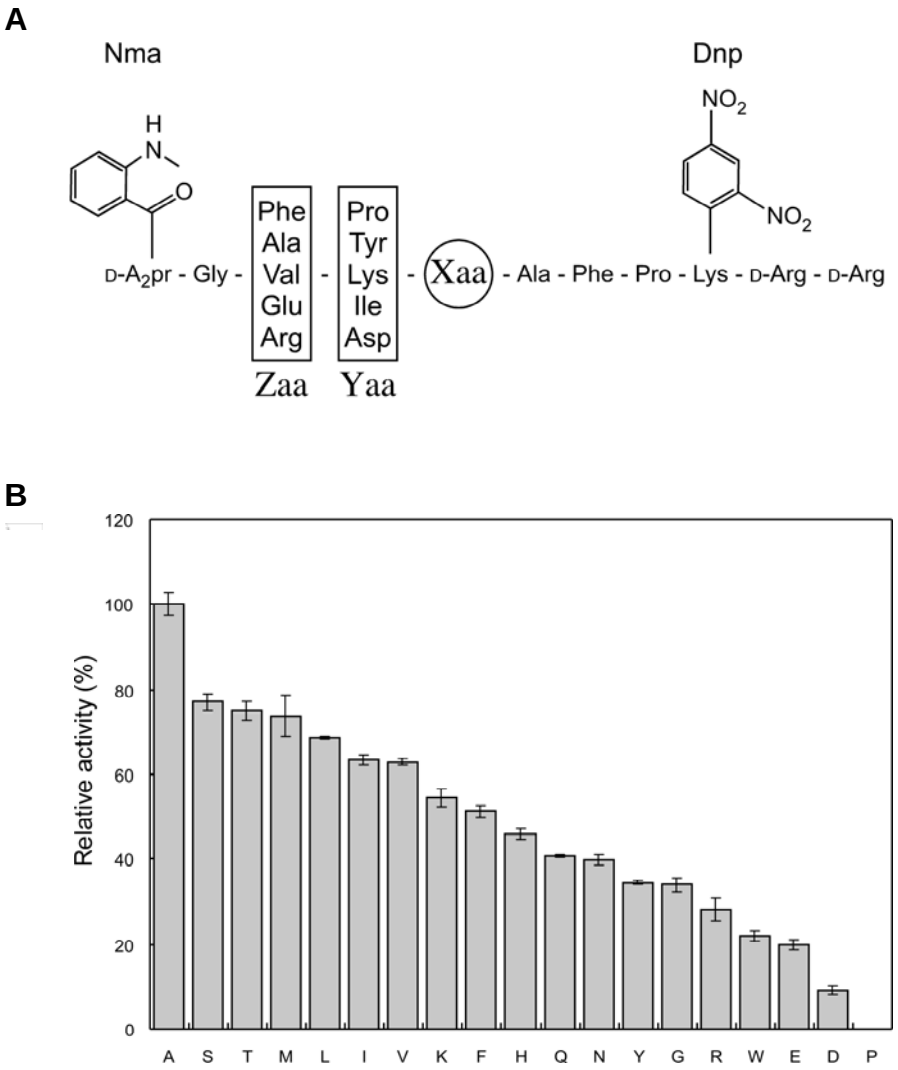


Fig. 7. (A) Structure of the peptide substrates from a FRETS peptide library utilized in B and C. (B) Peptidase activity towards various substrates from the FRETS peptide library. Amino acid residues at the Xaa position are indicated. Each substrate was examined at a concentration of 30 μM. (C) The major cleaved products of ten substrates (Xaa = Ala, Ser, Thr, Met, Leu, Ile, Val, Lys, Phe and Gly) were detected by LC-mass spectrometry. Relative quantities are indicated to the right of the sequences.

C

Xaa = Ala

P4	P3	P2	P1	P1'	
-	G	R	I ▼	A	100%
G	V	K	A ▼	A	94%
G	A	K	A ▼	A	58%
G	F	K	A ▼	A	54%
G	V	Y	A ▼	A	51%
G	A	Y	A ▼	A	46%

Xaa = Met

P4	P3	P2	P1	P1'	
-	G	R	I ▼	M	100%
G	V	K	M ▼	A	60%
-	G	F	I ▼	M	43%
G	A	K	M ▼	A	43%
G	F	K	M ▼	A	40%
G	A	Y	M ▼	A	38%

Xaa = Thr

P4	P3	P2	P1	P1'	
G	V	K	T ▼	A	100%
R	I	T	A ▼	F	88%
-	G	R	I ▼	T	81%
F	I	T	A ▼	F	74%
V	I	T	A ▼	F	67%
A	I	T	A ▼	F	64%

Xaa = Ser

P4	P3	P2	P1	P1'	
-	G	R	I ▼	S	100%
G	V	K	S ▼	A	48%
-	G	F	I ▼	S	41%
G	V	Y	S ▼	A	34%

Xaa = Leu

P4	P3	P2	P1	P1'	
-	G	R	I ▼	L	100%
-	-	G	V ▼	X	69%
G	F	K	L ▼	A	37%
G	A	K	L ▼	A	35%
G	V	K	L ▼	A	32%
G	A	Y	L ▼	A	31%

Xaa = Val

P4	P3	P2	P1	P1'	
-	-	G	V ▼	X	100%
-	G	R	I ▼	V	88%
G	V	K	V ▼	A	64%
G	A	K	V ▼	A	60%
G	F	K	V ▼	A	53%
G	A	Y	V ▼	A	44%

Xaa = Ile

P4	P3	P2	P1	P1'	
-	G	R	I ▼	I	100%
-	-	G	V ▼	X	97%
G	V	K	I ▼	A	83%
G	A	K	I ▼	A	82%
R	I	I	A ▼	F	49%
-	-	G	A ▼	X	48%

Xaa = Phe

P4	P3	P2	P1	P1'	
-	G	R	I ▼	F	100%
-	-	G	V ▼	X	75%
F	K	F	A ▼	F	75%
R	I	F	A ▼	F	66%
A	I	F	A ▼	F	50%
-	G	V	I ▼	F	48%

Xaa = Lys

P4	P3	P2	P1	P1'	
-	G	R	I ▼	K	100%
-	G	F	I ▼	K	81%
-	G	V	I ▼	K	63%
-	G	A	I ▼	K	57%
-	-	G	V ▼	X	44%
A	I	K	A ▼	F	32%

Xaa = Gly

P4	P3	P2	P1	P1'	
-	G	R	I ▼	G	100%
G	V	K	G ▼	A	72%
G	V	Y	G ▼	A	47%
-	G	F	I ▼	G	41%
G	A	Y	G ▼	A	36%
-	G	V	I ▼	G	32%

DISCUSSION

In this chapter, the author has identified and characterized the catalytic domain of a second membrane-bound peptidase, $SppB_{Tk}$, from the hyperthermophilic archaeon *T. kodakaraensis*. Closely related homologs are present on the genomes of a range of archaea and bacteria. However, the distribution of the $SppB_{Tk}$ homologs does not provide strong clues as to the function of the protein. One reasonable assumption is that as the $SppB_{Tk}$ homologs are present in a number of the chemoautotrophic methanogens and in the photoautotrophic cyanobacteria, it is not likely that they play a major role in the breakdown and metabolism of extracellular peptides for amino acid assimilation. They can thus be expected to act on proteins/peptides that are generated within the cell.

In order to gain insight on the function of $SppB_{Tk}$, the author has focused on the substrate specificity of the enzyme and compared its properties with those of $SppA_{Tk}$. Presuming that the substrate specificities of the recombinant $\Delta N54SppA_{Tk}$ and $\Delta N28SppB_{Tk}$ reflect those of the native $SppA_{Tk}$ and $SppB_{Tk}$, respectively, the results of Chapter 1 and this Chapter indicate that the two peptidases exhibit distinct preferences in substrate recognition. Although both peptidases do not cleave sequences in the near vicinity of acidic residues, $SppA_{Tk}$ prefers hydrophobic/aromatic residues at the P-3 site and residues with relatively small side chains at the P-1 site. $SppB_{Tk}$ prefers basic (and hydrophobic/aromatic) residues at the P-2 site and aliphatic hydrophobic residues at the P-1 site. As shown for $SppA_{Tk}$ in Chapter 1, the author examined whether $SppB_{Tk}$ was able to cleave substrates longer than the MCA and FRETS substrates used for activity analyses. As shown in Fig. 8, incubation of $\Delta N28SppB_{Tk}$ with various protein substrates revealed that the enzyme harbored hydrolytic activity towards protein substrates.

The archaeal signal peptide shares a common architecture to those found in bacteria and eukaryotes, with an n-region, h-region and c-region (14, 15). As described in Chapter 1, the substrate specificity of $SppA_{Tk}$ suggests that the cleavage sites of this enzyme mostly reside in the h-region, the core region of the signal peptide abundant in hydrophobic residues. In contrast, if $SppB_{Tk}$ were to function in signal peptide degradation, the enzyme would be responsible for cleavage of the n-region, which harbors positively charged residues. As $SppB_{Tk}$ also recognizes hydrophobic/aromatic residues at the P-2 site, the enzyme may also play a role in assisting $SppA_{Tk}$ in the cleavage of the h-region. Although further genetic and *in vivo* studies will be necessary to identify the actual substrates of these peptidases and verify their physiological roles, the substrate specificities of the two membrane-bound peptidases can be expected to

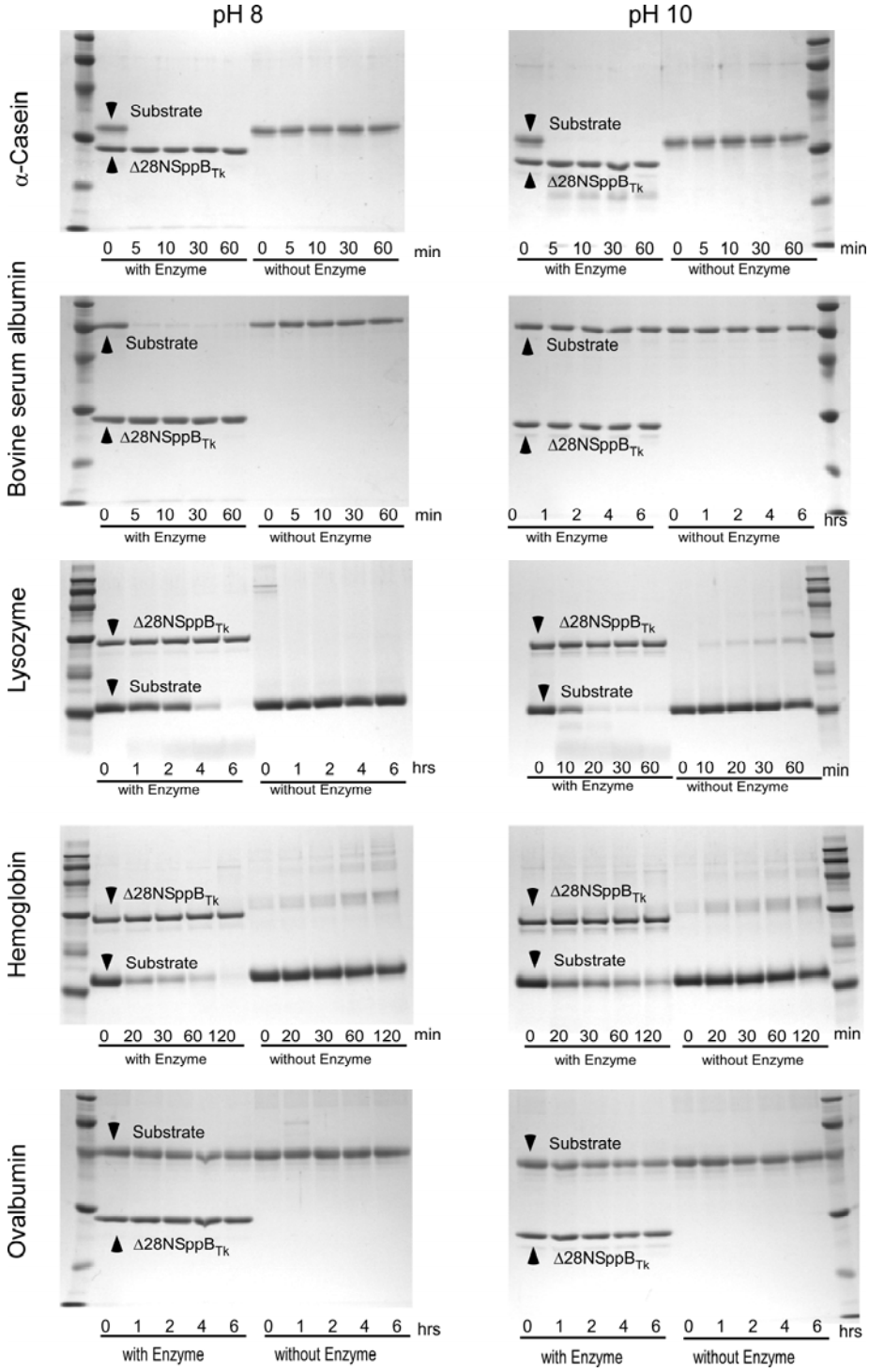


Fig. 8. Incubation was performed at 60˚C in a final volume of 1 ml with 1 μg of ΔN28SppB$_{Tk}$ and protein substrate (1 μg or 3 μg) in 50 mM CHES-NaOH (pH 10.0). After incubation, substrate degradation in the reaction mixture was confirmed by SDS-PAGE (12.5% or 15% concentration of acrylamide). As a control reaction, incubation was also performed without ΔN28SppB$_{Tk}$ in the reaction mixture.

efficiently complement one another in the initial breakdown of signal peptides at the cytoplasmic membrane (Fig. 9).

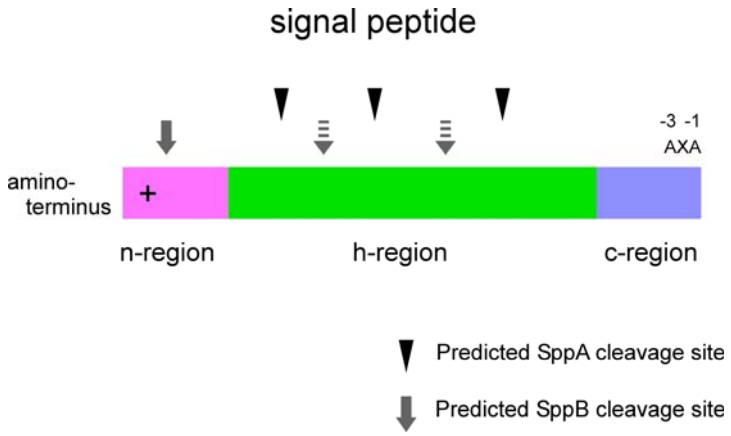


Fig. 9. A schematic diagram illustrating the proposed cleavage sites of $SppA_{Tk}$ and $SppB_{Tk}$ on the released signal peptide.

SUMMARY

In Chapters 1 and 2, the author identified and characterized a strong candidate for the signal peptide peptidase in the *Archaea* ($SppA_{Tk}$) from the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. In a search for additional enzymes involved in signal peptide degradation, the author realized that enzymes structurally related to the cytosolic peptidases functioning in signal peptide degradation in the *Bacteria* were not present in the *Archaea*. Instead, a gene encoding a putative membrane-bound peptidase with limited similarity to $SppA_{Tk}$ was found on the *T. kodakaraensis* genome. As in the case of $SppA_{Tk}$, the protein ($SppB_{Tk}$) harbored a putative membrane-spanning domain in its amino-terminal region. The catalytic domain of $SppB_{Tk}$ without its

membrane-spanning domain (ΔN28SppB$_{Tk}$) was produced in *Escherichia coli* and purified. The protein displayed a dodecameric oligomerization, and exhibited peptidase activity towards the substrate *Z*-Val-Lys-Met-MCA and Z-Leu-Leu-Leu-MCA. Western blot analysis using antisera raised against the purified recombinant protein confirmed that SppB$_{Tk}$ is associated with the cytoplasmic membrane in *T. kodakaraensis*. Analyses with a FRETS peptide library revealed that ΔN28SppB$_{Tk}$ displays a strong preference towards basic amino acids at the P-2 site and hydrophobic (but not aromatic) residues at the P-1 position. This is in contrast to ΔN54SppA$_{Tk}$, which preferred hydrophobic residues at the P-3 position and residues with relatively small side chains at the P-1 position. As in the case of ΔN54SppA$_{Tk}$, ΔN28SppB$_{Tk}$ was capable of cleaving proteins, indicating that its activity was not limited to oligopeptide degradation. As archaeal signal peptides are comprised from an n-domain (with basic residues) and an h-domain (with hydrophobic residues), the substrate specificities of SppA$_{Tk}$ and SppB$_{Tk}$ can be expected to efficiently complement one another in the initial breakdown of signal peptides at the cytoplasmic membrane.

REFERENCES

1. **Hussain, M., Ichihara, S. & Mizushima, S.**, Mechanism of signal peptide cleavage in the biosynthesis of the major lipoprotein of the *Escherichia coli* outer membrane. *J. Biol. Chem.*, 257, 5177-5182, 1982.
2. **Ichihara, S., Beppu, N. & Mizushima, S.**, Protease IV, a cytoplasmic membrane protein of *Escherichia coli*, has signal peptide peptidase activity. *J. Biol. Chem.*, 259, 9853-9857, 1984.
3. **Pacaud, M.**, Purification and characterization of two novel proteolytic enzymes

in membranes of *Escherichia coli*. Protease IV and protease V. *J. Biol. Chem.*, 257, 4333-4339, 1982.

4. **Novak, P. & Dev, I. K.**, Degradation of a signal peptide by protease IV and oligopeptidase A. *J. Bacteriol.*, 170, 5067-5075, 1988.

5. **Novak, P., Ray, P. H. & Dev, I. K.**, Localization and purification of two enzymes from *Escherichia coli* capable of hydrolyzing a signal peptide. *J. Biol. Chem.*, 261, 420-427, 1986.

6. **Bolhuis, A., Matzen, A., Hyyryläinen, H.-L., Kontinen, V. P., Meima, R., Chapuis, J., Venema, G., Bron, S., Freudl, R. & van Dijl, J. M.**, Signal peptide peptidase- and ClpP-like proteins of *Bacillus subtilis* required for efficient translocation and processing of secretory proteins. *J. Biol. Chem.*, 274, 24585-24592, 1999.

7. **Rahman, R. N. Z. A., Fujiwara, S., Takagi, M. & Imanaka, T.**, Sequence analysis of glutamate dehydrogenase (GDH) from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1 and comparison of the enzymatic characteristics of native and recombinant GDHs. *Mol. Gen. Genet.*, 257, 338-347, 1998.

8. **Fukui, T., Eguchi, T., Atomi, H. & Imanaka, T.**, A membrane-bound archaeal Lon protease displays ATP-independent proteolytic activity towards unfolded proteins and ATP-dependent activity for folded proteins. *J. Bacteriol.*, 184, 3689-3698, 2002.

9. **Tanskul, S., Oda, K., Oyama, H., Noparatnaraporn, N., Tsunemi, M. & Takada, K.**, Substrate specificity of alkaline serine proteinase isolated from photosynthetic bacterium, *Rubrivivax gelatinosus* KDDS1. *Biochem. Biophys. Res. Commun.*, 309, 547-551, 2003.

10. **Rawlings, N. D., Tolle, D. P. & Barrett, A. J.**, MEROPS: the peptidase database. *Nucleic Acids Res.*, 32 Database issue, D160-D164, 2004.

11. **Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B.**, Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science*, 296, 2215-2218, 2002.

12. **Weihofen, A., Lemberg, M. K., Friedmann, E., Rueeger, H., Schmitz, A., Paganetti, P., Rovelli, G. & Martoglio, B.**, Targeting presenilin-type aspartic protease signal peptide peptidase with γ-secretase inhibitors. *J. Biol. Chem.*, 278, 16528-16533, 2003.

13. **Weihofen, A., Lemberg, M. K., Ploegh, H. L., Bogyo, M. & Martoglio, B.**, Release of signal peptide fragments into the cytosol requires cleavage in the transmembrane region by a protease activity that is specifically blocked by a novel cysteine protease inhibitor. *J. Biol. Chem.*, 275, 30951-30956, 2000.

14. **Paetzel, M., Karla, A., Strynadka, N. C. J. & Dalbey, R. E.**, Signal peptidases. *Chem. Rev.*, 102, 4549-4579, 2002.

15. **Pohlschröder, M., Dilks, K., Hand, N. J. & Wesley Rose, R.**, Translocation of proteins across archaeal cytoplasmic membranes. *FEMS Microbiol. Rev.*, 28, 3-24, 2004.

CHAPTER 4

Identification of the amino acid residues essential for proteolytic activity of the membrane-bound peptidase $SppB_{Tk}$ from *Thermococcus kodakaraensis*

INTRODUCTION

As described in Chapter 3, the author identified a novel membrane-bound peptidase in *Thermococcus kodakaraensis* distinct to the previously characterized $SppA_{Tk}$. The structural and biochemical properties of $SppA_{Tk}$ supported a role of the protein as a signal peptide peptidase (SPP) in *T. kodakaraensis*, initiating the degradation of signal peptides after their release from the precursor form of proteins to be translocated to the cell surface or extracellular matrix. As in the case of $SppA_{Tk}$, $SppB_{Tk}$ also harbors a membrane-spanning domain, followed by a putative peptidase domain, which was biochemically shown to exhibit peptidase activity in the previous chapter. The substrate specificity of $SppB_{Tk}$ raised the possibilities that the protein cooperates with $SppA_{Tk}$ in the signal peptide degradation process in which the former recognizes sequences near the positively charged n-region of signal peptides, whereas the latter mainly cleaves the hydrophobic h-region (Chapter 3, Fig. 9).

The author has mentioned in Chapter 3 that $SppB_{Tk}$ homologs are classified as unassigned serine peptidases of the S49 family. Although they are found in a large number of both archaea and bacteria, none of the proteins besides $SppB_{Tk}$ has been characterized. Moreover, the $SppB_{Tk}$ homologs display a remarkably high similarity with one another; the homolog from the cyanobacterium *Synechococcus* sp. JA-2-3B'a(2-13) is 55.7% identical in primary structure with $SppB_{Tk}$. The residues that

constitute the catalytic center of these enzymes have not been determined. An alignment with representative $SppA_{Tk}$ and $SppB_{Tk}$ homologs indicates that the nucleophilic serine of $SppA_{Tk}$ (Ser162) aligns with a serine residue of $SppB_{Tk}$ (Ser130) that is conserved among $SppB_{Tk}$ homologs (Fig. 1). However, the counterpart of the catalytic base (Lys214) in $SppA_{Tk}$ cannot be found or predicted in $SppB_{Tk}$. In this chapter, the author carried out a detailed site-directed mutagenesis study in order to determine the amino acid residues that take part in the catalysis of $SppB_{Tk}$ and its homologs, which comprise a large family of archaeal/bacterial peptidases.

```
                                                  ▼
APho   TIKGVLVWIESPGGFVGPVREIYIELKKLDYIKPVIAYVSGYAYSGGYYIACAAKEIVAN
APab   TIKGVLLWIESPGGYVGPVREIYNELKKLGYLKPIVAYVSGYAYSGAYYIACAAREIIAE
APfu   SIRGVLLWIDSPGGYIGPVRAIYKEVKELAYIKPIVAYISGYATSGGYYIACGADKIIAD
ATko   SVGGVLLWIESPGGVVGPVINIHSEIKKLSLVKPVVAYSGDIIASGGYYIAVGAQKIVAS
               *  ***       *    *                **   **  *  *
BTko   PID---LIIHTPGGLVLAATQIARALKEHPAETRVIVPH--YAMSGGTLIALAADKIIMD
BPab   PID---LIIHTPGGLVLAATQIAKALKDHPAETRVIVPH--YAMSGGTLIALAADKIIMD
BPho   PID---LIIHTPGGLVLAATQIAKALKDHPAETRVIVPH--YAMSGGTLIALAADKIIMD
BPfu   PID---LIIHTPGGLVLAATQIAKALKDHPAETRVIVPH--YAMSGGTLIALAADKIIMD

                                          ▼
APho   PLADVGSIGVIYVHFNAEKYYEMNGIEVEVFKTGPYKDMGADWRKLTPEERKIVQTQIQT
APab   PLSEVGSIGVIYVHFNAEEYYKMNGIEVEVFKTGPYKDMGADWRGLTPEERDIIKNEIQT
APfu   PYAQVGSIGVIYVHFNAQKYYEMNGIEVEVFKTGPYKDMGADWRGLKPEEREIIQKQIDV
ATko   PLAEVGSIGVIYVHYDLEKNYEMNGIKVNVFKTGKHKDMGAEWRDLTPEEREKITEMVNT
       *    *                   *   *   *
BTko   PNAVLGPVDPQLGQYPAPS------ILRAVEKKG--PEKVDDQTLILADVAEKAIKQVQD
BPab   PHAVLGPVDPQLGQYPAPS------IIRAVERKG--PDKVDDQTLILADVAEKAIKQVRD
BPho   PHAVLGPVDPQLGQYPAPS------IVRAVEKKG--VDKVDDQTLILADVAEKAIRQVRD
BPfu   PHAVLGPVDPQLGQYPAPS------IIKAVEQKG--AEKVDDQTLILADVAKKAIKQVQD

APho   YFNDFLQVVSEGRNMTVEDVKKFATGRTWFAKDVNGTLVDKLGDLDLALKELLKLIGAKK
APab   YFNDFLEVVSEGRNMTINETKKFATGRAWFAKDVNGTLVDKLGDFDVALKELLKLIGAKK
APfu   YFKTFLDVVMEGRNLNETKVKEYADGRAWFAYEVNGTLIDDIGDLQYAIEETKKLANLKS
ATko   YFQAFISAVSEGRNMTIDEVKNFSTGETWFAENVTGALVDELGGMDTAIDVLEKLMNVSG
                  *            *               *
BTko   FVFSLLKDKY-GEEKARELAQILTEGRWTHDYPITVDHARELG-LNVETDVPEEVYALME
BPab   FIYDLLKDKY-GDEKAKELAKILTEGRWTHDYPITVEEARKLG-LNVSTDVPEEVYALME
BPho   FIFNLLKDKY-GEEKAKELAQVLTEGRWTHDYPITVEEAKKLG-LNVSTDVPEEVYALME
BPfu   FLYDLLKDKY-GEEKARELAQILTEGRWTHDYPITVEHARELG-LEVDTNVPEEVYALME
```

Fig. 1. An alignment of $SppA_{Tk}$ and $SppB_{Tk}$ homolog sequences of the Thermococcales. Residues conserved among the eight sequences are indicated with asterisks. The two residues essential for activity of $SppA_{Tk}$ (Chapter 2) are indicated with arrowheads. Sequences aligned are those of the $SppA_{Tk}$ and $SppB_{Tk}$ from *T. kodakaraensis* (ATko, BTko), homologs from *Pyrococcus horikoshii* (APho, BPho), *Pyrococcus abyssi* (APab, BPab), and *Pyrococcus furiosus* (APfu, BPfu).

MATERIALS AND METHODS

Strains, media, and plasmids

Escherichia coli strains used for plasmid construction, amplification and gene expression are described in Chapter 2. In order to obtain plasmid DNA sensitive to DpnI digestion, *E. coli* JM109 was used. *E. coli* strains were cultivated in LB medium as described in Chapter 1.

DNA manipulation, sequence analysis, and site-directed mutagenesis

Procedures for the isolation and purification of plasmid DNA, as well as DNA sequencing, are described in Chapter 1. Sequence comparisons and alignments were performed with the ClustalW program provided by the DNA Data Bank of Japan. Site-directed mutagenesis was carried out using a QuikChange XL site-directed kit as described in Chapter 2. The template used was an expression plasmid constructed for wild-type ΔN28SppB$_{Tk}$, a protein with an amino-terminal truncation of 27 amino acid residues. An artificial Met residue was inserted in the position of the native Phe28 (Chapter 3). The primers used to incorporate each mutation are shown in Table 1. After sequence confirmation, the plasmids were introduced into *E. coli* BL21-CodonPlus(DE3)-RIL cells for gene expression.

Expression and purification of wild-type and mutant ΔN28SppB$_{Tk}$ proteins

The recombinant *E. coli* cells harboring expression plasmids for wild-type and mutant ΔN28SppB$_{Tk}$ proteins were grown in LB medium supplemented with 0.01% glucose, and gene expression was induced with 0.1 mM isopropyl-β-D-thiogalacto-pyranoside at the late exponential growth phase. After a further 4 h incubation,

Table 1. Primers used for site-directed mutagenesis.

Mutation	Primer sequences
H55A	5′-ACTGTAATAGCAATGATCGCCAGGCAGGAGAGCATAGGA-3′
	5′-TCCTATGCTCTCCTGCCTGGCGATCATTGCTATTACAGT-3′
R56A	5′-GTAATAGCAATGATCCACGCCCAGGAGAGCATAGGACTC-3′
	5′-GAGTCCTATGCTCTCCTGGGCGTGGATCATTGCTATTAC-3′
E58A	5′-GCAATGATCCACAGGCAGGCCAGCATAGGACTCTTTGGG-3′
	5′-CCCAAAGCGTCCTATGCTGGCCTGCCTGTGGATCATTGC-3′
D75A	5′-AAGTTCATAAGCATCGAGGCCAGCGAGGAAGTGCTCAGA-3′
	5′-TCTGAGCACTTCCTCGCTGGCCTCGATGCTTATGAACTT-3′
E77A	5′-ATAAGCATCGAGGACAGCGCCGAAGTGCTCAGAGCAATC-3′
	5′-GATTGCTCTGAGCACTTCGGCGCTGTCCTCGATGCTTAT-3′
R81A	5′-GACAGCGAGGAAGTGCTCGCCGCAATCAGGAGCGCACCA-3′
	5′-TGGTGCGCTCCTGATTGCGGCGAGCACTTCCTCGCTGTC-3′
H97A	5′-CCGATAGACCTGATCATCGCCACGCCTGGTGGACTAGTT-3′
	5′-AACTAGTCCACCAGGCGTGGCGATGATCAGGTCTATCGG-3′
H116A	5′-GCGAGGGCGCTCAAGGAGGCCCCGGCTGAGACGCGCGTC-3′
	5′-GACGCGCGTCTCAGCCGGGGCCTCCTTGAGCGCCCTCGC-3′
S130A	5′-GTCCCACACTACGCCATGGCCGGCGGCACACTCATAGCA-3′
	5′-TGCTATGAGTGTGCCGCCGGCCATGGCGTAGTGTGGGAC-3′
T133A	5′-TACGCCATGAGCGGCGGCGCCCTCATAGCACTCGCCGCT-3′
	5′-AGCGGCGAGTGCTATGAGGGCGCCGCCGCTCATGGCGTA-3′
D140A	5′-CTCATAGCACTCGCCGCTGCCAAGATTATCATGGATCCA-3′
	5′-TGGATCCATGATAATCTTGGCAGCGGCGAGTGCTATGAG-3′
D145A	5′-GCTGACAAGATTATCATGGCCCCAAACGCAGTCCTCGGC-3′
	5′-GCCGAGGACTGCGTTTGGGGCCATGATAATCTTGTCAGC-3′
D154A	5′-GCAGTCCTCGGCCCAGTTGCCCCACAGCTTGGTCAGTAC-3′
	5′-GTACTGACCAAGCTGTGGGGCAACTGGGCCGAGGACTGC-3′
S164A	5′-GGTCAGTACCCAGCTCCGGCCATACTAAGGGCTGTCGAG-3′
	5′-CTCGACAGCCCTTAGTATGGCCGGAGCTGGGTACTGACC-3′
D179A	5′-GGCCCGGAGAAGGTTGACGCCCAGACCTTAATCCTGGCC-3′
	5′-GGCCAGGATTAAGGTCTGGGCGTCAACCTTCTCCGGGCC-3′
D186A	5′-CAGACCTTAATCCTGGCCGCCGTTGCTGAAAAGGCCATA-3′
	5′-TATGGCCTTTTCAGCAACGGCGGCCAGGATTAAGGTCTG-3′
T225A	5′-CTCACGGAGGGCAGATGGGCCCACGACTACCCGATAACC-3′
	5′-GGTTATCGGGTAGTCGTGGGCCCATCTGCCCTCCGTGAG-3′
H226A	5′-TGGACGGCCGACTACCCGATAACCGTTGACCACGCCAGG-3′
	5′-CCTGGCGTGGTCAACGGTTATCGGGTAGTCGGCCGTCCA-3′

cells were collected and washed with 50 mM Tris-HCl (pH 8.0), and resuspended in the same buffer. Methods for cell disruption and protein purification are the same as those described for the wild-type ΔN28SppB$_{Tk}$ protein in Chapter 3.

Enzyme activity measurements

Standard activity measurements were performed at 50˚C in 1 ml of 50 mM CHES (pH 10.0) with 0.5 μg of purified protein and 50 μM *Z*-Val-Lys-Met-MCA as described in Chapter 3. The final concentration of dimethyl sulfoxide used to dissolve the substrate was constant at 10% of the reaction mixture.

Circular dichroism spectroscopy of wild-type and mutant enzymes

Each protein sample was prepared in 25 mM Tris-HCl (pH 8.0), 75 mM NaCl at a protein concentration of 0.1 mg ml^{-1}. A J-820 spectropolarimeter (Jasco) was used to measure ellipticity as a function of wavelength from 250 to 200 nm in 0.2-nm increments using a 0.1-cm cylindrical quartz cuvette. The samples were scanned fifty times and averaged. The mean molar ellipticity $[\theta]$ (deg cm^2 dmol^{-1}) was calculated from the equation $[\theta] = \theta / 10\ nCl$, where θ is the measured ellipticity in millidegrees, C is the molar concentration of enzyme subunits, l is the path length in centimeters, and n is the number of residues per subunit.

RESULTS

Distribution of SppB$_{Tk}$ homologs in the *Archaea* and *Bacteria*

In order to determine which residues should be selected for site-directed mutagenesis and subsequent biochemical analyses, the author aligned all archaeal and

bacterial $SppB_{Tk}$ homologs found in the National Center for Biotechnology Information (NCBI) database. In the *Archaea*, highly related homologs were found in 24 of the 53 genomes. Among the *Crenarchaeota*, only *Aeropyrum pernix* and *Hyperthermus butyricus*, both members of the Desulfurococcales, harbored an $SppB_{Tk}$ homolog. Many members of the *Euryarchaeota* harbored $SppB_{Tk}$ homologs, but were absent on the genomes of the Thermoplasmatales and haloarchaea. In addition, closely related homologs were present on a variety of bacterial genomes. A phylogenetic analysis of all archaeal and bacterial $SppB_{Tk}$ homolog sequences is shown in Fig. 2, indicating that the archaeal and bacterial sequences, with only a few exceptions, can be phylogenetically distinguished.

Sequence alignment of archaeal and bacterial $SppB_{Tk}$ homologs

The sequences of the archaeal and bacterial $SppB_{Tk}$ homologs were aligned, and the result is shown in Fig. 3. Reflecting their high structural relationship, a remarkable number of residues (>25) were conserved among the 53 sequences. As was indicated in Fig. 1, sequences in the vicinity of the catalytic Ser162 of $SppA_{Tk}$ aligned well with the sequence near Ser130 of $SppB_{Tk}$. Ser130 was one of the amino acid residues that was completely conserved among all $SppB_{Tk}$ homologs (Fig. 3), making it an attractive candidate for the nucleophile that initiates peptide bond hydrolysis. In addition to this residue, a second serine residue (Ser164) was also conserved among all homologs, and was thus also selected as a target of site-directed mutagenesis. Completely conserved residues that were candidates for the general base were His55, Arg56, His97, Lys190 and His226. Conserved acidic residues that are components of the Ser-His-Asp(Glu) catalytic triad of many serine proteases are Glu58, Asp75,

Asp154 and Asp179. These nine residues were also selected for site-directed mutagenesis. Plasmids were designed and constructed to express the eleven mutant proteins with each selected residue individually replaced with an alanine residue. Although the possibilities were considered low, the author presumed that the activity levels of the mutant proteins of the latter nine residues would also provide valuable information regarding the catalytic mechanism of $SppB_{Tk}$ even if the enzyme were not a serine protease. The selected residues would cover the essential residues of aspartic proteases (Asp/Glu) and also amino acid residues commonly found in the metalloproteases (His/Glu). In addition to the eleven residues, the author constructed a number of proteins with mutations in other highly, but not completely, conserved residues such as Glu77, Arg81, His116, Thr133, Asp140, Asp145, Asp186, and Thr225. In all cases, single amino acids were replaced with an alanine residue.

Production and purification of mutant $SppB_{Tk}$ proteins via site-directed mutagenesis

The mutations described above were incorporated into the expression vector for $\Delta N28SppB_{Tk}$. Sequence analysis confirmed that only the intended mutations were introduced into the genes. Recombinant mutant proteins were produced in *E. coli* and purified from the cell-free extracts by heat treatment at 80˚C for 10 min, followed by ammonium sulfate fractionation, hydrophobic interaction chromatography, anion exchange chromatography, and gel filtration chromatography. Although the author was able to obtain each mutant protein in a soluble form, the amount of protein produced in the *E. coli* cells varied with each mutant. The apparent homogeneity of each protein after the purification procedure was examined by SDS-PAGE.

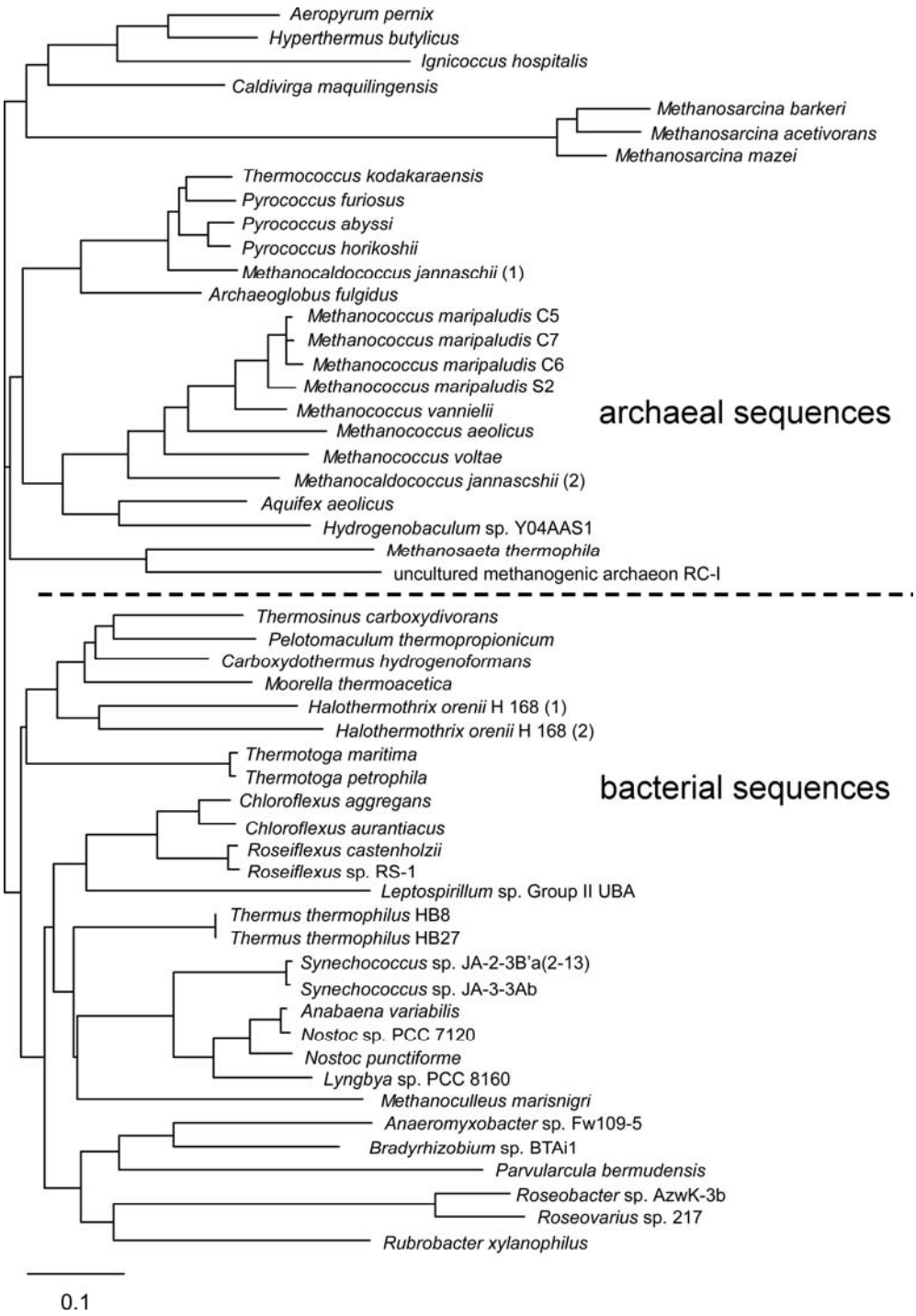


Fig. 2. A phylogenetic analysis of archaeal and bacterial $SppB_{Tk}$ homolog sequences.

```
Aperni   SMEKKYGWRVITMIHREERVSFFGIPLRRFIDIDDSEAVIRAIRSTPPEKPIALILHTPG
Hbutyl   ----------MTMIHREERISFFGIPIRRFIDIEDSEAVLRAIRTTPPNKPIALILHTPG
Ihospi   RIEEKYGYRIITLIHRQERVGFFGIPFYRFIDVEDSEAVIRAIHNTPKDVPIMMILHTPG
Cmaqui   IIERKYGYRVITMIHRQEKVGFLGIPVYRYIDIEDSEAVIRAIRTTPPSMPIMLILHTPG
Pfurio   KMARKRNSTVITMIHRQESIGFFGIPVYKFISIEDSEEVLRAIRMAPKDKPIDLIIHTPG
Pabyss   KLARKRNSTVITLIHRQESIGLFGIPVYRFISMEDSEEVLRAIRMAPKDKPIDLIIHTPG
Phorik   RLARKRNSTVITLIHRQESIGLFGIPVYRFISIEDSEEVLRAIRMAPKDKPIDLIIHTPG
Mjann1   ELSNKRNSTVITMIHRQESIGLFGIPVYKFITIEDSEEILRAIRAAPKDKPIDLIIHTPG
Afulgi   RIGMKRGSNVITMIHRQESIGFLGIPIYRFIDIDDSEKVLRAIRSTPKDKPIDLIIHTPG
Msathe   KLGKRRGSLVITLIHRQEVISLLGLPLARYIDIDDSEEVLRAIRSAPNDVPIDIILHTPG
Uncult   RISRERKSQVITLIHRQETIAFLGIPLSRYIDIDDSEEVLRAIRTTPPDVPIDLIMHTPG
Mmarp1   QFEAVRKTRAIVMIHRQEQLALFGIPLYKYITIEDSEEILRAIRMTPEDMPIDLILHTPG
Mmarp2   QFEAVRKTRAIVMIHRQEQLALFGIPLYKYITIEDSEEILRAIRMTPEDMPIDLILHTPG
Mmarp3   QFEAVRKTRTIVMIHRQEQLALFGIPLYKYITIEDSEEILRAIRLTPEDMPIDLILHTPG
Mmarp4   QFEAVRKTRAIVMIHRQEQLALFGIPLYKYITIEDSEEILRAIRMTPEDMPIDLILHTPG
Mvanni   NFETSRKARAIVMIHRQEQLALFGIPFYRFINIEDSEEILRAIRMTPGDMPIDLILHTPG
Maeoli   KLEKERKTRVIAMIHRQETLALFGIPLYRFINMEDSEEVLRAIRMTSDDVPIDLILHTPG
Mvolta   QLEVSRGSRVIVMIHRQEQLAFFGLPLYKFISIEDSEEVLRAIRLTPENMPIDLILHTPG
Mjann2   EIERQRGTRVIAMIHRQEALTFLGIPIYKFITIEDSEEILRAIRLTPEDMPIDLIIHTPG
Aaeoli   KLEEKRKSRVITMIHRQEQIGFLGIPIFRFMTIEDSERVLRAIRMTPDDMPIDLIIHTPG
HydrSp   LIEKKYNARVITMIHRQEGLSFFGFAFMKFINIEDSEQVLRAIRMTPEDMPIVMILHTPG
Tcarbo   QFEQKRGSRLITMIHRQEAISLLGVPISRYINIEDSEHILRAIRLTPDDMPIDLVLHTPG
Chydro   DFETKRKSRLITMIHRQETFSFLGLPVSRYINIEDSEQILRAIRLTPDDMPIDLVLHTPG
Ptherm   QIELKRKSRLITLIHRQESISILGIPISRYINIEDSEAVLRAIRLTPDDMPIDLVLHTPG
Horen1   KIEDKTNSKLIALIHRQEALSFLGIPFRRFINIEDSEDILRAIRMTSDDKKIVIVIHTPG
Moothe   TFEKRRNSRLVTLIHRQEALSFLGIPLSRYINIEDSEQILRAIRLTPDDMPIDLVLHTPG
Horen2   KLEKERNSRVITMIHRQEILSFIGIPFTRFINIEDSEQILRAIRSTPDEKPIDILLHTPG
Tmarit   EIEKKRNSRVITLIHRTESISFLGFPVRRYIDIEDSEEILRAIKLTPSDMPIDLILHTPG
Tpetro   EIEKKRNSRVITLIHRTESISFLGFPVRRYIDIEDSEEILRAIKLTPSDMPIDLILHTPG
Caggre   QLERKRKSRVIVLIHRQETMSLLGFPLVRYINIEDSEAVLRAIKMTDRDIPIDLILHTPG
Cauran   QLERKRKSRVIVLIHRQETMSLLGFPLVRYINIEDSEAVLRAIKMTDRDIPIDLILHTPG
Rcaste   RLEQQRQSRVIVLIHRQETLSLLGFPLVRYIDIDDSEAVLRAIKMTDKDVPIDLVLHTPG
RosfSp   RIEQQRQSRVIVLIHRQETLSLLGFPLVRYIDIDDSEAVLRAIKMTDRDVPIDLVLHTPG
Tther1   ELERKRKSRVITLIHRQEAVSFLGIPISRYINIDDSEQVLRAIRLTDKNVPIDLILHTPG
Tther2   ELERRRKSRVITLIHRQEAVSFLGIPISRYINIDDSEQVLRAIRLTDKNVPIDLILHTPG
SynSp1   ELEQKRGSRVILLIHRQESISFLGIPVSRYISIEDSEQVLRAIRLTPPNTPIDLILHTPG
SynSp2   ELEQKRGSRVILLIHRQESISFLGIPVSRYISIEDSEQVLRAIRLTPPNTPIDLILHTPG
Avaria   EFQQERKSRVILLIHRQESISLLGIPISRYITIEDSEQILRAIRLTPPDVPIDLILHTPG
NostSp   EFQQERKSRVILLIHRQESISLLGIPISRYITIEDSEQILRAIRLTPPDVPIDLILHTPG
Npunct   EFQQERKSRVILLIHRQESISFLGIPISRYITIEDSEQILRAIRLTPPDVPIDLILHTPG
LyngSp   EFEQKRNSRVILLIHRQESISLFGIPLSRYISIEDSEQVLRAIRLTPSDVPIDLILHTPG
Mmaris   DLEAGRRTRVVALIHRQERIGFLGIPLFRYIDINDSEEVLRAIRLTAPEMPIDLILHTPG
AnaeSp   RIERRRKSRVILLVHRQETMSLLGFPLMRYIDINDSEDVLRAIELTDPEVPLDIVLHTPG
BradSp   QIERDRNSRVILLVHRQETMRLLGFPLMRYIDVNDSEDVLRAIHMTDDDVPLDIVLHTPG
Pbermu   RTQRARKSRIIAIVHRQEPMGLLGIPQLRYIDLNDAEDVLNAIRSTPAGTPLEIILHTPG
RosbSp   AIERKRRSRVITMIHRQERRSFFGVNVSRMIDLEDAQSIISAIKATPANTPIDLVLHTPG
RosvSp   MIERRRKSRVITMIHRQERRSLFGFNMTNTIGMEDAQSIISAIKATPRDTPIDLVLHTPG
Rxylan   RLASERGSTVITLIHRQETMSLLGFPIFRHIDIDDSEGVLNAIRETPSDGPIDIVLHTPG
LeptSp   HLETTRGTKVIAMVHRQETMSLFGFPVLRYINIEDSEEILRAIKMTDPDTPIDLILHTPG
Mbarke   AMEKSHGCKVLTMIHRREAISLFGIPAYQYIDEEDAEQILRWIRKYK-DYPLELILHTPG
Maceti   AMESRHGCKVLTMIHRREAISFFGLPAYQFIDEEDAEQILRWIRKYR-DSPLELILHTSG
Mmazei   AMEKHHGYKVLTMIHRREMISLFGLPAYQSIDEEDAEQVLRWIRKYR-DYPLELILHTPG
Tkodak   QLARKRNSTVIAMIHRQESIGLFGIPVYKFISIEDSEEVLRAIRSAPKDKPIDLIIHTPG
                      ** *     *          * #   # *             ** *
```

Fig. 3 (1/4)

```
Aperni   GLVLAASQIARALKRHRGRKIVIVPHYAMSGGTLIALAADEIRMDPNAVLGPLDPQLSAG
Hbutyl   GLVLAASQIAMALKRHPGKKIVIVPHYAMSGGTLIALAADEILMDPNAVLGPLDPQLALG
Ihospi   GMVLAASQIAKALHDHPAKKVAVVPHYAMSGGTLIALAADEIWMGPAAALGPLDPQVPVA
Cmaqui   GLVLAASQIARALKSHPAKKIVVVPHYAMSGGTLIALAADEIVMDQNAVLGPLDPQLGGP
Pfurio   GLVLAATQIAKALKDHPAETRVIVPHYAMSGGTLIALAADKIIMDPHAVLGPVDPQLGQ-
Pabyss   GLVLAATQIAKALKDHPAETRVIVPHYAMSGGTLIALAADKIIMDPHAVLGPVDPQLGQ-
Phorik   GLVLAATQIAKALKDHPAETRVIVPHYAMSGGTLIALAADKIIMDPHAVLGPVDPQLGQ-
Mjann1   GLVLAATQIAKALKAHPAETRVIVPHYAMSGGTLIALAADKIIMDENAVLGPVDPQLGQ-
Afulgi   GLVLAATQIAKALHDHPAKTTVIVPHYAMSGGTLIALAADEILIDPHAVLGPVDPQLMN-
Msathe   GLALAATQIALALKNHPARTSVIIPHYAMSGGTLIALAVDEIIMDPNAALGPVDPQLGDQ
Uncult   GIALAATQIALALKSHPARKTVIVPHYAMSGGTLIALAADEILMDPHAVLGPVDPQLAGK
Mmarp1   GLVLASEQIATALKEHKAKTTVIIPHYAMSGGSLIALAADEIIIDKNAVMGPVDPQVGQ-
Mmarp2   GLVLASEQIATALKEHKAKTTVIIPHYAMSGGSLIALAADEIIIDKNAVMGPVDPQVGQ-
Mmarp3   GLVLASEQIATALKEHKAKTTVIIPHYAMSGGSLIALAADEIIIDKNAVMGPVDPQVGQ-
Mmarp4   GLVLASEQIAIALKEHKAKTTVIIPHYAMSGGSLIALAADEIIIDKNAVMGPVDPQIGQ-
Mvanni   GLVLASEQIATALKEHKAKTTVIIPHYAMSGGSLIALSADEIIMDKNAVMGPVDPQVGQ-
Maeoli   GLVLASEQIAMALKEHKAKTTVIIPHYAMSGGSLISLAVDEIIMDKNAVMGPVDPQIGQ-
Mvolta   GLVLASEQIASALMEHKAKTTVIIPHYAMSGGSLIALAADEIIMDKNAVMGPVDPQIGQ-
Mjann2   GLALASEQIALALKEHKAKTTVIIPHYAMSGGSLIALAADEIIMDKNAVMGPVDPQIGQ-
Aaeoli   GLALAATQIANALVKHKAPVRVIVPHYAMSGGTLIALAADEIIMDENAVLGPVDPQIGN-
HydrSp   GLALAASQIASALAKHKSKVIVIIPHYAMSGGTLIALSADEITMDHNAVLGPVDPQIGQ-
Tcarbo   GLVLASEQIAHALMRHPAKVTVFVPHYAMSGGTMIALAADEIVMDCNAVLGPVDPQLGQ-
Chydro   GLVLAAEQIAEALRKHPAKVTVFIPHYAMSGGTLIALAADEIVMDENAVAGPVDPQLGE-
Ptherm   GLVLASEQIARALQKHPAKVTVFVPHYAMSGGTMIALAADEIVMDENAVLGPVDPQLGE-
Horen1   GLVLAAEQIAHAIKKHPSKVTVIVPHYAMSGGTLLALAADEIVMDENAVLGPVDPQIGN-
Moothe   GLVLAAEQIAHAILKHPAKVTVYVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGE-
Horen2   GLVLAAEQIAMAIKKHPAPVRVIVPHYAMSGGTLIALAADEIIMDKNAVLGPVDPQIGQ-
Tmarit   GLVLAAEQIARALKMHKGKVTVFVPHYAMSGGTLIALAADEIIMDENAVLGPLDPQIGN-
Tpetro   GLVLAAEQIARALKMHKGKVTVFVPHYAMSGGTLIALAADEIIMDENAVLGPLDPQIGN-
Caggre   GLVLASEQIARALRRHPAKVTVFVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGQ-
Cauran   GLVLAAEQIARALTKHAAKVTVFVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGQ-
Rcaste   GLVLAAEQIASALRKHPAKVTVFVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGQ-
RosfSp   GLVLAAEQIASALRKHPAKVTVFVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGQ-
Tther1   GLVLAAEQIAEALLRHPAKVTVFVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGQ-
Tther2   GLVLAAEQIAEALLRHPAKVTVFVPHYAMSGGTLIALAADEIVMDENAVLGPVDPQLGQ-
SynSp1   GLVLATEQIARALIRHPAKVTVFVPHYAMSGGTMLALAADEIVMDANAVLGPVDPQLGG-
SynSp2   GLVLATEQIARALIRHPAKVTVFVPHYAMSGGTMLALAADEIVMDANAVLGPVDPQLGG-
Avaria   GLVLATEQIARALIRHSAKVTVFVPHYAMSGGTMLALAADEIVMDANAVLGPVDPQLGN-
NostSp   GLVLATEQIARALIRHSAKVTVFVPHYAMSGGTMLALAADEIVMDANAVLGPVDPQLGN-
Npunct   GLVLATEQIARALIRHQAKVTVFVPHYAMSGGTMLALASDEIIMDANAVLGPVDPQLGN-
LyngSp   GLVLATEQIAHALIRHQAKVTVFVPHYAMSGGTMLALASDEIVMDENAVLGPIDPQLGN-
Mmaris   GLILSSEQIAMALRRHKGKVTVFVPHYAMSGGTLICLAADEVVMDENAVLGPVDPRIGE-
AnaeSp   GVVLAALQIARAERDHKGKVTVIVPHYAMSGGTLIALSADEILMSRHAVLGPVDPQLGQ-
BradSp   GLVLAALQIARAIRAHKAKVTVFVPHYAMSGGTLIALAADEIVMCRHSVLGPIDPQLGQ-
Pbermu   GLVLPALQIARAIKAHAGPKTVFVPHYAMSGGTLIALAADNIILNDHAVLGPIDPQIGG-
RosbSp   GLVLAAMQIARAVETHPAKVTVFVPIYAMSGGTLIALAADEIVMGEFSMLGPIDPQIMG-
RosvSp   GLVLAAMQIARAVEAHPAKVTIFVPIYAMSGGTLIALAADEIVMGEFSMLGPIDPQIMG-
Rxylan   GMVLAASQIAEALAEHRGPVRAFVPHYAMSGGTLIALAADEIHVDPHAALGPVDPQLGG-
LeptSp   GLVLASTQIAHALSNRKAPVTVFVPHFAMSGGTLIALSANRIVMDENAVLGPVDPQLGE-
Mbarke   GQLHSSIQIARALRRHSKNTKVIIPHYSMSGGTIIALAANEIVMDRDAVIGPIDPQIGDF
Maceti   GQLHASIQIARALKNHSKKTRVLIPHYSLSGGTIIALAADEIVMDKDAVIGPIDPQVGDL
Mmazei   GQLHASIQIARALKNHPKKTRVLIPHYSMSGGTIIALAADEIVMDKDAVIGPIDPQVGDP
Tkodak   GLVLAATQIARALKEHPAETRVIVPHYAMSGGTLIALAADKIIMDPNAVLGPVDPQLGQ-
         *      *** *   #        *    ***#   *  #    #     ** **
```

Fig. 3 (2/4)

```
Aperni  PTG--PAVPAPSVVKVARMKGKD-AQDTTLILADVAEKAIEEMREVITELLKDKMGV---
Hbutyl  PQG--PVVPAPSIVKVAKMKGDK-ASDTTLVVADVAEKAIREMQELIVYLLRDKMGE---
Ihospi  TPAGPMHVPSPSVVKVAEEKGKE-ANEYFLVHADVAKKALNEMLEFVTYILKDKVGE---
Cmaqui  GGVY---YPAPSILRAVEVKGRDKVDDETLILADVAEKSLRQVKELVMELLRGKVND---
Pfurio  -------YPAPSIIKAVEQKGAEKVDDQTLILADVAKKAIKQVQDFLYDLLKDKYGE---
Pabyss  -------YPAPSIIRAVERKGPDKVDDQTLILADVAEKAIKQVRDFIYDLLKDKYGD---
Phorik  -------YPAPSIVRAVEKKGVDKVDDQTLILADVAEKAIRQVRDFIFNLLKDKYGE---
Mjann1  -------YPAPSIVKAVEQKGADKADDQTLILADIAKKAINQVQNFVYNLLKDKYGE---
Afulgi  -------YPAPSILKVVEKKEPKDIDDKTLIMADIAEKAINQVRETVFNLLKDKMDE---
Msathe  TGA----YPATSILKVVEKKKIDEIDDKTLILAEEARKAVEQMKALLRRIVCSDCDE---
Uncult  DGA----FAASSILSVLRQKPAEKIADTTFILADDARKAQEQMKELVRIIIKGKCGE---
Mmarp1  -------YPAASIINAINTKYTDELDDETLILGDISRKAIFQVKEFVYEILKDKMG----
Mmarp2  -------YPAASIINAINTKYTDELDDETLILGDISRKAIIQVKEFVYEILKDKMG----
Mmarp3  -------YPAASIINAINTKYTDELDDETLILGDISRKAIVQVKEFVYEMLKDKMG----
Mmarp4  -------YPAASIINAINTKYTDELDDETLILGDIARKAIVQVKEFVYDILKDKMG----
Mvanni  -------YPAASIISAINTKYVDELDDETLILGDISRKAIDQVKEFVYDILKDKLG----
Maeoli  -------YPAASILNVIDTKYVDELEDETLILGDIAKKAINQVGEFVYWLLKDKMS----
Mvolta  -------YPAASILSVLDKKYISEIDDETLILADISKKAINQVKDYVCHILKNKVG----
Mjann2  -------YPAASILEAYYRK-GEKVSDETLILVDISKKAIKQMEEFVYELLKDKYG----
Aaeoli  -------MPAASILKVLEKKDPKDIDDQTLIMADVSEKAIKQMVDCLVDLLTKNGMD---
HydrSp  -------MPAASILKVLDVKKPEDIDDETMIMADVSKKAIEQMKSYVYELLKKKGHP---
Tcarbo  -------YPAASILKAVKQKSIDRVDDNTLIMADIAEKAIKQVEDFVARLLSDKMD----
Chydro  -------YPAASILKVVEQKGPDKVEDRTLIMADIAKKALAQVENTVVELLSDKMD----
Ptherm  -------FPAVSVLEVVRQKGRDRVDDRTLMLAEIASKAVSQVKEFLYFLLSDKME----
Horen1  -------YPAPSIVKVVEEKSVDEIDDKTLILGDVAQKAIVQIKDLVKKLLEGKMD----
Moothe  -------YPAASIIKVIEEKDKNKIDDRILILGDIARKAMNQVRESLVELLAEKME----
Horen2  -------YPAVSILKTASTKNKDELDDETLILADVASKAVHQLQNFLYLLLEDRME----
Tmarit  -------MPAPSILAAVKKKDVNEVDDQTLILADIAEKAIRQVKEFVVEILSDKVS----
Tpetro  -------MPAPSILAAVKKKDVNEVDDQTLILADIAEKAIRQVKEFVVEILSDKVP----
Caggre  -------HPAASILRVLERKPISEIDDETLMMADIAEKAIRQVKHTVCELLSDKMSV---
Cauran  -------HPAASILSVLERKPLSEIDDETLMMADIAEKAIRQVKRTVCELLRDKMPV---
Rcaste  -------QPAASILKVLERKPISEIDDETLIMADIAEKAIRQVKATVIELLADRIGA---
RosfSp  -------QPAASILKVLERKPISEIDDETLIMADIAEKAIRQVKATVIELLADRMGA---
Tther1  -------YPAASIVKVLEKKPLSEIDDQTLILADVAEKALRQVKTTVKNLLKKHMPE---
Tther2  -------YPAASIVKVLEKKPLSEIDDQTLILADVAEKALRQVKTTVKNLLKKHMPE---
SynSp1  -------YAAASILQVVADKPIAEIDDQTLILADLARKAMQQVQNFVRDLLKDDVPKRKI
SynSp2  -------YAAASILQVVAEKPIAEIDDQTLILADLARKAMQQVQNFVRDLLKDDVPKRKI
Avaria  -------FPAASILKVVKDKPIGEIDDQTLIMADLAAKAIQQVQRFVRTLLKDNIPKQKV
NostSp  -------FPAASILKVVKDKPIGEIDDQTLIMADLAGKAIQQVQRFVRTLLKDNIPKQKV
Npunct  -------YPAASILKVIEDKPISEIDDQTLIMADLSRKAIGQVQRFVRTLLKDSIPKQKV
LyngSp  -------FPAASILKVLEDKPISEIDDQTLIMADISRKAIRQVQDFVRTLLEDNIPQQKI
Mmaris  -------YPAASLLRVPRLKPPEEIDDETFVLVDIAAKAQTQMREFVADLLRERMDAG--
AnaeSp  -------YPAASLLKVVSKKPIAEIDDQTLVLADVAEKAVNQIRESVQDLFTRSVSA---
BradSp  -------SPAASLLKVIEQKPIAEIDDQTLVLADVGRKAIAQVKAAAKELLTDKMPE---
Pbermu  -------LPAASIAHVTETKSPDTIEDFTWVLADVARKAQQQLERAAKDLLSGTVSP---
RosbSp  -------ISAASIIAARDAKPIEHVSDVALVLADVSDKAIAQVVRGAVEIMTPRMER---
RosvSp  -------ISAASIVAARDSKPVEHVSDIALVLADVSDKAIRQVKRGAVELMTPRLEQ---
Rxylan  -------YPAASILAAVEQ--AKEPEDRTLILADMSRKALRQVEDFVTGLLERRMEP---
LeptSp  -------FPAASILRVVEQKDRNEIDDKTLIMADIAEKAILQLNHSLERLLVKRMGP---
Mbarke  IRG---MYPAPSWIYAAETK-KEKADDTTLVMSDVSRKALKFTRNVAKELLEGKIQPGPA
Maceti  MRG---VFPAPSWIYAAETK-KEKAEDTTLVMSDISRKALKLTRNVAKELLEGKVEAGPE
Mmazei  IRG---VFPAPSWIHAAETK-KEDADDSTLVMSDISRKALRLTRNVAKELLEGKIQPDGK
Tkodak  -------YPAPSILRAVEKKGPEKVDDQTLILADVAEKAIKQVQDFVFSLLKDKYGE---
                 *              #      #   *
```

Fig. 3 (3/4)

```
Aperni  --EKARETAVLLTEGRWTHDYPITFEKAREIG-LPVSDEVPPEVYQLMELYPQAPANRPG
Hbutyl  --EKAWELAKKLTEGGWTHDYPITVEKARELG-LPVKTELPPEVYELMELYPQAPYNRPG
Ihospi  --EKAREIAEELVTGKYTHDHPLFYEDLVKLG-LPVKKEVPEEVWALMELYPQAQPQRPG
Cmaqui  --ERVEEIADKLVGGYYTHDYPITVEQLKEMG-FKVSTNVPPEVYELMDLYPQARTNRPG
Pfurio  --EKARELAQILTEGRWTHDYPITVEHARELG-LEVDTNVPEEVYALMELYKQPVRQRG-
Pabyss  --EKAKELAKILTEGRWTHDYPITVEEARKLG-LNVSTDVPEEVYALMELYKQPVRQRG-
Phorik  --EKAKELAQVLTEGRWTHDYPITVEEAKKLG-LNVSTDVPEEVYALMELYKQPIRQRG-
Mjann1  --EKAKELSKILTEGRWTHDYPITVEEAKELG-LDVDTNVPEEVYTLMELYKQPVRQRG-
Afulgi  --EKAREVAKILTEGRWTHDYPLTVEELRQLG-LKVSTDVPEEVYELMELYPQPMMQRHP
Msathe  --ETVNRIIEEMVSGKYTHDHPFLADDLRALLGDRVKTDVPQEVYELMALYRMDISRRRP
Uncult  --PNIETILAELVSGKYTHDYPIFPEKALELFGGCIRIGIPPAVYDLMDFYKMEMSSRRP
Mmarp1  -EEKAKYLSETLSTGKWTHDYPLTIRKLRELG-IEVNTDLPEIVYALFDLYRQPVN-QRP
Mmarp2  -EEKAKYLSETLSTGKWTHDYPLTIKKLKELG-IEVNTDLPEIVYALFDLYRQPVN-QRP
Mmarp3  -EEKAKYLSETLSTGKWTHDYPLTIKKLRELG-IEVNTDLSEIVYALFDLYRQPVN-QRP
Mmarp4  -EEKAKYLSETLSTGKWTHDYPLTINKLKDLG-IEVNTELPNIVYELFDLYRQPVN-QRP
Mvanni  -EEKAKNLSETLSSGKWTHDYPLTIKKLKELG-VEVTTDVPKEVYELFDLYRQPVN-QRP
Maeoli  -EEKAKYITELISTGVWTHDYPLTINKLKDMG-IKVNTNVPKTMYDLFDLYKQPTG-QKP
Mvolta  -FDKAKQLSEILATGKWTHDYPLYIEKLQNLG-VAINTKVPSEIYDLFDLYQQSSN-QRP
Mjann2  -DEKAKEIAKKLTSGTWTHDYPLTVSKLKELG-IEVNTNVPRKVYELLELYPQPMG-AKP
Aaeoli  -KEKAKKIAEELATGKFTHDYPLTVEYLKSLG-LPVNTNVPQEVYELMELYEQPMGSQPP
HydrSp  -DDVAKKIAEELSTGKFTHDYPLDVDQLKAMG-LNINTDVPEEVYELMELYDQPTNSQVP
Tcarbo  -AEKARELAKMLTEGRWTHDYPITCDKLKEMN-LPINPDLPPEIFELMELYPQPAQ-RRP
Chydro  -LEKAREVAKLLSQGTWTHDYPISAQQLKELG-LPVSTDLPREIYELMDLYPQPQG-RRP
Ptherm  -EARARELAELLASGTWTHDYPIDCEKLKEMG-LPVTVGLMSEIYYLMDLYPQPPQ-RRP
Horen1  -EEKVNRVADVLTEGYWTHDYPITVEKLKGLG-INVRTDLFKEIYQLMELYPQAGH-RRP
Moothe  -RQRAEELATVFSEGRWTHDYPIGVDQLIKMG-LPVKTGLPREIYQLMELYPQPAS-RRP
Horen2  -KEVVNRVVKNLSSGKFTHDFPLNLDVLEELG-IEVNTELPDEIYDLMELYPQPGW-GRP
Tmarit  -KEKAEKIADKLCSGYWTHDYPLNYEKLREMG-IQVKTDMPQEIYDLMDLYKQAEP-KRP
Tpetro  -KEKAEEIADKLCSGYWTHDYPLNYEKLREMG-IQVKTNMPQEIYDLMDLYKQAEP-KRP
Caggre  --EKAEEVAHTLASGVWTHDYPITVREARELG-LPISTDMPEEIYQMMALYPQTAQRRPS
Cauran  --ERAEEVAHTLASGVWTHDYPITVSEARELG-LPISTEVPEEIYQIMALYPQTAQRRPS
Rcaste  --EKAEHIATMLATGVWTHDYPISVREARELG-LPVSTDMPSLVYQLMGLYPQTAQRRPS
RosfSp  --EPAEHIATMLATGVWTHDYPISVREARELG-LPVSTDMPPLIYQLMGLYPQTAQRRPS
Tther1  --EKAEEVATLLSQGTWTHDYPIDVEQARSLG-LPVSTEMPIEVYELMELYPQAQGQRPS
Tther2  --EKAEEVATLLSQGTWTHDYPIDVEQARSLG-LPVSTEMPIEVYELMELYPQAQGQRPS
SynSp1  APERIEPLIQYLTSGQVTHDYPITAEEAQRLG-LPISTELPPQIYALMDLYPQAAIGRPS
SynSp2  APERIEPLIQYLTSGQVTHDYPITAEEAQRLG-LPISTELPPQIYALMDLYPQASMGRPS
Avaria  NPENIESIIEALTTGRVTHDYPITVEEATEMG-LPITVGLPHSIYELMDLYPQPQGGRPS
NostSp  NPENIESIIEALTTGRVTHDYPITVEEATEMG-LPITVGLPRSIYELMDLYPQSQGGRPS
Npunct  LPENIESIIEALTTGRVTHDYPITIEEATEMG-LPVTVGLPHSIYDLMDLYPQAQGGRPS
LyngSp  DPSRIDGVINALTTGRITHDCPIMAEEAAKLG-LPVTPGLPKSIYYLMDLYPQSRGNRPT
Mmaris  ---RADSLARLLSGGTWTHDYPITFEQARELG-LPVTSGMPAGIYRLMDLFPQAMPRRPS
AnaeSp  --EKSRELAELLSTGTWTHDYPITFEQAQRLG-FKVRSDVPAEFMQLMSLYPQPVRRRP-
BradSp  --DKAEALAEKLSTGQWTHDYPISPTEAKDLG-LPVSTNMPDAVLELMTLYPQPVRSQGG
Pbermu  --NAAHAIAEELSSGRWTHDYPIDAAEAREIG-LHTSTEMPEEIAALMELFPDKLSKQSV
RosbSp  --GKAEELAKTLASGTWTHDYALTPAEARELG-LEVTVGMPSEILDLMKLFPAP-VKQSA
RosvSp  --AKAEQLAETLASGTWTHDYALTATEAGGLG-LPITVGLPPEILELMKLYPAP-VKQSA
Rxylan  --ERARKVARTLSSGVWTHDHPLTPRDLEELG-LPVRVGVPGEIHALMRLYPQPRGRQSS
LeptSp  --EKASQLAGVLSSGHFTHDYPLTYEELIKFD-LPVSTDVPEDVYLFMRLFPQPKQAVPS
Mbarke  GESRLDEVVEKLVSGEMIHSTPLSAGEAKKIG-LSISTDFPQDVHEFMKLFKPVKRSVES
Maceti  GKSRLDEVVEKLVSGEMIHSTPLSAGEAKELG-LSVSTDFPEDVHEFMKLFRPVKRNVEY
Mmazei  -EDRLEEVVEKLVSGEMIHSTPLSAREAKELG-LSVNTDFPEDVHDFMRLFRPVKKTVEY
Tkodak  --EKARELAQILTEGRWTHDYPITVDHARELG-LNVETDVPEEVYALMELYKQPVKQRG-
                        *  #*
```

Fig. 3 (4/4)

Fig. 3. An alignment of all archaeal and bacterial $SppB_{Tk}$ homolog sequences used to construct the phylogenetic tree in Fig. 2. Residues completely conserved in all 53 sequences are indicated with asterisks. Residues highly conserved in the sequences are indicated with # symbols. Residues subjected to site-directed mutagenesis are indicated with arrowheads. Abbreviations are as follows. Aperni, *Aeropyrum pernix*, Hbutyl, *Hyperthermus butylicus*, Ihospi, *Ignicoccus hospitalis*, Cmaqui, *Caldivirga maquilingensis*, Pfurio, *Pyrococcus furiosus*, Pabyss, *Pyrococcus abyssi*, Phorik, *Pyrococcus horikoshii*, Mjann1, *Methanocaldococcus jannaschii*, Afulgi, *Archaeoglobus fulgidus*, Msathe, *Methanosaeta thermophila*, Uncult, uncultured methanogenic archaeon RC-1, Mmarp1, *Methanococcus maripaludis* C5, Mmarp2, *Methanococcus maripaludis* C7, Mmarp3, *Methanococcus maripaludis* C6, Mmarp4, *Methanococcus maripaludis* S2, Mvanni, *Methanococcus vannielii*, Maeoli, *Methanococcus aeolicus*, Mvolta, *Methanococcus voltae*, Mjann2, *Methanocaldococcus jannaschii*, Aaeoli, *Aquifex aeolicus*, HydrSp, *Hydrogenobaculum* sp. Y04AAS1, Tcarbo, *Thermosinus carboxydivorans*, Chydro, *Carboxydothermus hydrogenoformans*, Ptherm, *Pelotomaculum thermopropionicum*, Horen1, *Halothermothrix orenii*, Moothe, *Moorella thermoacetica*, Horen2, *Halothermothrix orenii*, Tmarit, *Thermotoga maritima*, Tpetro, *Thermotoga petrophila*, Caggre, *Chloroflexus aggregans*, Cauran, *Chloroflexus aurantiacus*, Rcaste, *Roseiflexus castenholzii*, RosfSp, *Roseiflexus* sp. RS-1, Tther1, *Thermus thermophilus* HB8, Tther2, *Thermus thermophilus* HB27, SynSp1, *Synechococcus* sp. JA-2-3B'a(2-13), SynSp2, *Synechococcus* sp. JA-3-3Ab, Avaria, *Anabaena variabilis*, NostSp, *Nostoc* sp. PCC 7120, Npunct, *Nostoc punctiforme*, LyngSp, *Lyngbya* sp. PCC 8106, Mmaris, *Methanoculleus marisnigri*, AnaeSp, *Anaeromyxobacter* sp. Fw109-5, BradSp, *Bradyrhizobium* sp. BTAi1, Pbermu, *Parvularcula bermudensis*, RosbSp, *Roseobacter* sp. AzwK-3b, RosvSp, *Roseovarius* sp. 217, Rxylan, *Rubrobacter xylanophilus*, LeptSp, *Leptospirillum* sp. Group II UBA, Mbarke, *Methanosarcina barkeri*, Maceti, *Methanosarcina acetivorans*, Mmazei, *Methanosarcina mazei*, Tkodak, *Thermococcus kodakaraensis*.

Peptidase activity of the wild-type and mutant $\Delta N28SppB_{Tk}$ proteins

The author used the synthetic peptide *Z*-Val-Lys-Met-MCA as the substrate for examining activity levels of the purified proteins. The activity levels of each mutant protein were measured at 50˚C in the presence of a fixed concentration of this substrate (50 μM) and compared with that of the wild-type enzyme (Fig. 4). The author found that the mutation of Ser130 led to complete abolishment of peptidase activity under these conditions. The H226A mutant also exhibited no activity. Although to various extents, the author was able to observe peptidase activity in all other mutant proteins. The results strongly indicate that $\Delta N28SppB_{Tk}$ utilizes a Ser-His-Asp(Glu) catalytic

triad for its peptidase activity. Among the mutants with replacements in completely conserved acidic residues, the D154A protein displayed the lowest levels in activity (only 1.7% activity retained compared to the wild-type protein), suggesting that this residue is the third residue in the catalytic triad. The only other completely conserved residue whose mutation led to over a 90% decrease in activity was His55 (2.4% activity compared to the wild-type protein). Among the highly (but not completely) conserved residues, the replacements of Glu77 (5.2%) and Asp186 (4.8%) had the largest effects. In particular, the residue variation of Asp186 in the $SppB_{Tk}$ homologs is limited to Glu (Fig. 3). It can thus be presumed that a negatively charged residue in this position plays an important role in the catalytic activity of $SppB_{Tk}$ homologs.

To examine whether the single-residue replacements had unintended broad effects on the protein structure, the author analyzed the CD spectrum of each mutant protein. In all mutant proteins, including the three that exhibited least activity under standard conditions (S130A, H226A and D154A), the CD spectra were indistinguishable from that of the wild-type protein (Fig. 5). The results confirm that the single-residue replacements did not lead to significant changes in the secondary structures of the proteins, indicating that the changes in activity levels in the mutant proteins were direct consequences brought about by single-residue exchange.

DISCUSSION

From the activity levels of the 19 mutant proteins, the author can conclude that $\Delta N28SppB_{Tk}$ is a serine peptidase, utilizing Ser130 as the nucleophilic serine. The results also strongly suggest that His226 is the general base that acts in increasing the nucleophilicity of Ser130. The determination of the remaining acidic residue is less

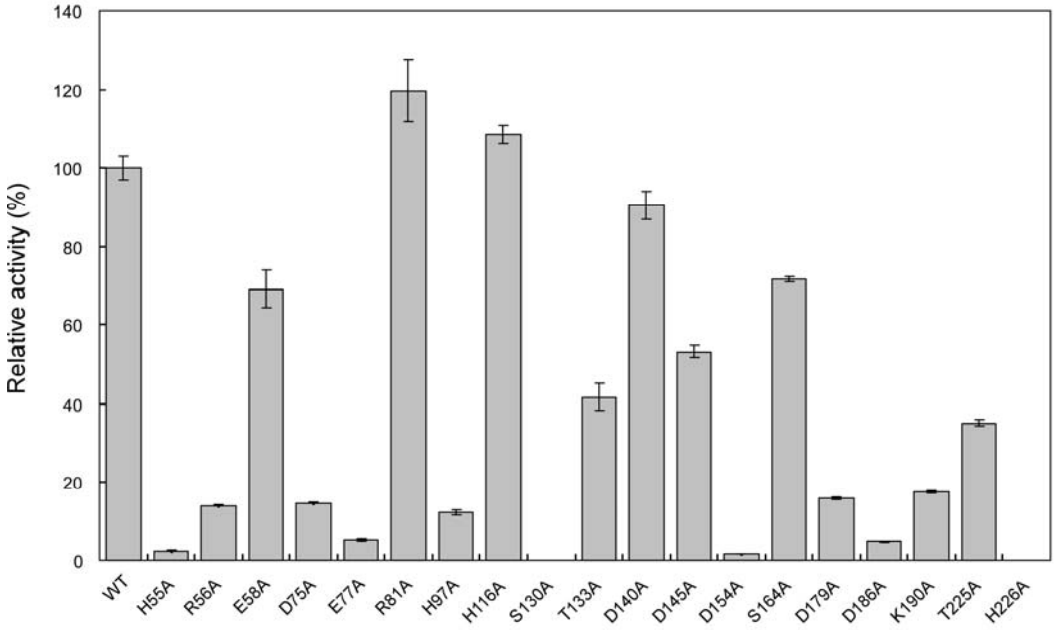


Fig. 4. Activity levels of the wild-type $SppB_{Tk}$ and 19 mutant proteins. The mutant proteins are indicated below the histogram and display the single residues replaced with alanine. The activity of the wild-type $SppB_{Tk}$ is designated as 100(%). Substrate concentration was fixed at 50 μM.

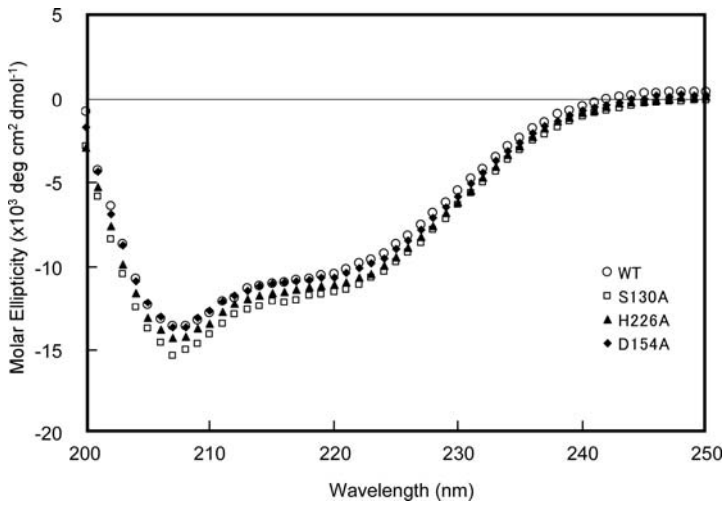


Fig. 5. Circular dichroism spectra of wild-type (WT), S130A, H226A and D154A proteins.

clear, as all mutant proteins other than S130A and H226A, although to various extents, exhibited peptidase activity. This is not surprising, as the serine and histidine residues are the most important residues in the Ser-His-Asp catalytic triad. It has been reported that replacing the serine or histidine residue in subtilisin or trypsin, both dependent on the catalytic triad, results in a 10^6-fold decrease in activity, whereas mutations in the Asp residue result in only a 10^4-fold decrease (1-4). There are also enzymes (5) and catalytic antibodies (6) that do not harbor an apparent Asp(Glu) residue, but display activity with only a Ser-His catalytic center. Judging solely from the activity levels of the mutant proteins with replacements in acidic residues, the third residue of the catalytic triad in $SppB_{Tk}$ is most likely Asp154. The mutations of these three residues, Ser130, His226 and Asp154, had the largest effects on activity, and the residues are completely conserved without exception in all $SppB_{Tk}$ homologs. The catalytic center of $SppB_{Tk}$ is thus distinct to that of $SppA_{Tk}$, which is dependent on a Ser-Lys catalytic dyad (Chapter 2).

The author noticed that an abundant number of His, Asp and Glu residues were important for activity. Besides the three residues proposed to form the catalytic triad, mutations in His55, Arg56, Asp75, Glu77, His97, Asp179 and Asp186 all led to proteins with over 80% decrease in activity. Multiple acidic residues may be interacting with the general base His226. The catalytic center of $SppB_{Tk}$ may be slightly distinct to those found in previously characterized serine proteases. This is supported by the fact that almost all conventional serine protease inhibitors such as phenylmethanesulfonate fluoride, diisopropyl fluorophosphate had no apparent effect on $SppB_{Tk}$ activity (Chapter 3). There are examples of serine proteases that harbor atypical catalytic centers composed of serine and histidine residues. The cytomegalovirus protease is dependent

on of a Ser-His-His or Ser-His-His-Asp active center for catalysis (7-9). In this structure (Ser132-His63-His157-Asp65), Ser132 is the nucleophile and His63 (Nε) is the general base that increases the nucleophilicity of Ser132. Instead of forming a hydrogen bond with an Asp/Glu side chain, Nδ of His63 forms a hydrogen bond with Nε of His157. There is also a further possibility that Nδ of His157 forms a hydrogen bond with Asp65, comprising a catalytic tetrad (8). A role similar to that of His157 would explain the importance of residues such as His55 and His97 in $SppB_{Tk}$, as their presence is important for activity but in no way can complement the function of His226. As no similarity can be observed between the primary structures of $SppB_{Tk}$ and cytomegalovirus protease, clarification of the roles of the individual residues will have to await the elucidation of the three-dimensional structure of $SppB_{Tk}$.

SUMMARY

In Chapter 3, the author examined the biochemical properties of a membrane-bound peptidase ($SppB_{Tk}$) from the hyperthermophilic archaeon, *Thermococcus kodakaraensis*. The substrate specificity of $SppB_{Tk}$ was distinct to that of $SppA_{Tk}$ (examined in Chapters 1 and 2), and raised the possibilities that the two peptidases efficiently complement one another in the initial breakdown of signal peptides at the cytoplasmic membrane of *T. kodakaraensis*. In this chapter, the author carried out a detailed site-directed mutagenesis study on the catalytic domain of $SppB_{Tk}$ ($\Delta N28SppB_{Tk}$), and determined the amino acid residues that contributed to the peptidase activity of the enzyme. Fifteen residues that were completely or highly conserved among the 53 $SppB_{Tk}$ homologs found in various archaeal and bacterial genome sequences were selected as targets for mutagenesis, focusing on amino acid residues

that have been shown to be involved in the catalysis of previously characterized peptidases/proteases. Although the replacement of a number of residues to alanine led to dramatic decreases in activity, mutation of Ser130 and His226 resulted in complete abolishment in activity. Among the acidic residues that were examined, Asp154 apparently contributed most to ΔN28SppB$_{Tk}$ activity. Circular dichroism studies on S130A, H226A and D154A confirmed that the mutations did not trigger broad structural changes in the respective mutant proteins. The results of this chapter indicate that the peptidase activity of SppB$_{Tk}$ is dependent on a Ser-His-Asp catalytic triad, distinct to the Ser-Lys catalytic dyad that is repsonsible for the activity of SppA$_{Tk}$.

REFERENCES

1. **Carter, P. & Wells, J. A.**, Dissecting the catalytic triad of a serine protease. *Nature*, 332, 564-568, 1988.
2. **Corey, D. R. & Craik, C. S.**, An investigation into the minimum requirements for peptide hydrolysis by mutation of the catalytic triad of trypsin. *J. Am. Chem. Soc.*, 114, 1784-1790, 1992.
3. **Craik, C. S., Roczniak, S., Largman, C. & Rutter, W. J.**, The catalytic role of the active site aspartic acid in serine proteases. *Science*, 237, 909-913, 1987.
4. **Paetzel, M. & Dalbey, R. E.**, Catalytic hydroxyl/amine dyads within serine proteases. *Trends Biochem. Sci.*, 22, 28-31, 1997.
5. **Wei, Y., Schottel, J. L., Derewenda, U., Swenson, L., Patkar, S. & Derewenda, Z. S.**, A novel variant of the catalytic triad in the *Streptomyces scabies* esterase. *Nat. Struct. Biol.*, 2, 218-223, 1995.

6. **Zhou, G. W., Guo, J., Huang, W., Fletterick, R. J. & Scanlan, T. S.**, Crystal structure of a catalytic antibody with a serine protease active site. *Science*, 265, 1059-1064, 1994.

7. **Qiu, X., Culp, J. S., DiLella, A. G., Hellmig, B., Hoog, S. S., Janson, C. A., Smith, W. W. & Abdel-Meguid, S. S.**, Unique fold and active site in cytomegalovirus protease. *Nature*, 383, 275-279, 1996.

8. **Shieh, H.-S., Kurumbail, R. G., Stevens, A. M., Stegeman, R. A., Sturman, E. J., Pak, J. Y., Wittwer, A. J., Palmier, M. O., Wiegand, R. C., Holwerda, B. C. & Stallings, W. C.**, Three-dimensional structure of human cytomegalovirus protease. *Nature*, 383, 279-282, 1996.

9. **Tong, L., Qian, C., Massariol, M.-J., Bonneau, P. R., Cordingley, M. G. & Lagacé, L.**, A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease. *Nature*, 383, 272-275, 1996.

PART II

Studies on a sugar transporter in *Thermococcus kodakaraensis* and development of a gene disruption system based on antibiotic resistance

CHAPTER 5

Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance

INTRODUCTION

As described in the GENERAL INTRODUCTION, hyperthermophiles have attracted much attention from an evolutionary viewpoint as they occupy the deepest lineages within the phylogenies of both *Archaea* and *Bacteria* based on ribosomal RNA sequences (1, 2). Hyperthermophiles are also focused upon as a source of (thermo)stable enzymes that have the potential for application in a broad range of technologies (3, 4). There are now 26 complete genome sequences and many more in progress, providing a wealth of primary structural data from which we can estimate the presence or absence of various metabolic and regulatory mechanisms. However, although biochemical and structural analyses of hyperthermophile proteins are proceeding at a rapid pace, genetic studies to examine gene function *in vivo* are still limited in number.

In contrast to the hyperthermophilic archaea, a wealth of gene disruption and shuttle vector systems has been developed for the mesophilic archaea. In the halophilic archaea, stable shuttle vectors have been developed (5, 6), and homologous recombination has been demonstrated (7, 8). In the methanogenic archaea, shuttle vectors and/or gene disruption systems have been developed in *Methanococcus maripaludis* (9), *Methanococcus voltae* (10), various *Methanosarcina* species (11) and *Methanobacterium thermoautotrophicum* (now *Methanothermobacter*

thermautotrophicus) (12). In the hyperthermophilic archaea, however, only two systems have been reported so far: one for *Thermococcus kodakaraensis* (13, 14) from the *Euryarchaeota* and the other for *Sulfolbus solfataricus* from the *Crenarchaeota* (15). Both systems rely on homologous recombination. The former system utilizes various host strains with amino acid/nucleotide auxotrophy and corresponding marker genes that complement the auxotrophy. The latter utilizes a *lacS*-deficient host strain and a modified but active *lacS* marker gene with selection based on lactose-dependent growth. The two systems have proved to be powerful tools in examining gene function in the respective strains (16-19) and can be expected to provide further genetic evidence that will help in understanding the physiological roles of genes in these and closely related organisms. In this study, the author aimed to develop a gene disruption system in hyperthermophiles using antibiotics and a marker gene that would confer resistance to transformant cells. This would relieve the necessity to prepare auxotrophic host cells and also allow selection of transformants in a nutrient-rich medium. Thus, the methodology should not only provide a convenient alternative for gene disruption in *T. kodakaraensis* but also be helpful in establishing gene disruption systems in other hyperthermophilic archaea. The author examined the possibilities of utilizing the mevinolin system established in the halophilic archaea (7, 20). Mevinolin, along with its analog simvastatin, is a specific inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, an enzyme essential for archaeal membrane lipid biosynthesis (21). HMG-CoA reductases have been extensively examined from a number of archaeal species (22, 23). An over-expression construct of the HMG-CoA reductase gene can be expected to be applicable as a marker gene. Additionally, as the gene is originally present in the hyperthermophile, there is no need for concern about the thermostability

of the marker gene product. As all archaeal strains are presumed to require the function of HMG-CoA reductase for lipid and membrane generation, the system described in this study has the potential for application in all hyperthermophilic archaea.

MATERIALS AND METHODS

Strains, media, and plasmids

T. kodakaraensis KOD1 and the mutant strains were cultivated under anaerobic conditions at 85˚C in a nutrient-rich medium, ASW-YT (Chapter 1), supplemented with various organic substrates or elemental sulfur when appropriate. In the case of plate culture, instead of elemental sulfur and $Na_2S{\cdot}9H_2O$, 2 ml of a polysulfide solution (10 g of $Na_2S{\cdot}9H_2O$ and 3 g of sulfur flowers in 15 ml of H_2O) per liter and Gelrite (10 g $liter^{-1}$) were added to solidify the medium. When simvastatin was added to the medium, simvastatin was dissolved in ethanol, and the amount of the solution added was adjusted so that the ethanol concentration in the medium was constant at 0.1% (vol/vol).

DNA manipulation and sequence analysis

Escherichia coli strain DH5α and pUC18/pUC19 were used for DNA manipulation and sequencing. *E. coli* strains were cultivated in LB medium as described in Chapter 1. All enzymes, reagents and apparatus used for DNA purification, plasmid construction and sequence confirmation are described in the MATERIALS AND METHODS section of Chapter 1.

Construction of the gene disruption vectors

Two disruption vectors, pUDapu and pUDmal, were constructed for the

targeted disruption of the *T. kodakaraensis* amylopullulanase gene *apu* (*apu*$_{Tk}$) and the sugar transporter gene cluster including *apu*$_{Tk}$, respectively, via double-crossover homologous recombination. Over-expression cassettes for the HMG-CoA reductase gene from *T. kodakaraensis* (*hmg*$_{Tk}$) were constructed by replacing the native promoter with a putative promoter region (-554 to -4) of the glutamate dehydrogenase gene (24). The region -3 to -1 was replaced by 5'-CAT-3' in order to incorporate an NdeI site for fusion of the promoter to the coding region of *hmg*$_{Tk}$. Cassettes were designed so that one had SmaI sites at both ends, while another had an XbaI site upstream of the promoter and a BamHI site downstream of *hmg*$_{Tk}$. The two cassettes were inserted into pUC18 and sequenced. For construction of the *apu*$_{Tk}$ disruption plasmid, a DNA fragment including *apu*$_{Tk}$ along with its flanking regions (about 1,000 bp) was amplified from the genomic DNA of *T. kodakaraensis* KOD1 with the primer set APU-F1 and APU-TRANS-R1 (5'-AATTCAGAACGGCAAGCTCTACGTAACAGACGGCA-3' and 5'-GCGTCGTAGATGTCCTCGGGCCTTATGCCGAAGAT-3', respectively) and inserted into pUC18 at the HincII site. An inverse PCR was then carried out to amplify the flanking regions and pUC18, thereby removing the coding region of gene. The primers used were APU-R2 and APU-F2 (5'-CTTATCACCTCACTCTTTAAGG CCTCCAACAGTGA-3' and 5'-AGAGGGTGGCGGAATCTGCGGCCCGGCGTT CCTCG-3', respectively). The DNA fragment was ligated with the *hmg*$_{Tk}$ over-expression cassette excised with SmaI and designated pUDapu. For disruption of the sugar transporter gene cluster, DNA fragments of the 5'- and 3'- flanking regions (about 1,000-bp) of the gene cluster were amplified with the primer pair TRANS-F1 and TRANS-R2 (5'-AGTTCTCAAATCGGACCTTCCGCCGATGGAAAAGT-3' and 5'-TGTTTATCACCTAGTTATCTCGTTGCATTTGAGTA-3', respectively) and the

pair TRANS-F2 and APU-TRANS-R1 (TRANS-F2, 5'-TCCCCAGGATCCGGCGGT GGTGAAGAGGGTGGCGG- 3'). The 5'-flanking region was inserted into pUC19 at the HincII site, followed by insertion of the over-expression cassette in the XbaI and BamHI sites. The 3'- flanking region was then inserted in the BamHI and SmaI sites, resulting in the plasmid pUDmal.

Transformation of *T. kodakaraensis* KOD1

Transformation procedures were performed as described previously (13, 14), but the host strain used in this study was the wild-type *T. kodakaraensis* KOD1. After transformation, cells were cultivated in ASW-YT liquid medium supplemented with 0.2% (wt/vol) elemental sulfur (ASW-YT-S^0) in the presence of 4 μM simvastatin at 85˚C for 12 h. The cells were further grown in ASW-YT-S^0 liquid medium with 8 μM simvastatin at 85˚C and spread on ASW-YT (polysulfide) plate medium containing 4 μM simvastatin and incubated at 85˚C. Genomic DNA was isolated from the transformants and analyzed by PCR and Southern blot analysis.

Southern blot analysis

A digoxigenin-DNA labeling and detection kit (Roche Diagnostics, Basel, Switzerland) was used according to the manufacturer's instructions. The probes within the coding regions of hmg_{Tk} and apu_{Tk} were amplified, respectively, with the primer pair HMG-F and HMG-R (5'-TGAGAACATCGGGCACTACTCAATAGATCCCAACC-3' and 5'-ACCAACGAGGTTCTTGCGGTAGTTCACCTCGGCTA-3', respectively) and the pair APU-F and APU-R (5'-CTCAACGACAAGACCCTTGAAATCCTAG CGGAGAA-3' and 5'-GGCTCATCTTATCTTTGTTTTCCATGAGGGCCTTT-3',

respectively). The probe within the coding region of the *malE* gene of *T. kodakaraensis* (*malE*$_{Tk}$) was amplified with the primers MalE-F and MalE-R (5'-CACTTCCCCGACCGAGACCACTACTACCTCACCCA-3' and 5'-CTGCTGG GTGTTGTAGTCGGCAGTCGGGGCCATGT-3', respectively). The probe within the *hmg* gene from *Pyrococcus furiosus* (*hmg*$_{Pf}$) was amplified with the primers PfHMG-F and PfHMG-R (5'-AAAGCACATTGGCCACTACTCAATTGATCCAAACG-3' and 5'-ACCCACTAAGTTCTTTAGGTAGTTTACTTCAGCGA-3', respectively), and the probe corresponding to the promoter region of the glutamate dehydrogenase gene was amplified with the primers GDHp-F and GDHp-R (5'-ATATCCCACCTC CGATTCCGTTGGTATTTAATCGG-3' and 5'-TACCACCTCATTTCGGTAATC TGCGAGGTTAACTT-3', respectively). Genomic DNA from the wild-type and gene disruption mutant strains was digested with PvuII.

Growth properties of *T. kodakaraensis* KOD1 and mutant strains

T. kodakaraensis KOD1 and the mutant strains were grown in ASW-YT-S^0 medium at 85˚C for 12 h and inoculated into 15 ml of ASW-YT-S^0 or ASW-YT medium supplemented with 0.5% (wt/vol) sodium pyruvate, a 0.5% (wt/vol) concentration of a specific maltooligosaccharide (3, 4, 5, 6, or 7 glucose units), 0.5% (wt/vol) amylose (polysaccharide consisting of glucose connected solely by α-1,4-glycosidic bonds), or 0.5% (wt/vol) pullulan (polysaccharide consisting of maltotriose units connected by α-1,6-glycosidic bonds). Cell densities (optical density at 660 nm) were measured at appropriate intervals with a UV spectrometer mini photo 518R (Taitec, Koshigaya, Japan). In order to estimate resistance toward simvastatin, cells were cultivated in 15 ml of ASW-YT-S^0 supplemented with 1, 5, 10, or 20 μM

simvastatin.

Measurements of HMG-CoA reductase activity

Activity measurements were performed at 60˚C in a final volume of 1 ml containing cell extracts, 200 μM NADPH, and 0.5 mM HMG-CoA (Sigma, St. Louis, MO, USA) in 50 mM potassium phosphate buffer (pH 7.0). The consumption of NADPH was monitored at 340 nm by a UV-visible light spectrophotometer (UV-1600PC; Shimadzu, Kyoto, Japan). Cell extracts were prepared as follows. *T. kodakaraensis* KOD1 and the disruptants were cultivated in ASW-YT-S^0 medium at 85˚C for approximately 8 h. Cells were collected and sonicated on ice, and the supernatant after centrifugation (20,000 x *g* for 30 min at 4˚C) was used as the cell extract. Protein concentrations were determined with a protein assay kit (Bio-Rad) using bovine serum albumin as a standard.

RESULTS

Effect of various concentrations of simvastatin on the growth of *T. kodakaraensis* KOD1

As isopentenyl diphosphate is the major precursor for archaeal lipid membranes, the author supposed that inhibition of HMG-CoA reductase would have severe effects on the growth of *T. kodakaraensis* KOD1 (Fig. 1). The main concerns were whether the uptake and inhibitory effects of simvastatin were sufficient to allow use of the antibiotic at realistic concentrations and whether the compound was stable enough at temperatures of >80˚C to inhibit growth for several days, which is necessary for the formation of colonies.

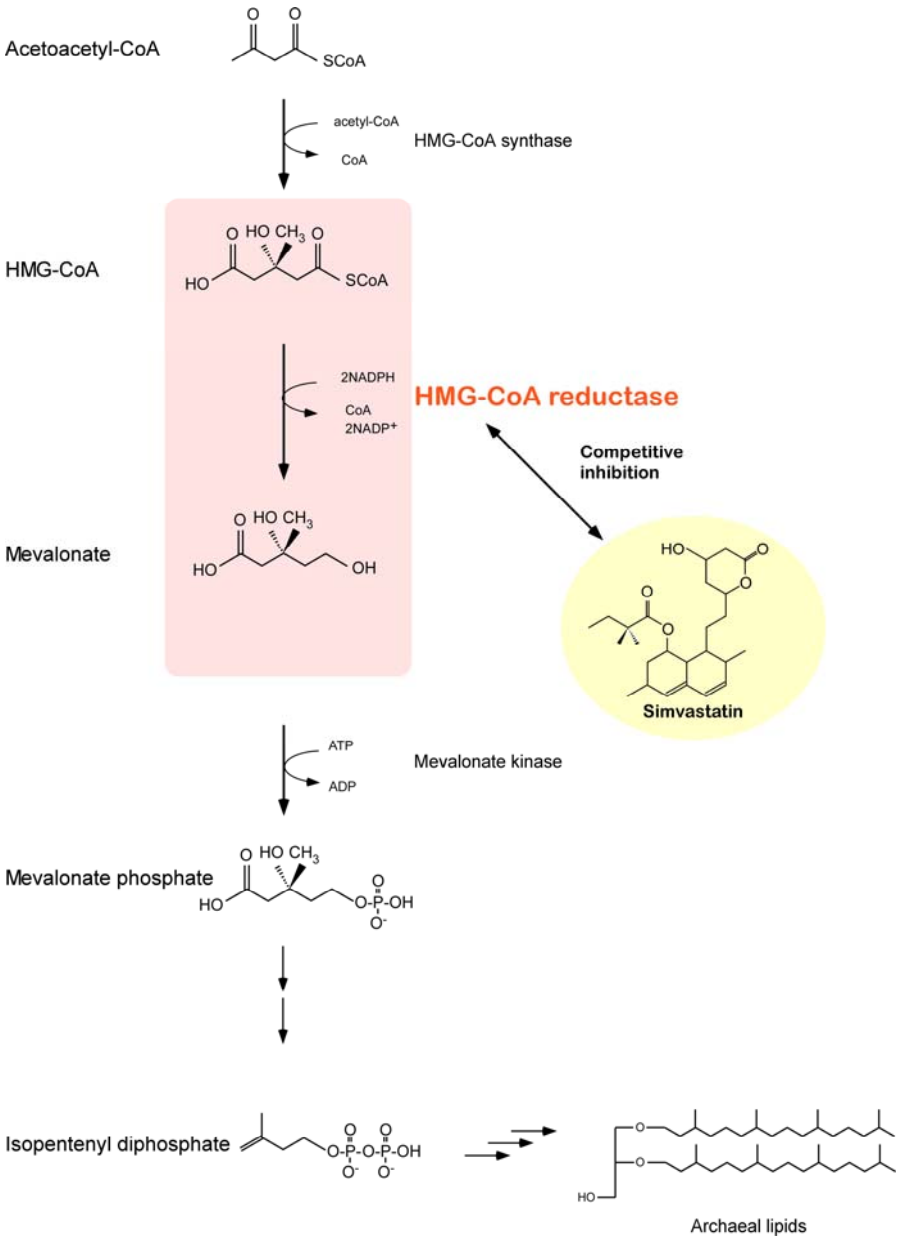


Fig. 1. The mevalonate pathway for isoprenoid lipid biosynthesis in the *Archaea*. The reaction catalyzed by HMG-CoA reductase is shaded in pink. Simvastatin is shaded in yellow. It should be noted that the two reactions converting mevalonate phosphate to isopentenyl diphosphate are distinct from the reactions in the classical mevalonate pathway (25).

The author examined the growth of *T. kodakaraensis* KOD1 in the presence of various concentrations of simvastatin in the nutrient-rich medium ASW-YT-S^0. Simvastatin was dissolved in ethanol, and the amount of ethanol added to the medium was constant at 0.1% (vol/vol). In ASW-YT-S^0 medium, *T. kodakaraensis* KOD1 cells reach the stationary phase within 24 h. No effect on growth was observed with the addition of ethanol alone. In the presence of 1 or 2 μM simvastatin, growth was observed only after 24 h, while 48 h was necessary for growth with 3 μM simvastatin. At concentrations of 4 or 5 μM simvastatin, growth was not observed for at least 5 days, indicating that these concentrations would be suitable for selecting transformants with resistance against simvastatin. The author also confirmed that these concentrations were sufficient to prevent colony formation of *T. kodakaraensis* KOD1 on nutrient-rich plate medium.

A cassette for the over-expression of the HMG-CoA reductase gene

As simvastatin is a competitive inhibitor of HMG-CoA reductase, the author expected that over-expression of its gene from *T. kodakaraensis* (hmg_{Tk}) would reduce the inhibitory effects of simvastatin on cell growth. Previous studies have indicated that the enzyme glutamate dehydrogenase is abundant in *T. kodakaraensis* cells grown in various media (49), suggesting that the gene (gdh_{Tk}) is under the control of a strong promoter. The author therefore utilized a 551-bp intergenic region between the coding regions of gdh_{Tk} (TK1431) and the adjacent gene TK1432 and fused the region upstream of hmg_{Tk} (see MATERIALS AND METHODS). This over-expression cassette (P_{gdh}-*hmg*) was used as the marker gene for construction of disruption plasmids (Fig. 2A).

Design and construction of the gene disruption plasmids

The genes disrupted in this study were a putative amylopullulanase gene (*apu*$_{Tk}$, or TK1774) and a gene cluster including *apu*$_{Tk}$ and three additional genes encoding the components of a sugar transporter (TK1771 to TK1773) of *T. kodakaraensis* (24) (Fig. 2B). In the latter stages of this study, the author discovered an error in the original genome sequence of TK1774 (an excess A at position 1,581,978 of the genome). The correct sequence leads to a protein with a change and elongation in sequence in the carboxy-terminal region from residue Asn1070 (see DISCUSSION). In this chapter the author will refer to the corrected *apu*$_{Tk}$ gene as TK1774* and, for simplicity, to the four-gene cluster (TK1771 to TK1774*) as *mal*$_{Tk}$.

Amylopullulanases, or type II pullulanases, exhibit both α-amylase and pullulanase activities and can therefore cleave both α-1,4- and α-1,6-glucosidic bonds (20). There are a number of other homologs on the genome (24), some with putative signal sequences for secretion, expected to harbor the ability to degrade α-linked polysaccharides. In particular, the TK1884 protein has been experimentally confirmed to exhibit α-amylase activity (26). On the other hand, in contrast to the two sugar transporters present in *P. furiosus* (Mal-I, PF1739 to PF1744, and Mal-II, PF1933 to PF1938) (27), only one putative gene cluster is found on the *T. kodakaraensis* genome (TK1771 to TK1775) (28). Based on primary structure similarity, the transporter from *T. kodakaraensis* corresponds to Mal-II, suggesting that it is specific to maltooligosaccharides with three or more glucose units. By disrupting *apu*$_{Tk}$, the author expected to gain insight into the actual degree of influence *apu*$_{Tk}$ has, among the multiple amylase homologs of *T. kodakaraensis*, on the degradation of various extracellular polysaccharides. Growth characteristics of the *mal*$_{Tk}$ disruptant were

expected to clarify the presence or absence of other sugar transporters as well as to provide information on the substrate specificity of the Mal_{Tk} transporter *in vivo*.

Similar to the design of gene disruption plasmids in a previously described system using *pyrF* or *trpE* as selectable markers (14), P_{gdh}-*hmg* was inserted between the 5'- and 3'- flanking regions (1,000 bp) of the target gene(s) (Fig. 2C). The plasmids pUDapu and pUDmal were used to transform wild-type *T. kodakaraensis* KOD1, and transformants were selected based on their resistance toward simvastatin.

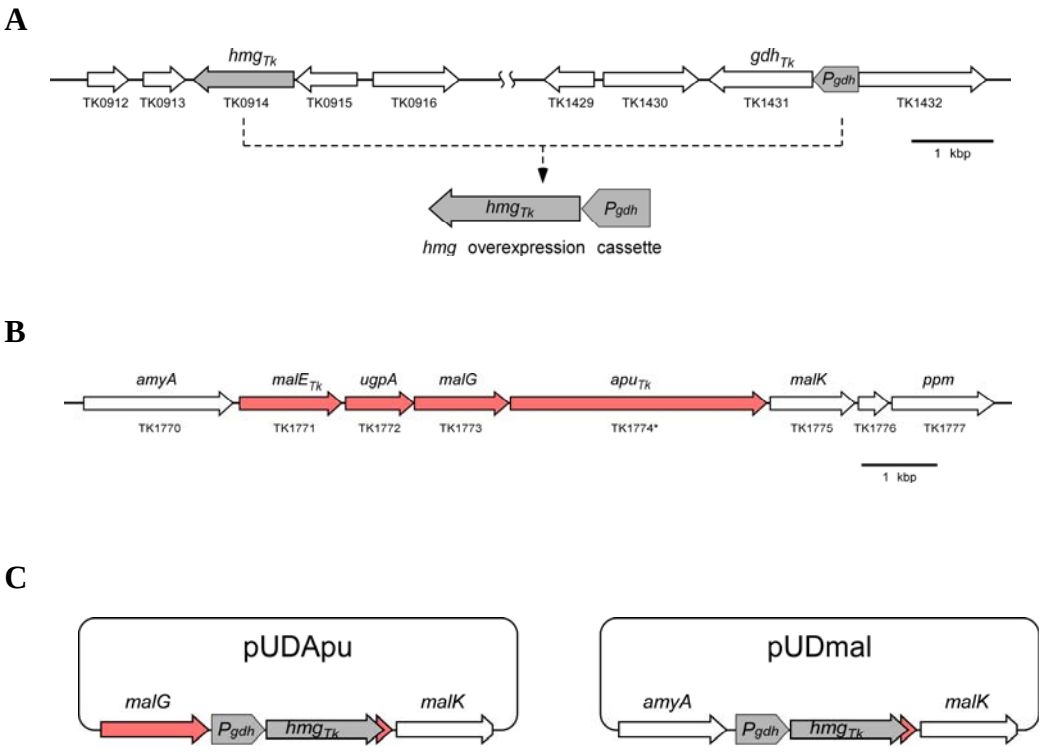


Fig. 2. (A) Design of the hmg_{Tk} over-expression cassette using the 5'-upstream flanking region of gdh_{Tk}. (B) Gene organization of the putative maltooligosaccharide transporter of *T. kodakaraensis*. TK1774* represents the correct sequence of the apu_{Tk} gene (see RESULTS section). Red arrows indicate the gene(s) disrupted in this study. (C) The two plasmids constructed for the disruption of the apu_{Tk} and mal_{Tk} loci via double-crossover recombination.

Isolation of the gene disruption strains Δ*apu*$_{Tk}$ and Δ*mal*$_{Tk}$

After transformation, cells were grown in ASW-YT-S^0 liquid medium in the presence of 4 μM simvastatin. Growth was observed with cells transformed with pUDapu and pUDmal but not for cells treated without plasmid. Cells were further inoculated in the same liquid medium with 8 μM simvastatin and then spread on plate medium with 4 μM simvastatin. Five colonies were selected for each gene disruption and grown in ASW-YT-S^0 medium. The author examined the *apu*$_{Tk}$ and *mal*$_{Tk}$ loci by PCR (Fig. 3). As expected, the author observed shorter amplified fragments from the transformants than from the wild-type strain, corresponding to the decrease in length brought about by the replacement of *apu*$_{Tk}$ and *mal*$_{Tk}$ by P$_{gdh}$-*hmg*.

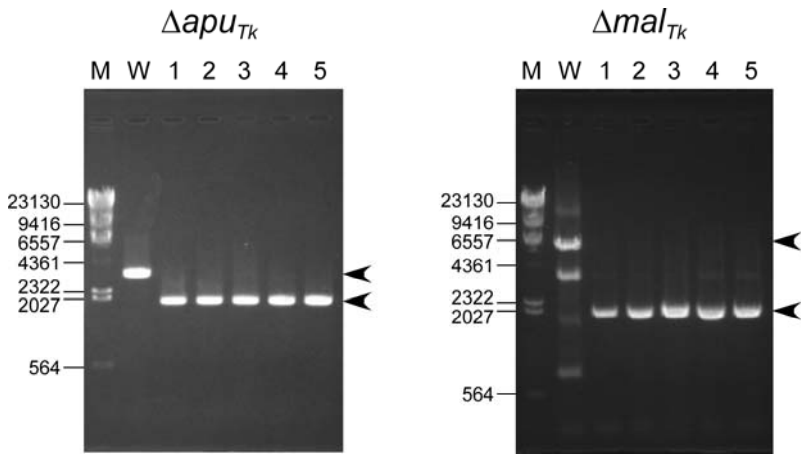


Fig. 3. Disruption of the *apu*$_{Tk}$ and *mal*$_{Tk}$ loci of *T. kodakaraensis*. PCR analyses of the *apu*$_{Tk}$ and *mal*$_{Tk}$ loci confirming gene disruption are shown. Primers were designed in the 5'- and 3'-flanking regions of the gene(s) to be disrupted. DNA size markers were run in lane M, and their sizes (bp) are indicated to the left of the gels. The results of PCR with wild-type *T. kodakaraensis* KOD1 and five individual transformants are indicated in lane W and lanes 1 to 5, respectively. The arrowheads to the right of the gels indicate the amplified fragments expected before and after recombination. The decreases in lengths of the amplified fragments reflect the differences in length between *apu*$_{Tk}$ (~3,500 bp) and P$_{gdh}$-*hmg* (~2,000 bp) and between *mal*$_{Tk}$ (~7,000 bp) and P$_{gdh}$-*hmg*. Nonspecific amplifications of DNA fragments observed for the wild-type *mal*$_{Tk}$ locus were due to the prolonged reaction time necessary to amplify the entire locus.

Extent of simvastatin resistance of the transformants

As the transformants harbored P_{gdh}-*hmg* on their genomes, the author examined their resistance against various concentrations of simvastatin (Fig. 4). One Δapu_{Tk} and one Δmal_{Tk} transformant were grown in the presence of 1, 5, 10, and 20 μM simvastatin, and their growth characteristics were compared with those observed in medium without simvastatin. Although the wild-type strain could not grow at all with 5 μM simvastatin, specific growth rates and cell yields of the transformants were still comparable to those in the absence of simvastatin. The degree of inhibition became prominent at higher concentrations, but the author found that concentrations over 20 μM were necessary to completely inhibit growth of the transformants. The author further examined the levels of HMG-CoA reductase activity in the wild-type and transformant cells. Specific activity in the cell extracts of wild-type cells was approximately 25 nmol min^{-1} mg^{-1}. In contrast, the level observed in the extracts of the Δapu_{Tk} strain was 760 nmol min^{-1} mg^{-1},

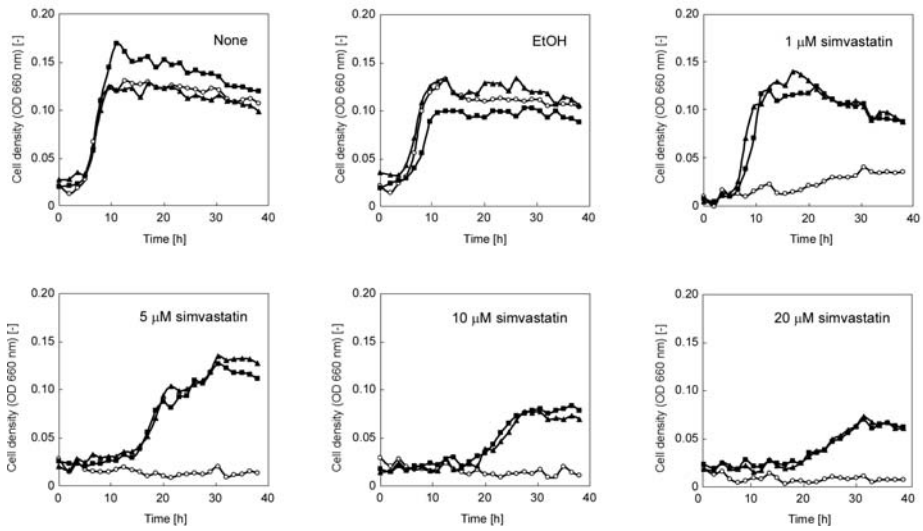


Fig. 4. Growth of wild-type *T. kodakaraensis* KOD1 and Δapu_{Tk} and Δmal_{Tk} mutant strains in the presence of various concentrations of simvastatin. Open circles, wild-type strain; solid squares, Δapu_{Tk} strain; solid triangles, Δmal_{Tk} strain; OD, optical density.

indicating an increase in activity of over 30-fold. The resistance against simvastatin and the increase in HMG-CoA reductase activity in the transformants are consistent with the presumption that simvastatin inhibits growth of *T. kodakaraensis* KOD1 by specifically inhibiting the activity of HMG-CoA reductase.

Phenotype analyses of the gene disruption strains

The author examined the growth characteristics of the Δapu_{Tk} and Δmal_{Tk} strains in various media and compared them with those of the wild-type strain (Fig. 5). No change in phenotype was observed when the three strains were grown on amino acids (ASW-YT-S^0) or amino acids and pyruvate (ASW-YT-pyruvate) as carbon sources. However, disruption of apu_{Tk} and mal_{Tk} brought about dramatic changes in phenotype when the strains were grown on various sugars. Disruption of the mal_{Tk} transporter abolished growth on all sugars examined. Although the Δapu_{Tk} strain displayed growth on a number of maltooligosaccharides, the strain could not grow on pullulan (Fig. 5I). Interestingly, while several α-amylase homologs (including TK1884) are present on the genome, the author found that disruption of apu_{Tk} led to a significant decrease in growth rates when strains were grown on amylose, a maltopolysaccharide consisting of only α-1,4-linkages (Fig. 5H). Another intriguing finding was that the disruption of apu_{Tk} had a greater detrimental effect on growth with shorter maltooligosaccharides, which was rather surprising as amylopullulanases are presumed to function in the breakdown of poly- or oligosaccharides. In contrast to the wild-type strain, no growth was observed for Δapu_{Tk} in the medium supplemented with maltotriose (Fig. 5C).

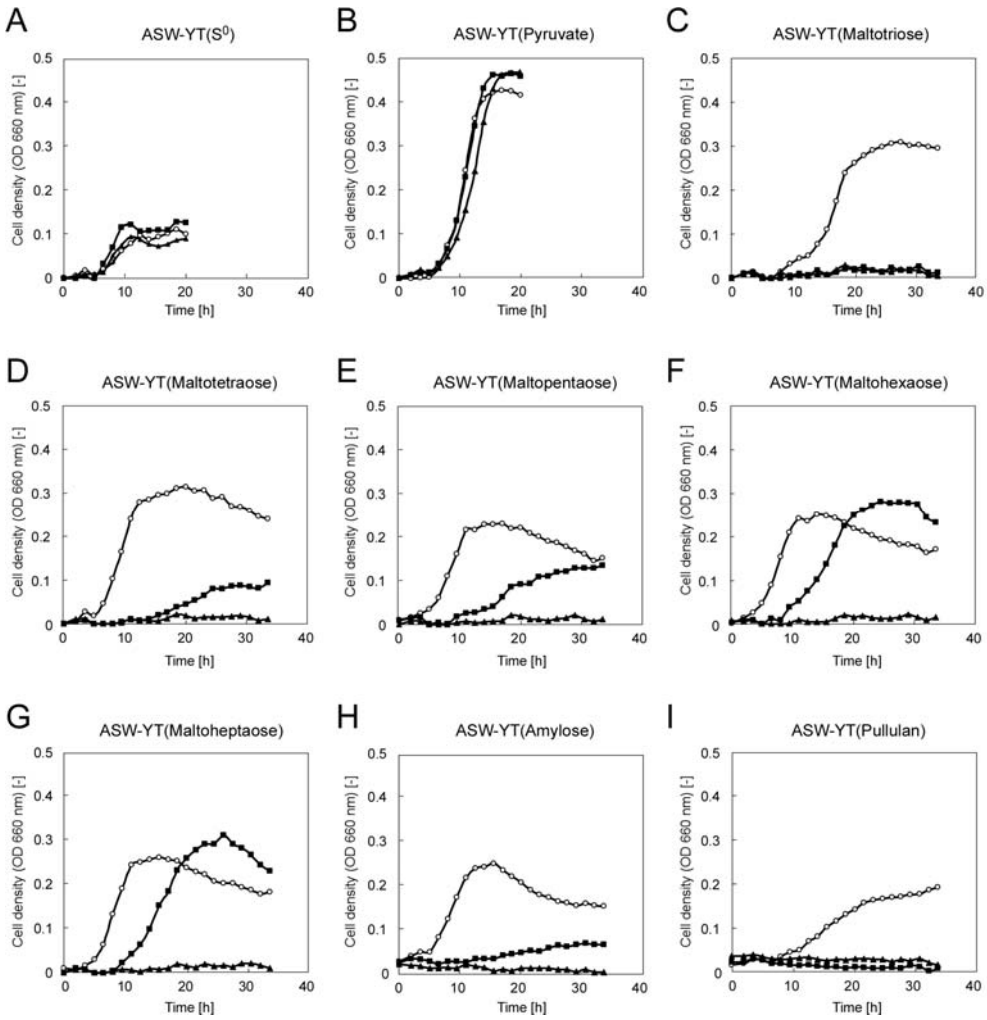


Fig. 5. Growth of wild-type *T. kodakaraensis* KOD1 and Δapu_{Tk} and Δmal_{Tk} mutant strains on various carbon sources. The carbon sources examined are indicated above each panel. Glucose and maltose were not examined, as the wild-type strain cannot utilize these sugars. Open circles, wild-type strain; solid squares, Δapu_{Tk} strain; solid triangles, Δmal_{Tk} strain; OD, optical density.

Recombination at the *hmg_{Tk}* locus

As the plasmid used in this study harbors the endogenous *hmg_{Tk}* and P_{gdh}, there will always be a possibility of these regions recombining with the corresponding native loci present on the genome. This property has actually been applied in developing a single-crossover insertion/pop-out recombination system using the *pyrF* gene as a selective marker (14). The author therefore examined whether recombination events had occurred at the *hmg_{Tk}* and/or P_{gdh} glutamate dehydrogenase locus. The author observed that in some strains the native *hmg_{Tk}* locus had been disturbed (data not shown). As the locus of the target gene is stable in a disrupted form via double-crossover recombination, the recombination at the *hmg_{Tk}* locus does not directly pose a problem. Nevertheless, the author examined possibilities to prevent or decrease the frequencies of unintended recombination. Linearizing the plasmids prior to introducing them into the cells would prevent single-crossover recombination. Utilizing an *hmg* gene from a heterologous host would also decrease the possibilities of single-crossover recombination as well as further recombination due to the presence of two identical regions on the genome. The author found that both methods could be used for gene disruption. In examining the latter possibility, the author used the *hmg* gene from the closely related *P. furiosus* and constructed plasmids to disrupt the *apu_{Tk}* gene. In five randomly chosen transformants with resistance against simvastatin, the author clearly observed that double-crossover recombination had occurred at the *apu_{Tk}* locus (Fig. 6A and B) along with the appearance of a single copy of the *hmg_{Pf}* gene (Fig. 6C) and an additional copy of the *gdh* promoter (Fig. 6E), whereas neither the native *hmg_{Tk}* locus nor the *gdh* promoter regions were disturbed (Fig. 6D and E). The author further confirmed that the strains obtained using *hmg_{Pf}* as a marker gene (i) were able to grow in the presence of 20 μM

simvastatin, (ii) exhibited significant HMG-CoA reductase activities in the cell extracts (720 nmol min^{-1} mg^{-1}), and (iii) displayed the same growth characteristics toward various carbon sources as those shown in Fig. 5 (data not shown).

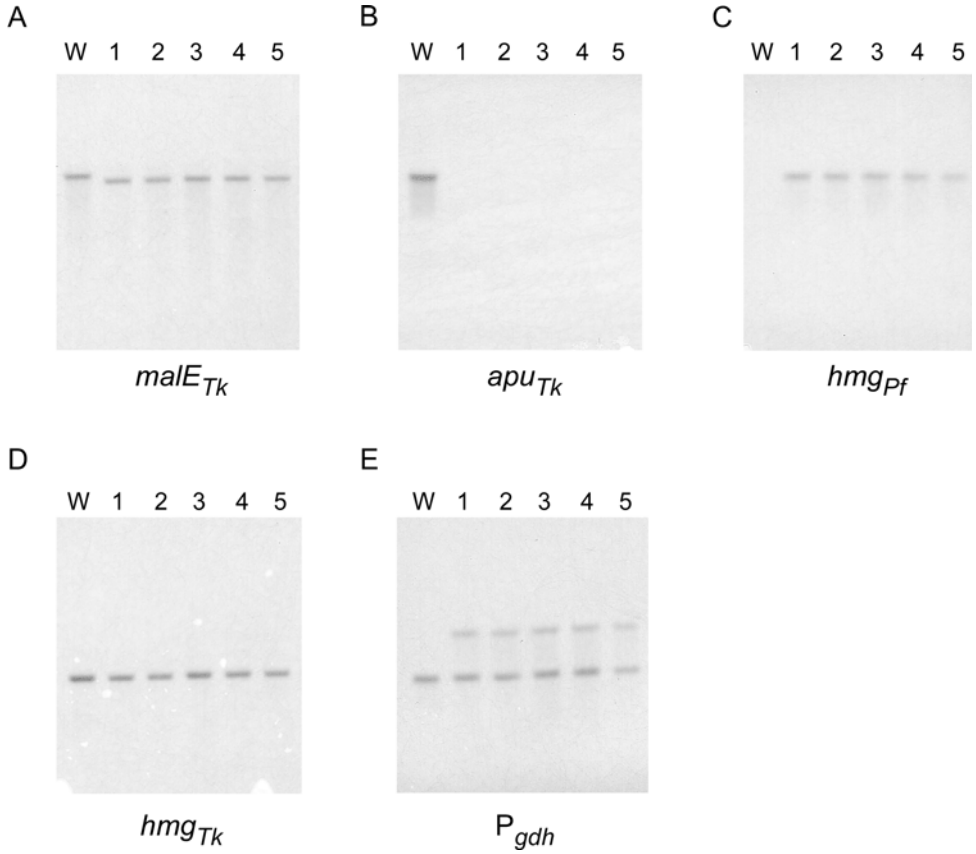


Fig. 6. Southern blot analyses on Δapu_{Tk} strains obtained with the hmg_{Pf} gene as a selectable marker. Genomic DNA from five selected Δapu_{Tk} strains and from wild-type *T. kodakaraensis* KOD1 (W) were subjected to Southern blot analyses using probes within the regions indicated below each membrane.

DISCUSSION

In this chapter, the author has developed a gene disruption system in *T. kodakaraensis* KOD1 based on resistance against antibiotics using simvastatin and an

over-expression cassette of hmg_{Tk}. The system has many advantages for initiating gene disruption studies in hyperthermophilic archaea. First, one does not need to construct a host strain with a particular defect or auxotrophy toward an amino acid. There is also no need for selection to be carried out in minimal medium. Positive selection of mutant strains is possible in nutrient-rich medium. If the genome sequence is available, the possibilities to disrupt genes with this system can be examined immediately. It should be noted that the author has not examined the stability of simvastatin under acidic conditions, which should be examined prior to application of the method on strains such as *Sulfolobus*. When the genome sequence is not available, it may be possible to use the genes of a closely related strain whose genome has been sequenced. As demonstrated in this study, the heterologous hmg_{Pf} gene was applicable for gene disruption in *T. kodakaraensis*. In order to avoid initial single-crossover recombination and possible recombination events afterwards due to the presence of two identical regions on the same chromosome, a heterologous marker may be more advantageous than an endogenous marker. The author hopes this methodology will promote gene disruption studies in a broader range of hyperthermophilic archaea.

The phenotypes of the Δapu_{Tk} and Δmal_{Tk} strains in various media not only provide valuable information on the physiological roles of the disrupted genes themselves but also allow us to estimate the contribution of other genes on the genome. The growth characteristics of the Δmal_{Tk} strain indicated that it is the only transporter in *T. kodakaraensis* involved in the uptake of the poly- and oligosaccharides examined in this study. As the Δapu_{Tk} strain could not grow at all on pullulan, it is most likely that Apu_{Tk} is the only relevant enzyme responsible for the extracellular hydrolysis of pullulan, in spite of the fact that another gene (TK0977), annotated as a type II pullulan

hydrolase, is present. From the results shown in Fig. 4H, the author also found that Apu_{Tk} plays a much greater role than expected in the cleavage of α-1,4-glycosidic linkages, suggesting that the experimentally verified α-amylase TK1884 protein (26) may not be the major amylose-degrading enzyme in *T. kodakaraensis*. This agrees well with the results of a transcriptome analysis of *P. furiosus* grown on starch, which revealed that PF1935*, the homolog of TK1774*, is the protein most up-regulated in the presence of starch (29).

An intriguing change in phenotype was observed in the Δapu_{Tk} strain grown on maltooligosaccharides. As the amylopullulanase was presumed to function in the degradation of poly- or oligosaccharides, the effects brought about by disrupting apu_{Tk} were expected to be greater with longer substrates. However, the results obtained with Δapu_{Tk} were just the opposite. This may be the result of polar effects brought about by insertion of the hmg_{Tk} over-expression cassette. With the disruption strategy taken in this study, the downstream genes, in particular, $malK_{Tk}$ (TK1775), would be under the control of the *gdh* promoter and would thereby disturb the stoichiometric expression of the transporter subunits. Another possibility is that Apu_{Tk} itself is also a component of the sugar transporter complex, resulting in a decrease in stability or efficiency of the complex when Apu_{Tk} is absent. This is consistent with the fact that the apu_{Tk} gene is clustered within the subunit genes of the transporter itself. The possibility that amylopullulanase resides on the cell surface has been proposed in closely related hyperthermophilic archaea (30-32). The enzyme from *Thermococcus hydrothermalis* harbors a carboxy-terminal extension with three domains in addition to the central catalytic domain (31). One of the domains contains motifs with similarity to the S-layer homology signature (33), found in several bacterial proteins anchored to the cell surface

(33-36). The second region is extremely rich in threonine residues and is followed by the third, putative transmembrane domain. This architecture resembles the carboxy-terminal regions of S-layer proteins of the haloarchaea, in which the Thr-rich regions are targets for *O*-linked glycosylation and the transmembrane domain serves as a cell surface anchor (37, 38). The two regions are found in a number of Thermococcales proteins annotated as periplasmic components of ABC-type dipeptide transport systems, further supporting the involvement of these domains in cell surface attachment (30) (Fig. 7). As amylopullulanases from *P. furiosus*, *P. abyssi*, and *T. kodakaraensis* also harbor these domains, it can be presumed that the amylopullulanases from the Thermococcales are attached to the cell surface. Further biochemical examination will be necessary to clarify whether these amylopullulanases have any additional function besides their roles in poly- and oligosaccharide hydrolysis.

```
                                                                                  Putative
                                   Thr(Ser)-rich region                           transmembrane domain
                                                                                                    ••••
TK1774* (Apu_Tk)  TETETPTKTTTTTTSSTETTTTSTTETATTTTTTTTTSPGGGGSGSGSTTTSTSPGTGGGEEGGGICGPAFLVGLAVVPLLLRRRR
TK1760  (DdpA)    TTTTTTSETTTSKTPTSEKTNTGSGSTTSSSN-----------------------------GGGICGPAVVVGLAVIPLLLRRRR
TK1804  (DdpA)    ETKTETQTTTTTSETETSTQTTSETETQTTTTETSEE-------------------------GGGICGPAFLVGLAVVPLLLRRRR

Apu_Th            TPTESPTETTTTTPSETTTTTSTTTGPSSTTTSTP---------------------------GGGICGPGIIAGLALIPLLLKRRN
Apu_Pf            TPTQTETQTPTETRTETKTPTETTTTTPTETKETPTQTTTTQPARTET--------------QGGICGPGLIVLLAALGVL-RRRS
Apu_Pa            TGTTTTRTPTKTSTPTPEKTTPKTTKTKTETKESPSPSQTPPSAGAPPSGEERTTQK-----TGGICGPAFVLVIVAIVAIARKRF
Apu_Tl            SETETPTETESPTPSETSSTVSPSSTSSPSPTE-----------------------------TGGICGPAALVGLALIPLLLRRRW
```

Fig. 7. Carboxy-terminal regions of various amylopullulanase proteins from the Thermococcales, along with the corresponding regions of periplasmic components of two putative ABC-type dipeptide transport systems of *T. kodakaraensis*. All of these proteins harbor a threonine (or serine)-rich region, followed by a putative transmembrane domain and a stretch of basic residues (indicated by circles) at the extreme carboxy-terminus. The subscripts of the amylopullulanase proteins identify the source organism as follows (accession number): Apu_{Tl}, *Thermococcus litoralis* (BAC10983); Apu_{Th}, *Thermococcus hydrothermalis* (AAD28552); Apu_{Pf}, *Pyrococcus furiosus* (ABA33719); Apu_{Pa}, *Pyrococcus abyssi* (CAB49104). Accession numbers for TK1760 and TK1804 are BAD85949 and BAD85993, respectively.

SUMMARY

The author has developed a gene disruption system in the hyperthermophilic archaeon *Thermococcus kodakaraensis* using the antibiotic simvastatin and a fusion gene designed to over-express the 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase gene (hmg_{Tk}) with the glutamate dehydrogenase promoter. With this system, the author disrupted the *T. kodakaraensis* amylopullulanase gene (apu_{Tk}) or a gene cluster which includes apu_{Tk} and genes encoding components of a putative sugar transporter. Disruption plasmids were introduced into wild-type *T. kodakaraensis* KOD1 cells, and transformants exhibiting resistance to 4 μM simvastatin were isolated. The transformants exhibited growth in the presence of 20 μM simvastatin, and the author observed a 30-fold increase in intracellular HMG-CoA reductase activity. The expected gene disruption via double-crossover recombination occurred at the target locus, but the author also observed recombination events at the hmg_{Tk} locus when the endogenous hmg_{Tk} gene was used. This could be avoided by using the corresponding gene from *Pyrococcus furiosus* (hmg_{Pf}) or by linearizing the plasmid prior to transformation. While both gene disruption strains displayed normal growth on amino acids or pyruvate, cells without the sugar transporter genes could not grow on maltooligosaccharides or polysaccharides, indicating that the gene cluster encodes the only sugar transporter involved in the uptake of these compounds. The Δapu_{Tk} strain could not grow on pullulan and displayed only low levels of growth on amylose, suggesting that Apu_{Tk} is a major polysaccharide-degrading enzyme in *T. kodakaraensis*.

REFERENCES

1. **Stetter, K. O.**, Hyperthermophilic procaryotes. *FEMS Microbiol. Rev.*, 18, 149-158, 1996.
2. **Stetter, K. O.**, Extremophiles and their adaptation to hot environments. *FEBS Lett.*, 452, 22-25, 1999.
3. **Atomi, H.**, Recent progress towards the application of hyperthermophiles and their enzymes. *Curr. Opin. Chem. Biol.*, 9, 166-173, 2005.
4. **Egorova, K. & Antranikian, G.**, Industrial relevance of thermophilic Archaea. *Curr. Opin. Microbiol.*, 8, 649-655, 2005.
5. **Cline, S. W. & Doolittle, W. F.**, Transformation of members of the genus *Haloarcula* with shuttle vectors based on *Halobacterium halobium* and *Haloferax volcanii* plasmid replicons. *J. Bacteriol.*, 174, 1076-1080, 1992.
6. **Lam, W. L. & Doolittle, W. F.**, Shuttle vectors for the archaebacterium *Halobacterium volcanii*. *Proc. Natl. Acad. Sci USA.*, 86, 5478-5482, 1989.
7. **Krebs, M. P., Mollaaghababa, R. & Khorana, H. G.**, Gene replacement in *Halobacterium halobium* and expression of bacteriorhodopsin mutants. *Proc. Natl. Acad. Sci. USA*, 90, 1987-1991, 1993.
8. **Peck, R. F., DasSarma, S. & Krebs, M. P.**, Homologous gene knockout in the archaeon *Halobacterium salinarum* with *ura3* as a counterselectable marker. *Mol. Microbiol.*, 35, 667-676, 2000.
9. **Porat, I., Kim, W., Hendrickson, E. L., Xia, Q., Zhang, Y., Wang, T., Taub, F., Moore, B. C., Anderson, I. J., Hackett, M., Leigh, J. A. & Whitman, W. B.**, Disruption of the operon encoding Ehb hydrogenase limits anabolic CO_2 assimilation in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.*, 188,

1373-1380, 2006.

10. **Bardy, S. L. & Jarrell, K. F.**, Cleavage of preflagellins by an aspartic acid signal peptidase is essential for flagellation in the archaeon *Methanococcus voltae*. *Mol. Microbiol.*, 50, 1339-1347, 2003.

11. **Metcalf, W. W., Zhang, J. K., Apolinario, E., Sowers, K. R. & Wolfe, R. S.**, A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. USA*, 94, 2626-2631, 1997.

12. **Worrell, V. E., Nagle, D. P., Jr., McCarthy, D. & Eisenbraun, A.**, Genetic transformation system in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. *J. Bacteriol.*, 170, 653-656, 1988.

13. **Sato, T., Fukui, T., Atomi, H. & Imanaka, T.**, Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. *J. Bacteriol.*, 185, 210-220, 2003.

14. **Sato, T., Fukui, T., Atomi, H. & Imanaka, T.**, Improved and versatile transformation system allowing multiple genetic manipulations of the hyperthermophilic archaeon *Thermococcus kodakaraensis*. *Appl. Environ. Microbiol.*, 71, 3889-3899, 2005.

15. **Worthington, P., Hoang, V., Perez-Pomares, F. & Blum, P.**, Targeted disruption of the α-amylase gene in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J. Bacteriol.*, 185, 482-488, 2003.

16. **Atomi, H., Matsumi, R. & Imanaka, T.**, Reverse gyrase is not a prerequisite for hyperthermophilic life. *J. Bacteriol.*, 186, 4829-4833, 2004.

17. **Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H. & Imanaka, T.**,

Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis*. *Mol. Microbiol.*, 61, 898-909, 2006.

18. **Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H. & Imanaka, T.**, Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles. *J. Bacteriol.*, 186, 5799-5807, 2004.

19. **Schelert, J., Dixit, V., Hoang, V., Simbahan, J., Drozda, M. & Blum, P.**, Occurrence and characterization of mercury resistance in the hyperthermophilic archaeon *Sulfolobus solfataricus* by use of gene disruption. *J. Bacteriol.*, 186, 427-437, 2004.

20. **Wendoloski, D., Ferrer, C. & Dyall-Smith, M. L.**, A new simvastatin (mevinolin)-resistance marker from *Haloarcula hispanica* and a new *Haloferax volcanii* strain cured of plasmid pHV2. *Microbiology*, 147, 959-964, 2001.

21. **Cabrera, J. A., Bolds, J., Shields, P. E., Havel, C. M. & Watson, J. A.**, Isoprenoid synthesis in *Halobacterium halobium*. Modulation of 3-hydroxy-3-methylglutaryl coenzyme A concentration in response to mevalonate availability. *J. Biol. Chem.*, 261, 3578-3583, 1986.

22. **Bochar, D. A., Brown, J. R., Doolittle, W. F., Klenk, H.-P., Lam, W., Schenk, M. E., Stauffacher, C. V. & Rodwell, V. W.**, 3-Hydroxy-3-methylglutaryl coenzyme A reductase of *Sulfolobus solfataricus*: DNA sequence, phylogeny, expression in *Escherichia coli* of the *hmgA* gene, and purification and kinetic characterization of the gene product. *J. Bacteriol.*, 179, 3632-3638, 1997.

23. **Kim, D.-Y., Stauffacher, C. V. & Rodwell, V. W.**, Dual coenzyme specificity

of *Archaeoglobus fulgidus* HMG-CoA reductase. *Protein Sci.*, 9, 1226-1234, 2000.

24. **Rahman, R. N. Z. A., Fujiwara, S., Takagi, M. & Imanaka, T.**, Sequence analysis of glutamate dehydrogenase (GDH) from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1 and comparison of the enzymatic characteristics of native and recombinant GDHs. *Mol. Gen. Genet.*, 257, 338-347, 1998.

25. **Grochowski, L. L., Xu, H. & White, R. H.**, *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J. Bacteriol.*, 188, 3192-3198, 2006.

26. **Tachibana, Y., Leclere, M. M., Fujiwara, S., Takagi, M. & Imanaka, T.**, Cloning and expression of the α-amylase gene from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and characterization of the enzyme. *J. Ferment. Bioeng.*, 82, 224-232, 1996.

27. **Koning, S. M., Konings, W. N. & Driessen, A. J. M.**, Biochemical evidence for the presence of two α-glucoside ABC-transport systems in the hyperthermophilic archaeon *Pyrococcus furiosus. Archaea*, 1, 19-25, 2002.

28. **Fukui, T., Atomi, H., Kanai, T., Matsumi, R., Fujiwara, S. & Imanaka, T.**, Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes. *Genome Res.*, 15, 352-363, 2005.

29. **Lee, H.-S., Shockley, K. R., Schut, G. J., Conners, S. B., Montero, C. I., Johnson, M. R., Chou, C.-J., Bridger, S. L., Wigner, N., Brehm, S. D., Jenney, F. E., Jr., Comfort, D. A., Kelly, R. M. & Adams, M. W. W.**, Transcriptional and biochemical analysis of starch metabolism in the

hyperthermophilic archaeon *Pyrococcus furiosus. J. Bacteriol.*, 188, 2115-2125, 2006.

30. **Albers, S.-V., Koning, S. M., Konings, W. N. & Driessen, A. J. M.**, Insights into ABC transport in archaea. *J. Bioenerg. Biomembr.*, 36, 5-15, 2004.

31. **Erra-Pujada, M., Debeire, P., Duchiron, F. & O'Donohue, M. J.**, The type II pullulanase of *Thermococcus hydrothermalis*: molecular characterization of the gene and expression of the catalytic domain. *J. Bacteriol.*, 181, 3284-3287, 1999.

32. **Koning, S. M., Albers, S.-V., Konings, W. N. & Driessen, A. J. M.**, Sugar transport in (hyper)thermophilic archaea. *Res. Microbiol.*, 153, 61-67, 2002.

33. **Lupas, A., Engelhardt, H., Peters, J., Santarius, U., Volker, S. & Baumeister, W.**, Domain structure of the *Acetogenium kivui* surface layer revealed by electron crystallography and sequence analysis. *J. Bacteriol.*, 176, 1224-1233, 1994.

34. **Lemaire, M., Ohayon, H., Gounon, P., Fujino, T. & Béguin, P.**, OlpB, a new outer layer protein of *Clostridium thermocellum*, and binding of its S-layer-like domains to components of the cell envelope. *J. Bacteriol.*, 177, 2451-2459, 1995.

35. **Olabarría, G., Carrascosa, J. L., de Pedro, M. A. & Berenguer, J.**, A conserved motif in S-layer proteins is involved in peptidoglycan binding in *Thermus thermophilus*. *J. Bacteriol.*, 178, 4765-4772, 1996.

36. **Sára, M., Egelseer, E. M., Dekitsch, C. & Sleytr, U. B.**, Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus*

stearothermophilus PV72/p2. *J. Bacteriol.*, 180, 6780-6783, 1998.

37. **Lechner, J. & Sumper, M.**, The primary structure of a procaryotic glycoprotein. Cloning and sequencing of the cell surface glycoprotein gene of halobacteria. *J. Biol. Chem.*, 262, 9724-9729, 1987.

38. **Sumper, M., Berg, E., Mengele, R. & Strobel, I.**, Primary structure and glycosylation of the S-layer protein of *Haloferax volcanii*. *J. Bacteriol.*, 172, 7111-7118, 1990.

GENERAL CONCLUSIONS

In this study, the author has performed biochemical analyses on two membrane-anchored peptidases and genetic analyses on a sugar transporter in the hyperthermophilic archaeon, *Thermococcus kodakaraensis*.

Part I deals with the membrane-anchored peptidases of *T. kodakaraensis*. The author initiated the study in order to gain insight on how signal peptides are degraded in the *Archaea* after their release from the precursor secretion proteins. In bacteria, the enzymes responsible for the initial cleavage of signal peptides (signal peptide peptidase), and further degradation (oligopeptidase A and TepA) have been identified and characterized. Although closely related homologs were not present on the archaeal genomes, the author identified two genes on the *T. kodakaraensis* genome that encoded putative membrane-anchored peptidases ($SppA_{Tk}$ and $SppB_{Tk}$). Biochemical characterization of the catalytic domains of $SppA_{Tk}$ and $SppB_{Tk}$ revealed that these proteins indeed exhibited peptidase activity. Examinations on their substrate specificities revealed that both proteins do not recognize peptides with acidic residues, a feature that coincides with the fact that signal peptides do not harbor acidic residues. Furthermore, $SppA_{Tk}$ and $SppB_{Tk}$ displayed differences in substrate specificities that suggest the two enzymes complement one another in signal peptide fragmentation; $SppA_{Tk}$ prefers hydrophobic/aromatic residues at the P-3 position and residues with relatively small side chains at the P-1 position, whereas $SppB_{Tk}$ exhibited a strong preference for basic amino acid residues at the P-2 position and hydrophobic residues at the P-1 site. This suggests that $SppA_{Tk}$ is mainly involved in the cleavage of the h-domain of signal peptides rich in hydrophobic residues, while $SppB_{Tk}$ is mainly

responsible for the cleavage of the n-region, which contains basic and hydrophobic amino acid residues.

Although a large number of $SppA_{Tk}$ and $SppB_{Tk}$ homologs are present on the archaeal and bacterial genomes, no information had been available on their catalytic mechanisms. Through detailed site-directed mutagenesis studies, the author was able to determine the residues essential for proteolytic activity in both proteins. $SppA_{Tk}$ was found to exhibit activity through a Ser-Lys catalytic dyad comprised from Ser162 and Lys214. $SppB_{Tk}$ was dependent on a catalytic triad comprised from Ser130, His226 and Asp154. The results have clarified the catalytic mechanism of two relatively large groups of serine proteases.

In Part II, the author has performed a genetic study on a putative sugar transporter from *T. kodakaraensis*. Gene disruption of a gene cluster (TK1771-TK1774) and phenotype analysis of the disruptant strain revealed that the putative sugar transporter is the only transporter in *T. kodakaraensis* involved in the uptake of maltooligosaccharides. Gene disruption in this study was not performed with the previously developed system in *T. kodakaraensis* using auxotrophic mutant strains, but was based on a novel system utilizing antibiotic resistance.

As conventional antibiotics and antibiotic resistance marker genes cannot be used in hyperthermophiles due to their lack of thermostability, a strategy based on inhibition of a particular endogenous protein by an antibiotic and relieving the inhibition by over-expressing the protein was applied. The author found that gene disruption was possible using simvastatin, a specific inhibitor of 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, which is essential for archaeal membrane synthesis. An over-expression cassette of the HMG-CoA reductase

gene was used as the marker gene. As this system allows gene disruption and selection in nutrient-rich media, the author expects the system to be a valuable tool in future genetic studies on genes that are difficult to disrupt in minimal media. Furthermore, as the initial preparation of auxotrophic host strains is not necessary, this methodology should be helpful in developing gene disruption systems in other hyperthermophilic archaea.

LIST OF RELATED PUBLICATIONS

1. Reverse gyrase is not a prerequisite for hyperthermophilic life.
   Haruyuki Atomi, **Rie Matsumi,** and Tadayuki Imanaka.
   Journal of Bacteriology, 186(14), 4829-4833, 2004.

2. Complete genome sequence of the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1 and comparison with *Pyrococcus* genomes.
   Toshiaki Fukui, Haruyuki Atomi, Tamotsu Kanai, **Rie Matsumi**, Shinsuke Fujiwara, and Tadayuki Imanaka.
   Genome Research, 15(3), 352-363, 2005.

3. Biochemical properties of a putative signal peptide peptidase from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1.
   Rie Matsumi, Haruyuki Atomi, and Tadayuki Imanaka.
   Journal of Bacteriology, 187(20), 7072-7080, 2005.

4. Identification of the amino acid residues essential for proteolytic activity in an archaeal signal peptide peptidase.
   Rie Matsumi, Haruyuki Atomi, and Tadayuki Imanaka.
   The Journal of Biological Chemistry, 281(15), 10533-10539, 2006.

5. Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host/marker system based on antibiotic resistance.
   Rie Matsumi, Kenji Manabe, Toshiaki Fukui, Haruyuki Atomi, and Tadayuki Imanaka.
   Journal of Bacteriology, 189(7), 2683-2691, 2007.

6. Examination of the biochemical properties and catalytic residues of a novel membrane-bound peptidase from *Thermococcus kodakaraensis*.
   Rie Matsumi, Haruyuki Atomi, and Tadayuki Imanaka.
   In preparation.

OTHER PUBLICATIONS

1. Isolation and characterization of a novel poly(vinyl alcohol)-degrading bacterium, *Sphingopyxis* sp. PVA3.
   Atsushi Yamatsu, **Rie Matsumi**, Haruyuki Atomi, and Tadayuki Imanaka.
   Applied Microbiology and Biotechnology, 72(4), 804-811, 2006.

2. Atomic force microscopy dissects the hierarchy of genome architectures in eukaryote, prokaryote, and chloroplast.
   Ryosuke L. Ohniwa, Kazuya Morikawa, Joongbaek Kim, Toshiro Kobori, Koji Hizume, **Rie Matsumi**, Haruyuki Atomi, Tadayuki Imanaka, Toshiko Ohta, Chieko Wada, Shige H. Yoshimura and Kunio Takeyasu.
   Microscopy and Microanalysis, 13(1), 3-12, 2007.

3. Crystallization and preliminary X-ray crystallographic studies of the [NiFe] hydrogenase maturation proteins HypC and HypD.
   Satoshi Watanabe, **Rie Matsumi**, Haruyuki Atomi, Tadayuki Imanaka, and Kunio Miki.
   Acta Crystallographica Section F Structural Biology and Crystallization Communications, F63(6), 538-541, 2007.

4. Crystal structures of [NiFe] hydrogenase maturation proteins HypC, HypD, and HypE: insights into cyanation reaction by thiol redox signaling.
   Satoshi Watanabe, **Rie Matsumi**, Takayuki Arai, Haruyuki Atomi, Tadayuki Imanaka, and Kunio Miki.
   Molecular Cell, 27(1), 29-40, 2007.

5. Crystallization and preliminary X-ray crystallographic study of [NiFe]-hydrogenase maturation factor HypE from *Thermococcus kodakaraensis* KOD1.
   Takayuki Arai, Satoshi Watanabe, **Rie Matsumi**, Haruyuki Atomi, Tadayuki Imanaka, and Kunio Miki.
   Acta Crystallographica Section F Structural Biology and Crystallization Communications, F63(9), 765-767, 2007.